[ACCEPTED VERSION]

The Use of Technology in Sign Language Assessment: Results of a Pre-Pandemic Survey

Tobias Haug[1], Wolfgang Mann[2], & Franz Holzknecht[1]

[1]University of Teacher Education in Special Needs

Schaffhauserstrasse 239 / Postfach 5850

8050 Zurich

Switzerland

tobias.haug@hfh.ch / franz.holzknecht@hfh.ch

[2] University of Cologne

Faculty of Human Sciences

Education and Rehabilitation of the Deaf and Hard of Hearing

Klosterstr. 79b

50931 Cologne

Germany

w.mann@uni-koeln.de

Published version:

Haug, T., Mann, W., & Holzknecht, F. (2023). The use of technology in sign language testing: results of a pre-pandemic survey. *Sign Language Studies*, *23*(2), 243–281.

Abstract

This study is a follow up to previous research conducted in 2012 on computer-assisted language testing (CALT) which used a survey approach to investigate the use of technology in sign language assessment. The goal of the current study was in part to replicate the 2012 study and obtain updated information on the use of technology in sign language assessment. Additionally, the goal was to broaden the scope by also considering the use of automatic sign language recognition and generation in applied testing scenarios, through technologies that make use of artificial intelligence (AI). 32 sign language testing professionals participated in the current study. The results of the study confirm the findings from 2012, but also raise new issues for future sign language assessment, such as the use of automatic sign language recognition for automatic scoring and the problem of acquiring funding for developing and maintaining web-based sign language test platforms.

Keywords: sign language testing and assessment, computer-assisted language testing (CALT), web-based sign language testing, sign language technology

Introduction

Sign language assessment is a young sub-discipline that emerged within sign linguistics and deaf education (Haug, Mann, and Knoch 2022a). Interest in sign language assessment gained momentum in the 1990s with the development of tests to assess deaf children in bilingual schools (e.g., Hoffmeister 2000; Strong and Prinz 1997) as well as tests for linguistic research (e.g., Supalla et al. 1995), and some years later also tests to assess adult learners of a sign language (e.g., Caccamise and Samar 2009). For all of these types of tests, technology has always been an integral part of assessing sign languages (e.g., capturing sign language on video).

The presented study is a follow up to previous research done in 2012 (Haug 2015) who used a survey approach to investigate the use of technology in sign language assessment. The goal of the current study was to replicate the 2012 study and obtain updated information on this topic, but also to broaden the scope by also considering the use of automatic sign language recognition and generation in applied testing scenarios, through technologies that make use of artificial intelligence (AI).

The field of sign language learning and assessment has advanced considerably since the first survey was carried out in 2012, both in terms of research and application. This can be seen in part in the increasing number of research projects and publications on sign language assessment (e.g., Bochner et al. 2016; Haug et al. 2020; Hauser et al. 2015; Kanto, Syrjälä, and Mann 2020; Rinaldi et al. 2018), and conference and workshop contributions (e.g., Haug et al. 2018; Schönström and Holmström 2017), but also in the growing amount of practical training that is available, for example, related to sign language test development (e.g., Haug and Van den Bogaerde 2017). An increased interest in sign language assessment has also led to more dialog between experts from sign and spoken language assessment (Haug, Mann, and Knoch 2022b)

Over the past decade(s), technology has advanced rapidly and has impacted everyday life in many areas. Examples of this impact include the shift to exclusive online banking, the rising number of smart phone apps to pay online or in a shop (e.g., ApplePay), or the sharing of personal information via online devices, for example, the sharing of health-related information with insurance companies in return for a discount on health insurance. Yet, there is also a downside to these technological developments which has led to a more critical response in the media and research, specifically in relation to the impact of technological advances on privacy concerns or how the digital transformation of our societies affects job security (i.e., increasing automatization of industrial productions).

Similarly, there are both advantages and challenges related to the application of new technology within the context of sign language assessment. These may include, for example, enhanced sign language test delivery, streamlined recording of test results, automatic scoring and score reporting, as well as more efficient storage of and access to test results. As with technological developments in other domains, in sign language testing these developments also raise questions related to data protection as well as to ethical and legal concerns (e.g., the EU's General Data Protection Regulation [GDPR][i]). The investigation of these and other issues related to sign language assessment and technology was the focus of this study. The objectives were as follows:

(1) to gather data six years after the first survey from 2012 and compare the results on a set of selected variables (e.g., type of test delivery, scoring procedures),

(2) to extend the scope of the 2012 survey by also gathering information on the application of technologies related to AI, such as automatic sign language recognition, and

(3) to discuss advantages and challenges on the use of new technologies in sign language assessment.

Within the context of this article, we focus on different groups of learners, including deaf children acquiring a sign language as a first language, as well as deaf and/or hearing adults

learning a sign language as a second language. We use the term computer-assisted language testing (CALT) to refer to technology used in language assessment (e.g., Chapelle and Voss 2017; Suvorov and Hegelheimer 2014) with "a focus on technology for delivering tests and processing test takers' linguistic responses" (Chapelle and Voss 2017, 149).

Literature Review

Due to the limited availability of publications on sign language assessment and the use of technology, we will base the literature review on studies related to spoken language assessment and technology. The literature review will touch upon topics that were operationalized in the survey.

Language Testing and the Role of Technology

Already before the COVID-19 pandemic, the use of technologies in our daily lives had impacted language testing practices (Chapelle and Douglas 2006). Traditionally, spoken language tests have been delivered via different technologies, such as pre-recorded audio or video recordings of the instructions and tasks, or they have been aided through phone calls (Douglas 2012; Sawaki 2012). Over the years, these technologies have been advanced further to make language testing more efficient, for example, through computer-based test delivery or automatic scoring. There has also been an increasing number of publications on the use of technology in (spoken) language testing (for a review, see Plakans 2018). As a result, the increased use of technology in language testing has created new fields of enquiry. For example, researchers have started to explore the influence of technology on the language ability construct (Isbell and Kremmel 2020), on task development (e.g., using video recordings in listening tests; Batty 2021). At the same time, they have investigated possible limitations of technologies used for scoring such as the differences between automated and human scoring (Douglas 2009; Suvorov and Hegelheimer 2014).

In comparison to spoken languages where the use of video technology for assessment purposes is a more recent development (Batty 2014; Ockey and Wagner 2018), video

technology has always been central to sign language learning and assessment due to its visual modality. As a result of the technological advances made over the last decades, web-based sign language test delivery through videos has become common. Despite this progress, there are still some (technical) challenges with this approach, including poor Internet connectivity, lack of technical support at test sites, or lack of storage for online video recordings (Haug 2015).

In sum, CALT has introduced several benefits to help making language assessment more efficient and to offer new ways of incorporating multimedia materials to create new, more authentic task types. However, CALT has also raised new questions related to the construct under investigation (i.e., test takers' computer familiarity) or potential technical problems (e.g., poor Internet connectivity).

Natural Language Processing and Language Assessment

Natural language processing (NLP), the automatic recognition and analysis of speech sounds or written words, found its way into language assessment (Chapelle and Voss 2017). Within the context of this study, we will only focus on automatic speech scoring (and scoring of signed productions) rather than on scoring of written language, since sign languages do not have a widely used written form (Boyes Braem 2001). The number of available tests that make use of speech scoring technology has increased, either as the sole means of scoring or combined with human rating (Zechner and Evanini 2019). Due to increasing computing power in recent years, advances "such as the use of Deep Neural Network algorithms for training automatic speech recognition models, have brought about impressive gains in the performance and reliability of systems that make use of this technology" (Zechner and Evanini 2019, 3). NLP applied for automatic scoring could also reduce the costs of language proficiency testing (Dodigovic 2015).

Automatic sign language processing is a subfield of NLP. A prerequisite for automatic sign language processing is automatic sign language recognition, which includes recognizing the form and the meaning of a sign produced by a human signer (Ebling, Camgöz, and Bowden

2022). Both technologies (i.e., automatic sign language recognition and processing) can be compared to automatic speech processing although they are still fairly young in comparison and have not been applied to sign language assessment more widely (Ebling et al. 2018).

Advantages of Computer-Assisted Language Testing

Immediate feedback for the test taker: One advantage of CALT is that test takers can receive immediate feedback after completing a test (Cheng 2009; Silye and Wiwczaroski 2002). This feedback may be provided in form of a score report that can be printed out, is sent by email, or presented on the computer screen.

