Summary of the International Exploratory Workshop

Methodological Challenges and Technical Innovations in Signed Language Assessment

1. Summary of the workshop

The presentations on the first day of the workshop served as a basis for the discussions in the three parallel sub-workshops on the second day and for the final discussion on the third day.

1.1 Summary of the presentations

Presentation 1: Key methodological issues in signed language test development for L1 and L2 learners

Robert Hoffmeister and Peter Hauser (both USA) reported on methodological issues in signed language test development for children/L1 learners and adults/L2 learners. For example, they talked about differences in the degree of language input that the tested deaf children had received. Only 5% of the deaf signing population can be considered “native signers”, the remaining 95% of the children receive a variety of (signed) language input. Another issue that the presenters stressed was the age at which deaf children receive their linguistic input, which has an impact on their language development and consequently on their success in the school context.

As for the adult learners, the shortage of research on the acquisition of a signed language in adult learners has an impact on test development. Little is known about the “typical” development of adult L2 learners of a signed language. The presenters further emphasized the importance of interdisciplinary cooperation during test development, such cooperation with psychologists, linguists, and members of the deaf community.

Presentation 2: What are differences and similarities in signed language test design for L1 and L2 learners?

Eveline Boers (The Netherlands) and Charlotte Enns (Canada) talked about the similarities and differences in test design for L1 and L2 learners of a signed language. One of the most striking differences is the age of the target groups, which has an impact on the testing materials used and, hence, on the test design. For example, images need to be age-appropriate when testing children, whereas more complex images like a map or a calendar can be used when testing adults. Another issue concerns the length of videos used in test tasks, i.e., a video for children should be shorter than one for adults. Moreover, the style/register of the signer on the video needs to match the target group. A test for children needs to be administered by an adult, whereas adults can be tested independently. Both presenters agreed on the potential of web-based testing formats. They also emphasized the need for future collaborations in signed language test development for both L1 and L2 learners. For example, they mentioned that it might be possible to share elicitation materials across signed languages.

Presentation 3: Applying a framework of computer-/mobile-assisted language testing (CALT) for spoken languages to signed language assessment

Tobias Haug (Switzerland) and Wolfgang Mann (UK) reported on their work in applying a CALT framework (Suvorov & Hegelheimer, 2014)\(^1\) to signed language test development. They focused on the following aspects of the framework:

1. Delivery format (computer- or web-based)

---

2. Media density (e.g., multimedia)
3. Scoring mechanism (human scoring, automatic scoring)
4. Task and response types (e.g., selected or constructed responses)

The presenters concluded that CALT offers a lot of potential for future applications in signed language testing, potential that has not yet been used for many signed language tests. For example, there are a number of web-based tests, but very few make use of, for example, online video recording (media density) or apply automatic score reporting for their (multiple-choice) tests. Also, many signed language tests often use selected-response tasks and response types, such as multiple-choice items, and very few tests make use of productive (e.g., short answers) or interactive (e.g., drag-and-drop) task types. The presenters emphasized that the framework they had applied can be used as a guidepost when thinking about the possibilities that web-based testing can offer. However, they recommended to always reflect critically on the extent to which, for example, automatic score reporting or more interactive task types can actually improve the authenticity, reliability and validity of a test.

Presentation 4: Automatic signed language recognition
Richard Bowden (UK) provided a comprehensive review