Logistics of test administration: With CALT, a test taker could, in theory, take a test at any time and in any location, provided they have access to a computer with Internet connection (Cheng 2009; Roever 2001). Time- and location-independent test administration is used both for low-stakes and high-stakes scenarios. One example of a high stakes test using CALT is the Duolingo English Test, which is recognized by many US universities as proof of English proficiency for admitting students whose native language is not English (e.g., Custer 2017). The Duolingo test is delivered completely online, and test takers are monitored via the webcam of their mobile device or computer. However, this format has also been criticized for several reasons (e.g., Wagner and Kunnan 2015; Wagner 2020). For instance, prior to the COVID-19 pandemic, computer-delivered high-stakes tests such as the Test of English as a Foreign Language (TOEFL) iBT had to be administered in certified test centers to avoid cheating. However, many testing companies have offered at-home solutions during the pandemic (for a review of high-stakes online testing during the pandemic, see Isbell and Kremmel 2020).

Authenticity and interaction in language use: With CALT it is possible to incorporate multimedia resources such as texts, images, videos, and audio which allow more fluid interaction between the test and the test taker and, thus, can increase the authenticity of test items (Suvorov and Hegelheimer 2014). For example, using videos in a listening test can lead

to more authentic assessment by making non-verbal cues in an L2 conversation accessible (Wagner 2010), even though this particular approach is not without challenges (see Batty 2014).

Adaptive testing: In a "traditional" non-CALT testing scenario, all test takers receive the same set of test items, regardless of their ability level (Sawaki 2012). In contrast, with computer-adaptive testing, the "examinee's rough ability estimate is obtained based on his/her performance on an initial set of items" (Sawaki 2012, 427). Based on this initial performance the appropriate difficulty level of the next item is estimated and a new item is selected from an item bank. "The examinee's ability estimate is constantly updated as he or she answers items" (Sawaki 2012, 427). This allows for more flexible and, potentially, shorter testing time (Mizumoto, Sasao, and Webb 2019) depending on the test taker's performance (Ockey 2009). This, in turn, can have a positive impact on test taker motivation because test takers are less likely being confronted with questions which are either too difficult or too easy for them (Linacre 2000; Meunier 1993). Finally, another advantage of computer-adaptive testing is that test taker proficiency is measured more accurately (Tseng 2016) as item difficulty is tailored to ability (Chapelle and Douglas 2006).

Test security: Although test security is often mentioned as a shortcoming of CALT (see discussion below), CALT can also enhance security. For example, CALT enables the creation of large item banks, which enable large-scale tests to provide different test versions for each testing session and thus reduce the possibility that items or sections of the test can be memorized by the test takers and passed on to peers.

Challenges of Computer-Assisted Language Testing

Computer familiarity of test takers: When tests are delivered via a computer or mobile device, test providers assume that test takers have a certain level of computer familiarity to successfully take a test and not be disadvantaged (Choi, Kim, and Boo 2003; Roever 2001). However, previous studies investigating the influence of the presentation mode of a test (paper-based vs. computer-based) on adult test takers' performance in language assessments did not

find any effects of computer familiarity (e.g., G. Yu 2010; W. Yu and Iwashita 2021) and report that students tend to prefer computer delivery (Brunfaut, Harding, and Batty 2018).

Security issues: As briefly mentioned above, in high-stakes scenarios, tests are usually administered in test centers where test takers' identity can be checked (Mubarak Pathan 2012). This is one drawback of tests that are self-administered at home, where it may be easier for candidates to fake their identities or get proxy candidates take the test in their names (see, for example, the recent language test center scandal in the UK; Main and Watson 2022). Purpura and colleagues address potential security issues in test proctoring, which may become more prevalent due to the pandemic as more tests are taken from home (Purpura, Davoodifard, and Voss 2021). Another security concern of CALT is that items from a test could be memorized and passed on to people taking the test at a later date (Ockey 2009). This happened in 2002 for the computer-based version of the Graduate Record Examination (GRE) offered by Educational Testing Service (Ockey 2009, 839). For web-based testing, hacker attacks may also pose a threat (Ockey 2009).

Technical expertise and infrastructure: To develop a CALT system, inter-disciplinary collaboration is needed, not only for constructing the test, but also for maintaining the hardware and software (Roever 2001). The hardware in use (e.g., the size of the computer screen) should fit the purpose of the test, for example, screens need to be large enough for the intended task (Mubarak Pathan 2012). As technical expertise and infrastructure are expensive, not all test providers may be able to develop a CALT system.

Computer-Assisted Sign Language Assessment

Limited research exists on the use of CALT within a sign language assessment context. Sign languages are visual-spatial languages that make use of the three-dimensional space, using manual (signs) and non-manual (e.g., facial expression) linguistic signals simultaneously to convey meaning (e.g., Boyes Braem 1995). Perhaps unsurprisingly, this format makes it more difficult, compared to spoken languages, to develop automatic systems for sign languages that

can recognize and score performances of test takers. To our knowledge, only one study has made use of automatic sign language recognition within the context of assessing vocabulary knowledge in adult L2 learners (Ebling et al. 2018).

Another issue related to the use of CALT for sign language assessment is the constant need for video-recording the test taker's performance whereas in case of spoken language tests a voice recording is often sufficient. But this is only the case for productive sign language assessment.

Research Questions

The study addresses the following research questions (RQ):

- RQ1: What are the differences and similarities between the 2012 and 2018 sign language assessment surveys in terms of the impact of technology on:
    - sign language skills tested,
    - test purposes,
    - test formats,
    - and test development?
- RQ2: What are advantages and challenges of using technologies in sign language assessment?
- RQ3: Which technologies are currently used in sign language testing?
- RQ4: What do sign language test developers and practitioners envisage for the future in this area?

Methodology

Survey Instrument

To answer the research questions of the current study, an updated version of the questionnaire from 2012 was used (Haug 2015). In this newer version most open-ended items of the 2012 questionnaire were changed into closed questions with pre-formulated responses,

drawing on the categorization of the responses from the 2012 survey. The final version of the questionnaire for the current study consisted of five parts (A-E), with a total of 26 questions:

- Part A: Background information about the survey participants (6 questions)
- Part B: Technical issues (e.g., problems with recording sign language production; skills tested, test purposes, the format used for test delivery, Internet security) (10 questions)
- Part C: Advantages and challenges that sign language test developers see regarding the use of CALT in sign language testing (2 questions)

Parts D and E included questions which were not part of the 2012 survey:

- Part D: The use of automatic sign language recognition and signing avatars (6 questions)
- Part E: Future scenarios of sign language testing (2 questions)

The questionnaire contained a combination of closed and open-ended questions. Closed-ended questions were used for items with pre-determined, limited answers (e.g., work position within an institution or ranking of possible advantages using CALT for sign language testing) whereas open-ended questions were used to gather more detailed information about the survey participants' opinion on a certain topic (e.g., to obtain more in-depth information about future scenarios). The survey was piloted with two colleagues involved in sign language testing. Their feedback contained only small changes to the wording of the instruction and questions. It took about 15 minutes to complete the survey. Similar to the approach used by Haug (2015), the survey was administered in English using LimeSurvey[ii], a free online survey tool. The complete survey is available at https://signlanguages.eu/index.php/366143?lang=en.

Procedure

The questionnaire was announced through different channels, including two different mailing lists in the international sign linguistics community, a Facebook group on sign linguistics, and through personal contacts. The survey was launched in summer 2018. Two reminders were posted within four weeks. The survey was closed in fall 2018.

Participants

Participants included language testing professionals worldwide who are involved in sign language test development and research. A total of 32 test developers from 18 different countries completed the survey (Table 1). The sample was considerably larger in comparison to the 2012 survey, which had been completed by 19 test developers (Haug 2015). This reflects the growing interest in sign language assessments within the international community of sign language practitioners including countries beyond the USA and the UK (both of which have traditionally dominated the research on sign language).

Table 1: Survey respondents by country

Countries of survey participants 2012 (*N* = 19) and 2018 (*N* = 32)

Country	n 2012	n 2018
Australia	1	1
Austria	N/A	2
Belgium	N/A	1
Canada	1	2
Czech Republic	1	N/A
Ethiopia	1	1
Finland	N/A	1
France	2	N/A
Germany	1	2
Guinea	N/A	1
Iceland	1	1
Ireland	N/A	1
Italy	N/A	1
Netherlands	2	2
New Zealand	N/A	1
Spain	N/A	3
Sweden	N/A	1

Turkey	N/A	1
UK	1	5

Similar to the 2012 survey, the majority of participants in the 2018 survey (n = 27) were affiliated with a university or university of applied sciences whereas the remaining participants were working in a variety of settings, including research institute (n = 1), communication center (n = 1), adult learning center (n = 1), or in a school setting (n = 2).

Data Analysis

A combination of quantitative and qualitative analyses was used to evaluate the survey data. To analyze the closed questions, we used descriptive statistics to calculate the proportions of responses across different response options assigned to a question. Responses to open-ended questions were organized by themes such as such as "Difficulties with technical infrastructure" or "Poor Internet connectivity" and given codes by the first author. After coding all replies, the frequency counts of each response were summarized. These categories were double-checked by the second author to establish inter-rater reliability. Cases of disagreements were discussed and the themes were adjusted accordingly. For some questions, only those categories with the highest frequency count are discussed.

As mentioned above, the development of certain multiple-choice answers was informed by responses to the 2012 version of the survey. For instance, the item "What are advantages using new technologies in sign language testing and assessment?" was originally an open-ended question in the 2012 survey and was redesigned into a multiple-choice question in the 2018 version.

Results

Skills Tested, Test Purposes, Test Formats, and Test Development

Figure 1 shows a direct comparison of the results from the 2012 and 2018 surveys in terms of participants' answers with regards to the sign language skills that were being tested.

As can be seen in the figure, numbers from each survey showed overall similar response patterns: most respondents were testing both receptive and productive skills, although responses from the 2018 survey showed a small but marked preference towards testing sign language production (88% production vs. 69% reception). In comparison to the 2012 version, the revised survey also included "language interaction" as a possible response. Roughly a third of all respondents (*n* = 12) selected this option, which suggests that there is a notable interest in assessing this area of sign language use.

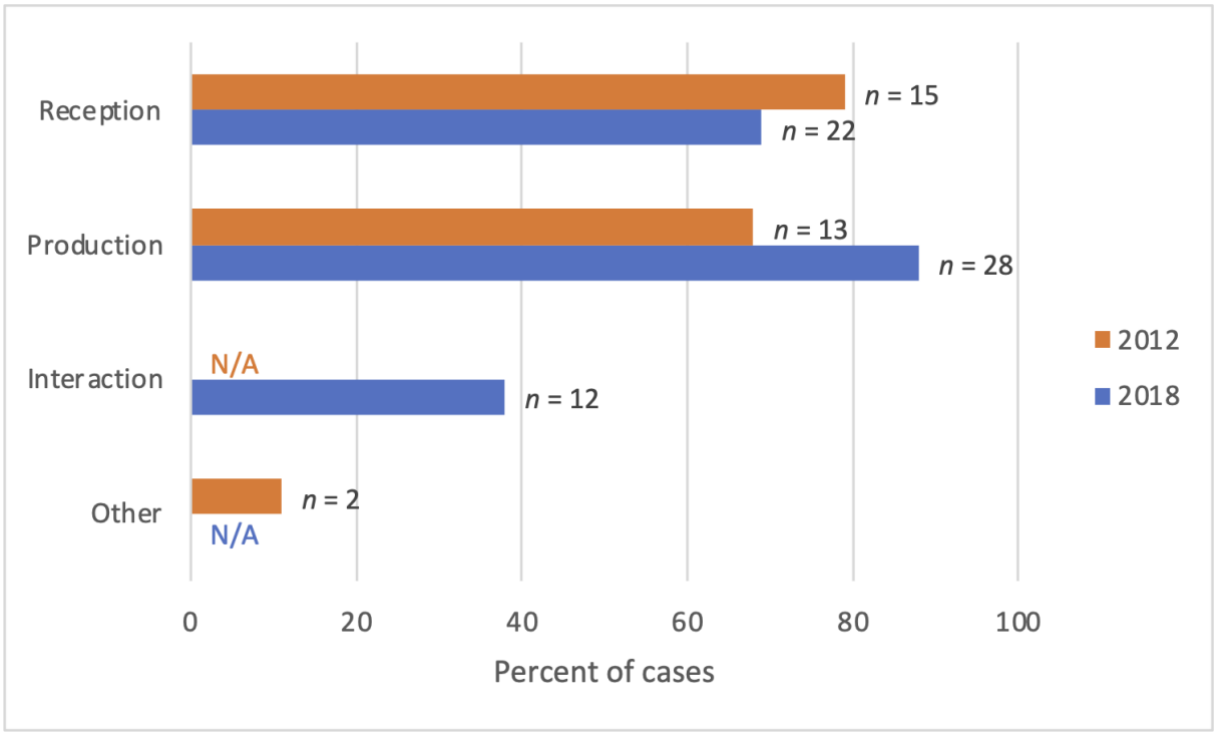

Figure 1

Sign language skills tested according to the 2012 and 2018 surveys

With regards to test purposes (Figure 2), as in the 2012 survey responses from 2018 showed an emphasis on testing language development of children, however proportionally this was a more frequent test purpose in 2012 (74% of respondents chose this option in 2012 compared to 41% in 2018). In comparison, assessing adult learners who acquire a sign language as their L2 (or additional language) was chosen as a test purpose by about 40 percent of respondents in both surveys, while sign language testing for linguistic research showed a

decrease between 2012 and 2018 (42% in 2012 vs. 13% in 2018). In 2018, 19 percent of respondent also indicated to use tests for other purposes not listed in the survey, including proficiency testing of sign language teachers (n = 1) and testing very specific aspects of sign language usage (n = 5) such as nonsense-sign repetition or mutual intelligibility tests between different sign languages.

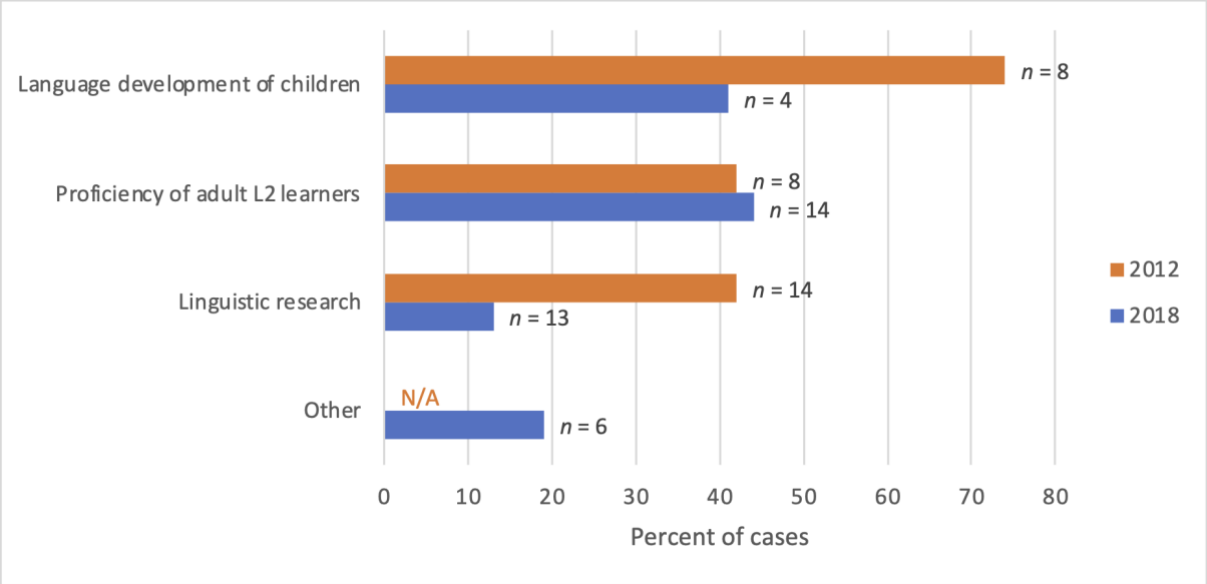

Figure 2

Test purposes in sign language assessment according to the 2012 and 2018 surveys

In terms of test formats, answers to the 2018 survey indicated a notable increase in respondents' use of video-supported and computer-based assessments compared to the 2012 version (Figure 3). Similarly, the number of web-based assessments increased between 2012 and 2018, albeit not as much. The integration of assessments in learning management systems such as Moodle, which is still new in sign language assessment, was also used by three respondents in 2018.

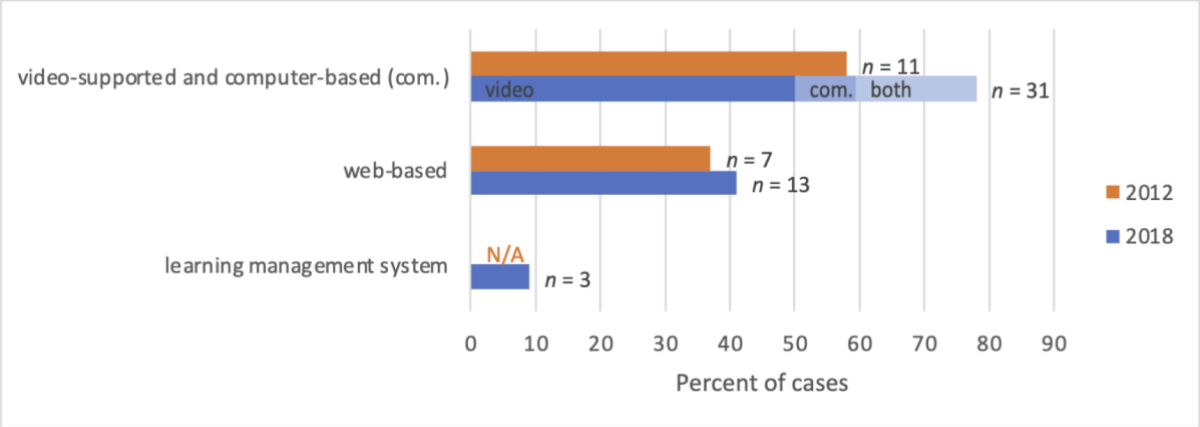

Figure 3

Test formats in sign language assessment according to the 2012 and 2018 surveys

Finally, the 2018 survey asked whether participants used tests for spoken languages and adapt them for sign language testing. A fourth of all respondents (n = 8) chose this option, which indicates a growing interest in this (sub-)area, particularly in countries where sign language research is less established than in the US or parts of Europe.

Hardware used for Sign Language Testing

The 2018 survey also included questions about hardware used for web-based testing (web-based testing was used by 13 respondents (41%), see Figure 3 above). Our results show a relatively balanced use across (mobile) devices. All respondents indicated that their web-based tests can be used on a laptop computer (n = 13), followed by tablets (n = 10) and mobile phones (n = 8).

Six of the respondents who indicated that they use web-based tests reported that these tests are hosted internally, five reported using external servers of a partner institution (e.g., another university or a non-profit organization, not a commercial provider), and two hosted their tests with a commercial provider. Only one respondent provided information on the financial resources (e.g., grant, income, department/institution) used to cover expenses linked to the development or maintenance of their web-based test.

Advantages using CALT in Sign Language Testing

In this question, as well as in the following question about challenges, the respondents were asked to rank statements about advantages (or challenges) of CALT for sign language testing. Participants did not have to rank all statements but could choose those which they regarded as most important. The responses were analyzed by calculating the weighted average rank across participants for each statement:

$$\frac{x_1 w_1 + x_2 w_2 + x_3 w_3 \ldots + x_n w_n}{Total\ number\ of\ responses}$$

where:

x = number of responses for the statement

w = weight of ranked position

Weights were assigned in descending order according to the total number of statements ranked by the participants. The total number of ranked statements was 11 for advantages and 12 for challenges. Thus, rank 1 was assigned a weight of 11 for advantages and a weight of 12 for challenges, rank 2 a weight of 10 for advantages and 11 for challenges, etc. with rank 11 for advantages and rank 12 for challenges corresponding to a weight of 1.

In this subsection, the results for advantages of CALT are presented (Figure 4). Seven participants ranked 11 statements[iii], 21 participants ranked 3 to 10 statements, and 1 participant ranked 2 statements. Three participants did not respond to this question. As shown in the figure, the highest ranked advantage of CALT for sign language assessment across participants was "ease of test administration" with a weighted average ranking of 8.90 out of a possible 11, followed by "standardized testing format improves reliability and validity" (8.39). In third position was the statement "easy to distribute and make test accessible" (8.27), with "user-friendliness" in close fourth position (8.25). Descriptive statistics for all statements including number and percentage assigned to each ranking position are included in the Appendix.

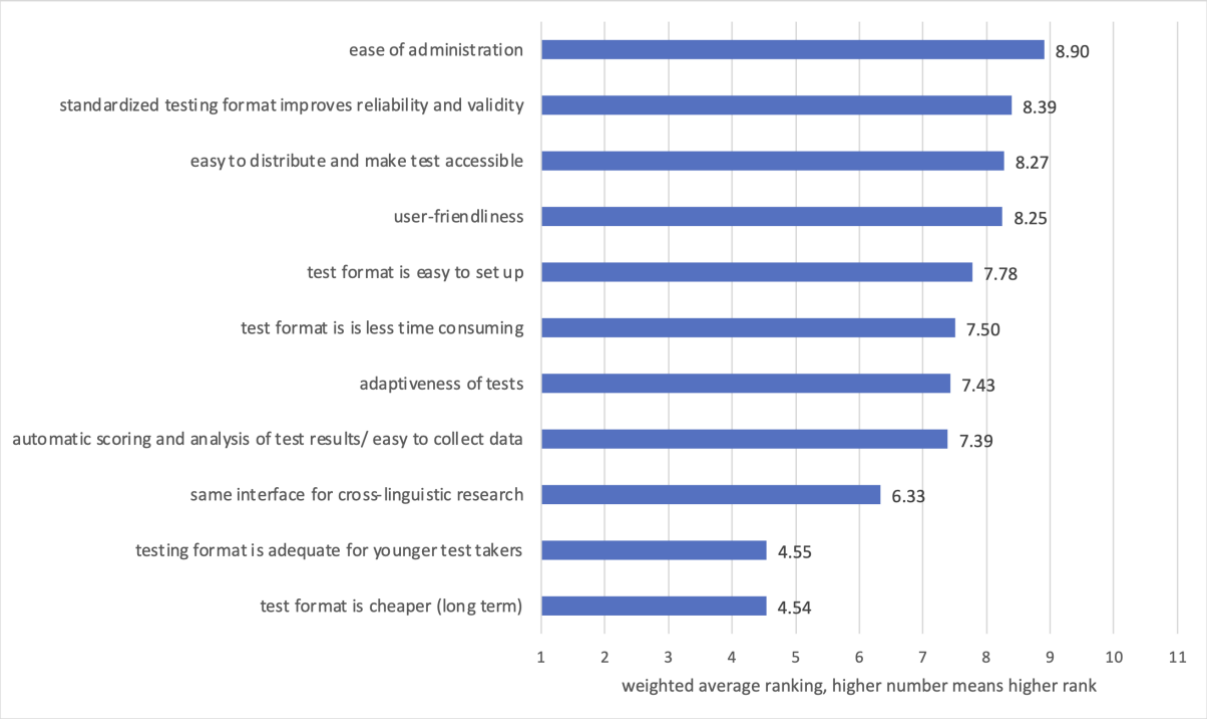

Figure 4

Weighted average ranking of advantages of CALT for sign language assessment

Challenges using CALT in Sign Language Testing

Figure 5 shows the weighted average ranking of the statements on challenges associated with CALT in sign languages. Six participants ranked 12 statements[iv], 15 participants ranked 3 to 7 statements, and 3 participants ranked 2 statements. Three participants only included one statement in their ranking, while five participants did not respond to the question. As shown in Figure 5, the four greatest challenges of CALT for sign language testing, according to the respondents, are "costs for developing such a test interface" with a weighted average ranking of 9.71 out of a possible 12, followed by "problems with quality of the videos" (9.64), "difficulties with technical infrastructure" (9.53), and "IT support at test site is missing" (9.44). Descriptive statistics for all statements including number and percentage assigned to each ranking position are again included in the Appendix.

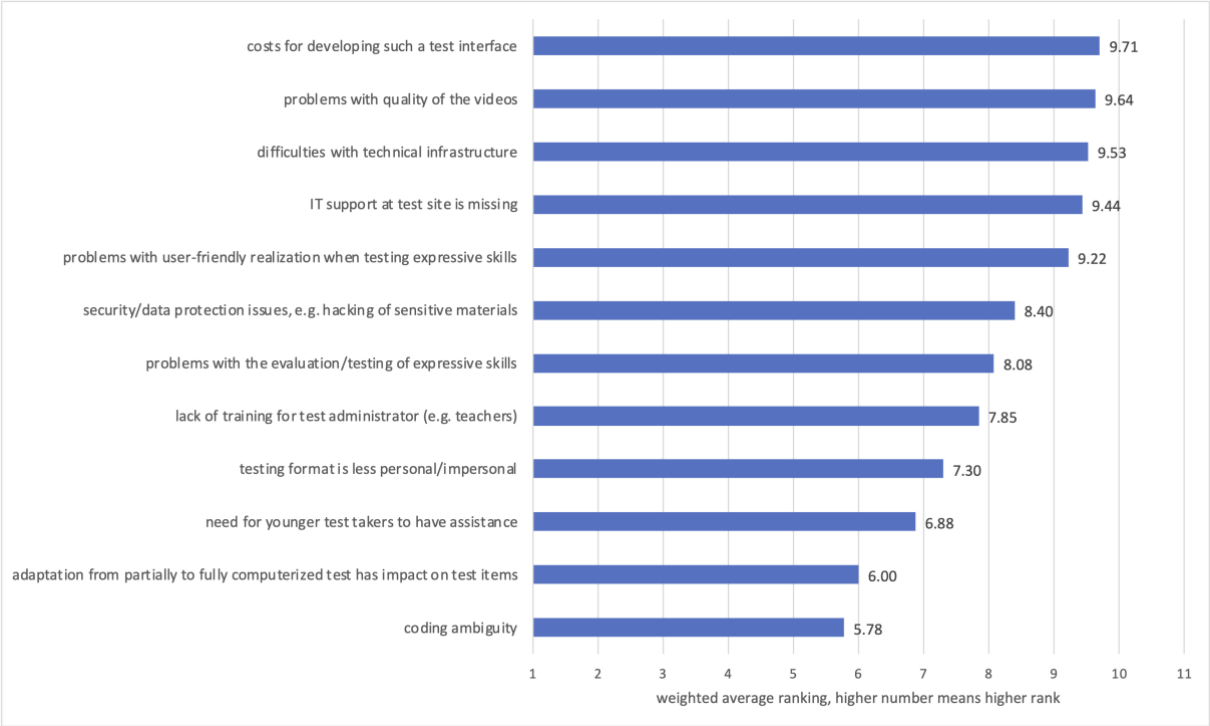

Figure 5

Weighted average ranking of challenges of CALT for sign language assessment

Automatic Sign Language Recognition and the Use of Avatars

Only one of the respondents indicated that they used automatic sign language recognition to assess sign language production of learners, which suggests that this type of technology was not yet available to most respondents at the time of data collection. This one respondent used technology for recognizing isolated signs but thought that the system should currently only be utilized as a practice tool for learners to aid classroom experience rather than in more formal testing scenarios. In addition, one respondent acknowledged recent advances in technology linked to automatic sign recognition that may soon allow at least partial processing of data, but they also pointed out that the quality of such recognition devices is currently not very good.

Similarly, only one respondent noted that they used signing avatars to present questions in an assessment scenario, thereby again indicating that this technology was not yet advanced enough for practical use in most testing contexts at the time of data collection.

Future Use of Sign Language Technologies in Sign Language Testing

Respondents were asked what they envisage for the future of CALT in sign languages. The 22 open answers were grouped according to common themes, with 2 answers applying to more than one theme, resulting in a total of 27 assigned responses. Out of these, 22 responses were related to advantages of CALT while the remaining 5 were related to associated challenges.

In terms of advantages, the most frequently mentioned theme was automatic recognition and evaluation of sign languages ($n = 9$), which respondents thought would not only help reduce teachers' workload but would also be useful for streamlining data collection. Another commonly mentioned advantage was that CALT could foster data mining and sharing, as well as sharing tests and test templates ($n = 6$). Two respondents also mentioned that a tool for commenting on learners' recorded videos would be useful for teachers. Other themes only mentioned once included the standardization of annotation, the possibility of remote assessment, the use of avatars for testing children, improved video quality, and improvements to survey tools such as Google Forms or SurveyMonkey in terms of integrating video.

With regards to challenges of CALT in sign languages, two respondents thought that the automated scoring of expressive tests will continue to be difficult. One respondent each mentioned that CALT may be useful for practice in classrooms but not for higher stakes testing, that developing CALT in sign languages is too expensive in many scenarios, and that teachers need to be trained in using new technologies.

Discussion

The comparison to the results from the 2012 and the current survey shows overall a similar response pattern regarding the testing of receptive and productive skills, with a small preference in the 2018 survey for testing productive skills. These results are not surprising as they show that different skills are being assessed (e.g., receptive, productive), reflecting sign language development in different groups of learners (e.g., deaf children, adults). The results

of the 2018 survey also demonstrate a notable interest in testing interaction, which was not included in the 2012 survey. It is encouraging that interaction seems to be seen as an important part of sign language education and testing, as this reflects the communicative competence paradigm in language education more generally (Canale and Swain 1980). Another area of language use that has recently gained more prominence through its inclusion in the CEFR Companion Volume (Council of Europe 2020), is mediation, and a future survey should also investigate its role in sign language teaching and testing.

As for test purpose, testing children was proportionally chosen more often in the 2012 than in the 2018 survey, testing adult L2 learners showed similar response patterns, and there was a decrease in tests for linguistic research. A possible explanation for the decrease of testing children in the 2018 survey might be the increased interest in sign language assessment in general with a subsequent specialization within sign language assessment. The same could apply for the decrease of testing for linguistic research: researchers don't rely as much on tests as method to obtain data for linguistic research as they used to a few years ago, more often corpus-linguistic approaches are used to conduct research on a specific sign language (e.g., for German Sign Language see Hanke 2019). A new finding is the information about testing specific aspects of sign language usage, such as a nonsense-sign repetition test. It will be interesting to see whether the testing of specific aspects, such as testing signing children for developmental language disorders, might be a consequence of an increased specialization within sign language assessment (for examples of such speciliazations, see Quinto-Pozos 2022 on assessing deaf children with developmental language disorder and Shield et al. 2022 for assessing sign language in children on the Autism Spectrum Disorder).

The results of the 2018 survey show that video-supported, computer-based, and web-based test formats have increased in usage in sign language assessment. This echoes a similar trend that has been observed for spoken language testing (Plakans 2018). At the same time, the use of videos is very common in sign language testing, including videos in comprehension tests

for spoken languages is a more recent development (e.g., Batty 2014). If this trend will continue after the pandemic in sign language assessment will be a matter of investigation in the future.

In 2018, survey participants were asked on which hardware their web-based tests can be used (n = 13). While all participants stated that their tests can be used on a laptop, ten and eight participants, respectively, indicated that their tests can also be used on a tablet or smart phone. This shows that test developers are more aware that their web-based tests should also be available on different (mobile) devices - which increases the flexibility of usage, for example, to test adult learners remotely at home. The trend towards using hand-held devices in sign language testing is likely to continue over the next years in different contexts, for example, for assessing deaf children in schools, as tablets are becoming more available in many educational settings (Haßler, Major, and Hennessy 2016). Since sign language tests always involve videos, attention needs to be paid to the size of the screen of hand-held devices so that signed items can still be understood easily. Most high-stakes English proficiency testing was conducted primarily at test centers before the COVID-19 pandemic, with the exception of the Duolingo English Test which could be taken completely online at home. In case testing companies will also provide at-home solutions for high-stakes testing in a post-pandemic world (for a review, see Isbell and Kremmel 2020), the format of test delivery on multiple devices might become an issue for future test development and research, including sign language assessment research.

The respondents who used web-based test formats also indicated where they hosted their tests, however, only one respondent provided information about the financial side of developing and hosting a web-based test. Thus, there is a lack of information on how sign language test developers manage to host and maintain sign language tests after the completion of a (research) project. Providing long term funding for web-based sign language tests is an important issue, which can only be resolved through collaboration across institutions or even countries (this might also explain why so few sign language tests are commercially available or offered by test

providers). One rare example is the sign language test portal run by the Deafness, Cognition and Language Research Centre[v] at University College London. A different approach could be that a consortium of institutions invests in the development of a sign language test platform that includes different kinds of test methods. The portal's code could be made available on an open-source platform so that institutions outside of the consortium can also host their tests on their own servers. Future developments would need to be financed by institutions using the platform/code.

In terms of advantages of CALT for sign language testing, respondents were asked to rank 12 statements in descending order of importance. The four top-ranked statements discussed here were "ease of test administration", "standardized testing format improves reliability and validity", "easy to distribute and make test accessible", and "user-friendliness". The most important advantage of CALT according to the participants is "ease of test administration". This mirrors discussions within the spoken language testing community, where test efficacy and expeditious test administration is generally regarded as a major advantage of CALT (Suvorov and Hegelheimer 2014). For sign languages, this advantage may be even more pronounced due to the necessary omnipresence of videos in most assessment formats.

The second most important advantage, "standardized testing format improves reliability and validity", refers to the central issue in all language testing endeavors. It is encouraging that sign language test developers regard this as an important advantage of CALT, especially seeing that today very few tests for sign language usage are validated and commercially available. Similar to language testing in spoken languages, standardized formats of computer- or web-based tests can contribute to the validation of sign language tests in the future (for spoken languages, see Mubarak Pathan 2012). However, it should be mentioned that CALT also introduces new challenges for establishing the validity of language tests, such as ensuring that test takers' computer familiarity does not interfere with the measurement of their language skills

(Suvorov and Hegelheimer 2014). This should be kept in mind whenever new computer-based assessment systems are introduced.

The third statement, "easy to distribute and make test accessible", refers to the flexibility of sign language test delivery. Using a computer-based test can streamline administration at different stages of data collection in a research project, for example, and can make on-site testing or remote testing at home more efficient. The same applies when a sign language test becomes commercially available. Depending on skills tested (e.g., vocabulary knowledge in a sign language, see Haug et al. 2022), a computer-based test can provide immediate feedback to the test taker (Cheng 2009) and can thus also be linked more directly to formative assessment to support learning.

The fourth most important advantage of CALT, according to the participants, was "user-friendliness". Thus, participants thought that computer-based tests were more user-friendly than conventional paper-based tests. However, it is important to keep in mind that user-friendliness depends on the setup of the test. Tests should be designed in a way that makes interaction with the interface easy and intuitive to avoid potential problems due to computer unfamiliarity in test takers. Recent research on spoken language testing in this area is encouraging, as it shows only minor differences in teenage and adult test takers' performances between paper-and-pencil vs. computer delivered tests (e.g., Brunfaut, Harding, and Batty 2018; W. Yu and Iwashita 2021). Test takers also prefer online test delivery to paper-and-pencil formats (Brunfaut, Harding, and Batty 2018). Computer familiarity might be less of a problem than it was 10 or 20 years ago as many children nowadays are more used to Internet devices from an early age (e.g., Danovitch 2019). While adaptive testing has been an issue of investigation and application for spoken languages for quite some time (e.g., Mizumoto, Sasao, and Webb 2019), the statement "adaptiveness of tests" was only ranked the 7th highest statement (of 11), most likely because it is a very new topics in sign language testing. To the best of our knowledge, only one adaptive test is currently available for a sign language (Bochner et al. 2016).

Respondents were also asked to rank statements associated with challenges of CALT. The four most often ranked statements were “costs for developing such a test interface”, “problems with quality of the videos”, “difficulties with technical infrastructure”, and “IT support at test site is missing”. The first statement, “costs for developing such a test interface”, links to an issue discussed above, i.e. that very little information is available on how sign language test developers finance web-based sign language tests. As the sign language community is fairly small it will remain difficult to obtain funding for the development and maintenance of a CALT system for sign languages. Therefore, as outlined above, it will be important that sign language test developers (and potential test providers) work together for setting up sign language test platforms that could work across different sign languages (e.g., Haug, Herman, and Woll 2015) and also implement technologies such as sign language recognition for automatic scoring that work sign language independent (e.g., Ebling et al. 2018).

The other challenges rated highest by the participants all relate to IT infrastructure and support. The statement “problems with quality of the videos” (rated second) is directly linked to the available technical infrastructure (rated third). Here it is crucial to ensure that well-functioning infrastructure is available before thinking about the implementation of computer-based or online testing. This has also been reported for spoken language testing (Roever 2001) and in the survey from 2012 (Haug 2015). Similarly, professional IT support at the testing site (rated fourth)) is a crucial variable for successful employment of CALT in sign languages and will hopefully become the norm in many contexts in the future.

Only one respondent provided information about the use of automatic sign language recognition, as part of natural language processing (NLP), for assessing sign language productions of isolated signs. This is a new topic in sign language assessment, but as technology improves it can be applied in a larger context (Ebling, Camgöz, and Bowden 2022). NLP has found its way into spoken language assessment some time ago (Chapelle and Voss 2017) and is used more and more in (semi)automatic scoring of speech tests (Zechner and Evanini 2019).

Automatic sign language recognition for the evaluation of sign language productions was also mentioned by nine respondents in the section on future scenarios of sign language technologies. Participants thought that the technology could be applied to reduce workload for teachers and to streamline data collection. Testing is an additional workload for the teachers in schools for the deaf, and semi-automatic scoring would leave time for other tasks in a classroom setting. The same is true for testing sign language proficiency in adult learners. Automated scoring can also reduce costs, as has been shown in research on spoken language testing (e.g., Dodigovic 2015). However, more research will be necessary to make automatic recognition and scoring of sign language productions more widely available.

Similarly, one respondent indicated that the automatic generation of sign language through avatars is not yet advanced enough to be incorporated into an applied testing scenario. But as the technology advances, signing avatars might be a valid option to generate sign language in tests for younger learners. In addition, the upcoming technology of creating photo-realistic avatars through the anonymization of human signers (Saunders, Camgöz, and Bowden 2021) could also have an impact on sign language assessment, for example, for anonymizing the production of a signing test taker for more objective rating.

Conclusion

The survey from 2018 shows a growing interest by sign language researchers and practitioners in the topic of using technologies for sign language assessment and it highlights some changes to the previous survey from 2012. The results show that CALT in sign languages is used for a variety of purposes in different contexts. The major advantages of using CALT in sign languages, according to the survey participants, are streamlined and more expeditious test administration and enhanced reliability and validity. While technological advances such as automatic sign language recognition for (semi)automatic scoring of signed productions can be seen as a chance for the future in applied testing scenarios, the cost of developing and maintaining a CALT system for sign languages remains to be a major challenge. Other

challenges such as a well-functioning IT infrastructure and technical support at the test site were also mentioned in both the 2012 and 2018 survey. Our goal is to monitor the development of CALT in sign language assessment in the future to update existing guidelines on the use of technology in sign language test development and administration (e.g., Haug et al. 2018).

References

Batty, Aaron Olaf. 2014. “A Comparison of Video- and Audio-Mediated Listening Tests with Many-Facet Rasch Modeling and Differential Distractor Functioning.” *Language Testing* 32 (1): 3–20. https://doi.org/10.1177/0265532214531254.

———. 2021. “An Eye-Tracking Study of Attention to Visual Cues in L2 Listening Tests.” *Language Testing* 38 (4): 511–35. https://doi.org/10.1177/0265532220951504.

Bochner, Joseph H., Vincent J. Samar, Peter C. Hauser, Wayne M. Garrison, Matt J. Searls, and Cynthia A. Sanders. 2016. “Validity of the American Sign Language Discrimination Test.” *Language Testing* 33 (4): 473–95. https://doi.org/10.1177/0265532215590849.

Boyes Braem, Penny. 1995. *Einführung in die Gebärdensprache und ihre Erforschung*. Vol. 11. Hamburg: Signum-Verlag.

———. 2001. “A Multimedia Bilingual Database for the Lexicon of Swiss German Sign Language.” *Sign Language & Linguistics* 4 (1–2): 133–43. https://doi.org/10.1075/sll.4.12.10boy.

Brunfaut, Tineke, Luke Harding, and Aaron Olaf Batty. 2018. “Going Online: The Effect of Mode of Delivery on Performances and Perceptions on an English L2 Writing Test Suite.” *Assessing Writing*, February. https://doi.org/10.1016/j.asw.2018.02.003.

Caccamise, Frank, and Vicent Samar. 2009. “Sign Language Proficiency Interview (SLPI): Prenegotiation Interrater Reliability and Rater Validity.” *Contemporary Issues in Communication Science and Disorders* 36: 36–47.

Canale, Michael, and Merrill Swain. 1980. “Theoretical Bases of Communicative Approaches to Second Language Teaching and Testing.” *Applied Linguistics* 1 (1): 1–47. https://doi.org/10.1093/applin/I.1.1.

Chapelle, Carol A., and Dan Douglas. 2006. *Assessing Language through Computer Technology*. Cambridge: Cambridge University Press.

https://doi.org/10.1017/CBO9780511733116.

Chapelle, Carol A., and Erik Voss. 2017. “Utilizing Technology in Language Assessment.” In *Language Testing and Assessment*, edited by Elana Shohamy, Iair G. Or, and Stephen May, 149–61. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-02261-1_10.

Cheng, Ke. 2009. “Computer-Based Language Testing: Perspectives Of The Past And The Future.” In *Proceedings of 2009 4th International Conference on Computer Science & Education*, 1764–68. IEEE. https://doi.org/10.1109/ICCSE.2009.5228276.

Choi, Inn-Chull, Kyoung Sung Kim, and Jaeyool Boo. 2003. “Comparability of a Paper-Based Language Test and a Computer-Based Language Test.” *Language Testing* 20 (3): 295–320. https://doi.org/10.1191/0265532203lt258oa.

Council of Europe. 2020. *Common European Framework of Reference for Languages: Learning, Teaching, Assessment. Companion Volume*. Strasbourg: Council of Europe Publishing.

Custer, Sara. 2017. “Duolingo English Test Gaining in US Admissions.” *The Pie News*, April 13, 2017. https://thepienews.com/news/duolingo-english-test-gaining-ground-in-us-admissions/.

Danovitch, Judith H. 2019. “Growing up with Google: How Children’s Understanding and Use of Internet-based Devices Relates to Cognitive Development.” *Human Behavior and Emerging Technologies* 1 (2): 81–90. https://doi.org/10.1002/hbe2.142.

Dodigovic, Marina. 2015. “Speech Recognition Technology in Language Testing: Does Taking the Test in an English-Speaking Environment Matter?” *Chinese Journal of Applied Linguistics* 38 (3). https://doi.org/10.1515/cjal-2015-0016.

Douglas, Dan. 2009. *Understanding Language Testing*. Oxon, UK: Hodder Education.

———. 2012. “Technology and Language Testing.” In *The Encyclopedia of Applied Linguistics*, edited by Carol A. Chapelle, wbeal1182. Oxford, UK: Blackwell

Publishing Ltd. https://doi.org/10.1002/9781405198431.wbeal1182.

Ebling, Sarah, Necati Cihan Camgöz, and Richard Bowden. 2022. “New Technologies in Second Language Signed Assessment.” In *The Handbook of Language Assessment Across Modalities*, edited by Tobias Haug, Wolfgang Mann, and Ute Knoch, 417–30. Oxford University Press. https://doi.org/10.1093/oso/9780190885052.003.0036.

Ebling, Sarah, Necati Cihan Camgöz, Penny Boyes Braem, Katja Tissi, Sandra Sidler-Miserez, Stephanie Stoll, Simon Hatfield, et al. 2018. “SMILE Swiss German Sign Language Data Set.” *11th Language Resources and Evaluation Conference (LREC 2018)*, 4221–29.

Hanke, Thomas. 2019. “Gebärdensprachgemeinschaft Und Gebärdensprachtechnologie.” PowerPoint presented at the Vortragreihe ISK, Interkantonale Hochschule für Heilpädagogik, Zurich, November 5.

Haßler, B., L. Major, and S. Hennessy. 2016. “Tablet Use in Schools: A Critical Review of the Evidence for Learning Outcomes: Tablet Use in Schools: A Critical Review.” *Journal of Computer Assisted Learning* 32 (2): 139–56. https://doi.org/10.1111/jcal.12123.

Haug, Tobias. 2015. “Use of Information and Communication Technologies in Sign Language Test Development: Results of an International Survey.” *Deafness & Education International* 17 (1): 33–48. https://doi.org/10.1179/1557069X14Y.0000000041.

Haug, Tobias, Aaron Olaf Batty, Martin Venetz, Christa Notter, Simone Girard-Groeber, Ute Knoch, and Mireille Audeoud. 2020. “Validity Evidence for a Sentence Repetition Test of Swiss German Sign Language.” *Language Testing* 37 (3): 412–34. https://doi.org/10.1177/0265532219898382.

Haug, Tobias, Sarah Ebling, Katja Tissi, Sandra Sidler-Miserez, and Penny Boyes Braem. 2022. “Development of a Technology-Assisted Assessment for Sign Language

Learning." *International Journal of Emerging Technologies in Learning (IJET)* 17 (06): 39–56. https://doi.org/10.3991/ijet.v17i06.26959.

Haug, Tobias, Rosalind Herman, and Bencie Woll. 2015. "Constructing an Online Test Framework, Using the Example of a Sign Language Receptive Skills Test." *Deafness & Education International* 17 (1): 3–7. https://doi.org/10.1179/1557069X14Y.0000000035.

Haug, Tobias, Wolfgang Mann, Eveline Boers-Visker, Jessica Contreras, Charlotte Enns, and Katherine Rowley. 2018. "Guidelines for Sign Language Test Development, Evaluation, and Use." http://www.signlang-assessment.info/tl_files/signlanguage/guidelines_sign_language_tests_haugetal_v1_2016-11-10.pdf.

Haug, Tobias, Wolfgang Mann, and Ute Knoch. 2022a. "Introduction: Why an Edited Volume on Signed and Spoken Language Assessment?" In *The Handbook of Language Assessment Across Modalities*, edited by Tobias Haug, Wolfgang Mann, and Ute Knoch, 1–12. Oxford University Press. https://doi.org/10.1093/oso/9780190885052.003.0001.

———, eds. 2022b. *The Handbook of Language Assessment Across Modalities*. 1st ed. Oxford University Press. https://doi.org/10.1093/oso/9780190885052.001.0001.

Haug, Tobias, and Van den Bogaerde. 2017. "Testing and Assessment in Classroom Environment, from the First Experiences to CEFR: Introduction." Presentation presented at the EALTA workshop: Assessing sign language in the clinical and classroom environment: Learning from experience exchange, ISSR, Rome, November 10.

Hauser, Peter C., Raylene Paludneviciene, Wanda Riddle, Kim B. Kurz, Karen Emmorey, and Jessica Contreras. 2015. "American Sign Language Comprehension Test: A Tool for Sign Language Researchers." *Journal of Deaf Studies and Deaf Education* 21 (1): 64–

69. https://doi.org/10.1093/deafed/env051.

Hoffmeister, Robert. 2000. “A Piece of the Puzzle: ASL and Reading Comprehension in Deaf Children.” In *Language Acquisition by Eye*, edited by Charlene Chamberlain, Jill Morford, and Rachel Mayberry, 143–63. Mahwah, NJ: Erlbaum.

Isbell, Daniel R., and Benjamin Kremmel. 2020. “Test Review: Current Options in at-Home Language Proficiency Tests for Making High-Stakes Decisions.” *Language Testing* 37 (4): 600–619. https://doi.org/10.1177/0265532220943483.

Kanto, Laura, Henna Syrjälä, and Wolfgang Mann. 2020. “Assessing Vocabulary in Deaf and Hearing Children Using Finnish Sign Language.” *The Journal of Deaf Studies and Deaf Education*, October, enaa032. https://doi.org/10.1093/deafed/enaa032.

Linacre, John Michael. 2000. “Computer-Adaptive Testing: A Methodology Whose Time Has Come.” *MESA Memorandum* 69: 1–58.

Main, Ed, and Richard Watson. 2022. “The English Test That Ruined Thousands of Lives.” BBC News. February 9, 2022. https://www.bbc.co.uk/news/uk-60264106.

Mizumoto, Atsushi, Yosuke Sasao, and Stuart A. Webb. 2019. “Developing and Evaluating a Computerized Adaptive Testing Version of the Word Part Levels Test.” *Language Testing* 36 (1): 101–23. https://doi.org/10.1177/0265532217725776.

Mubarak Pathan, Mustafa Mubarak. 2012. “Computer Assisted Language Testing (CALT): Advantages, Implications and Limitations.” *Researchvistas.Com* 1 (4): 30–45.

Ockey, Gary. 2009. “Developments and Challenges in the Use of Computer-Based Testing for Assessing Second Language Ability.” *The Modern Language Journal* 93: 836–47.

Ockey, Gary J., and Elvis Wagner. 2018. *Assessing L2 Listening: Moving towards Authenticity*. Vol. 50. Language Learning & Language Teaching. Amsterdam: John Benjamins Publishing Company. https://doi.org/10.1075/lllt.50.

Plakans, Lia. 2018. “Then and Now: Themes in Languages Assessment Research.” *Language Education and Assessment* 1 (1): 3–8. https://doi.org/10.29140/lea.v1n1.35.

Purpura, James E., Mashad Davoodifard, and Erik Voss. 2021. “Conversion to Remote Proctoring of the Community English Language Program Online Placement Exam at Teachers College, Columbia University.” *Language Assessment Quarterly*, January, 1–9. https://doi.org/10.1080/15434303.2020.1867145.

Quinto-Pozos, David. 2022. “Developmental Language Disorder and the Assessment of Signed Language.” In *The Handbook of Language Assessment Across Modalities*, edited by Tobias Haug, Wolfgang Mann, and Ute Knoch, 171–84. Oxford University Press. https://doi.org/10.1093/oso/9780190885052.003.0015.

Rinaldi, Pasquale, Maria Cristina Caselli, Tommaso Lucioli, Luca Lamano, and Virginia Volterra. 2018. “Sign Language Skills Assessed Through a Sentence Reproduction Task.” *The Journal of Deaf Studies and Deaf Education* 23 (4): 408–21. https://doi.org/10.1093/deafed/eny021.

Roever, Carsten. 2001. “Web-Based Language Testing.” *Language Learning & Technology* 5 (2): 84–94.

Saunders, Ben, Necati Cihan Camgöz, and Richard Bowden. 2021. “AnonySIGN: Novel Human Appearance Synthesis for Sign Language Video Anonymisation.” https://doi.org/10.48550/ARXIV.2107.10685.

Sawaki, Yasuyo. 2012. “Technology in Language Testing.” In *The Routledge Handbook of Language Testing*, edited by Glenn Fulcher and Fred Davidson, 426–37. London: Routledge.

Schönström, Krister, and Ingela Holmström. 2017. “Elicited Imitation Tasks (EITs) as a Tool for Measuring Sign Language Proficiency in L1 and L2 Signers.” Presentation presented at the 6th International ALTE Conference: Learning and Assessment: Making the connection, Bologna.

Shield, Aaron, Deborah Mood, Nicole Salamy, and Jonathan Henner. 2022. “Assessing Signed Language Development in Deaf/Signing Children with Autism Spectrum

Disorder." In *The Handbook of Language Assessment Across Modalities*, edited by Tobias Haug, Wolfgang Mann, and Ute Knoch, 131–44. Oxford University Press. https://doi.org/10.1093/oso/9780190885052.003.0012.

Silye, Magdolna, and Troy Wiwczaroski. 2002. "A Critical Review of Selected Computer Assisted Language Testing Instruments." http://www.date.hu/acta-agraria/2002-01i/fekete1.pdf.

Strong, Michael, and Philip M. Prinz. 1997. "A Study on the Relationship between American Sign Language and English Literacy." *Journal of Deaf Studies and Deaf Education* 2 (1): 37–46.

Supalla, Ted, Elissa Newport, Jenny L. Singleton, Samuel J. Supalla, Don Metlay, and Geoff Coulter. 1995. "An Overview of the Test Battery for American Sign Language Morphology and Syntax." Presentation presented at the Annual Meeting of the American Educational Research Association (AERA), San Francisco, CA, April 20.

Suvorov, Ruslan, and Volker Hegelheimer. 2014. "Computer-Assisted Language Testing." In *The Companion to Language Assessment*, edited by Antony John Kunnan, 594–613. Hoboken, NJ, USA: John Wiley & Sons, Inc. http://doi.wiley.com/10.1002/9781118411360.wbcla083.

Tseng, Wen-Ta. 2016. "Measuring English Vocabulary Size via Computerized Adaptive Testing." *Computers & Education* 97 (June): 69–85. https://doi.org/10.1016/j.compedu.2016.02.018.

Wagner, Elvis. 2010. "The Effect of the Use of Video Texts on ESL Listening Test-Taker Performance." *Language Testing* 27 (4): 493–513. https://doi.org/10.1177/0265532209355668.

———. 2020. "Duolingo English Test, Revised Version July 2019." *Language Assessment Quarterly* 17 (3): 300–315. https://doi.org/10.1080/15434303.2020.1771343.

Wagner, Elvis, and Antony John Kunnan. 2015. "The Duolingo English Test." *Language*

Assessment Quarterly 12 (3): 320–31. https://doi.org/10.1080/15434303.2015.1061530.

Yu, Guoxing. 2010. "Effects of Presentation Mode and Computer Familiarity on Summarization of Extended Texts." *Language Assessment Quarterly* 7 (2): 119–36. https://doi.org/10.1080/15434300903452355.

Yu, Wenjing, and Noriko Iwashita. 2021. "Comparison of Test Performance on Paper-Based Testing (PBT) and Computer-Based Testing (CBT) by English-Majored Undergraduate Students in China." *Language Testing in Asia* 11 (1): 32. https://doi.org/10.1186/s40468-021-00147-0.

Zechner, Klaus, and Keelan Evanini, eds. 2019. *Automated Speaking Assessment Using Language Technologies to Score Spontaneous Speech.* Milton: Routledge.

[i] http://data.europa.eu/eli/reg/2016/679/oj
[ii] www.limesurvey.org
[iii] The survey included 12 statements about advantages but 1 statement was not ranked by any of the participants ("Test format can be used as basic measurement for psycholinguistic experiments").
[iv] The survey included 13 statements about challenges but one statement was not ranked by any of the participants, that is, "Loss of linguistic information when children were asked to produce what they saw (bending over to computer)". The statement refers to a situation where a child might have bent over to the computer to see better what he/she should sign (i.e., to see what the task is) and that the child was also signing while bending over to the computer. This way the signing was not captured by the camera. The statement may not have been ranked because it was unclear to study participants.
[v] https://dcalportal.org/

Appendix

Descriptive statistics for the rank order of advantages of CALT for sign language testing

	1st	2nd	3rd	4th	5th	6th	7th	8th	9th	10th	11th	Total
ease of administration	7 33%	4 19%	4 19%	1 5%	1 5%	1 5%	2 10%	0 0%	1 5%	0 0%	0 0%	21 100%
standardized format improves rel. and val.	5 22%	4 17%	1 4%	8 35%	0 0%	2 9%	2 9%	0 0%	1 4%	0 0%	0 0%	23 100%
easy to distribute and make test accesible	4 18%	4 18%	3 14%	4 18%	2 9%	2 9%	1 5%	2 9%	0 0%	0 0%	0 0%	22 100%
user-friendliness	1 5%	6 30%	5 25%	2 10%	2 10%	2 10%	1 5%	0 0%	0 0%	1 5%	0 0%	20 100%
test format is easy to set up	5 28%	2 11%	2 11%	0 0%	3 17%	2 11%	1 6%	2 11%	0 0%	0 0%	1 6%	18 100%
test format is is less time consuming	1 6%	3 19%	3 19%	4 25%	1 6%	0 0%	0 0%	2 13%	1 6%	1 6%	0 0%	16 100%
adaptiveness of tests	1 7%	3 21%	1 7%	2 14%	2 14%	2 14%	2 14%	0 0%	0 0%	1 7%	0 0%	14 100%
automatic analysis / easy to collect data	3 13%	2 9%	3 13%	5 22%	3 13%	2 9%	1 4%	2 9%	0 0%	2 9%	0 0%	23 100%
potential to use same interface / database	0 0%	1 6%	4 22%	0 0%	5 28%	2 11%	2 11%	1 6%	1 6%	2 11%	0 0%	18 100%
format adequate for younger test takers	1 9%	0 0%	1 9%	1 9%	0 0%	2 18%	1 9%	0 0%	0 0%	0 0%	5 45%	11 100%
test format is cheaper (long term)	1 8%	0 0%	1 8%	0 0%	2 15%	0 0%	0 0%	1 8%	6 46%	1 8%	1 8%	13 100%

Descriptive statistics for the rank order of disadvantages of CALT for sign language testing

	1st	2nd	3rd	4th	5th	6th	7th	8th	9th	10th	11th	12th	Total
costs for developing such a test interface	4 24%	4 24%	3 18%	3 18%	1 6%	0 0%	0 0%	1 6%	0 0%	1 6%	0 0%	0 0%	17 100%
problems with quality of the videos	3 27%	4 36%	0 0%	1 9%	0 0%	1 9%	0 0%	2 18%	0 0%	0 0%	0 0%	0 0%	11 100%
difficulties with technical infrastructure	5 26%	3 16%	5 26%	3 16%	0 0%	1 5%	0 0%	0 0%	0 0%	1 5%	0 0%	1 5%	19 100%
IT support at test site is missing	6 38%	2 13%	2 13%	2 13%	0 0%	1 6%	1 6%	1 6%	0 0%	0 0%	0 0%	1 6%	16 100%
user-friendly realization of expressive skills	1 11%	1 11%	2 22%	1 11%	3 33%	1 11%	0 0%	0 0%	0 0%	0 0%	0 0%	0 0%	9 100%
security/data protection issues	2 13%	4 27%	3 20%	2 13%	0 0%	0 0%	0 0%	0 0%	1 7%	1 7%	1 7%	1 7%	15 100%
problems with the testing of expressive skills	1 8%	3 25%	1 8%	1 8%	0 0%	1 8%	4 33%	0 0%	0 0%	0 0%	1 8%	0 0%	12 100%
lack of training for test administrator	1 8%	2 15%	3 23%	1 8%	2 15%	1 8%	0 0%	0 0%	0 0%	0 0%	3 23%	0 0%	13 100%
testing format is less personal/impersonal	1 10%	1 10%	1 10%	0 0%	4 40%	0 0%	1 10%	0 0%	0 0%	0 0%	0 0%	2 20%	10 100%
younger test takers need assistance	2 25%	0 0%	0 0%	1 13%	0 0%	2 25%	0 0%	0 0%	0 0%	2 25%	1 13%	0 0%	8 100%
adaptation to computerized impacts test items	1 14%	0 0%	1 14%	0 0%	0 0%	0 0%	0 0%	1 14%	3 43%	1 14%	0 0%	0 0%	7 100%
coding ambiguity	0 0%	0 0%	0 0%	2 22%	1 11%	0 0%	2 22%	1 11%	2 22%	0 0%	0 0%	1 11%	9 100%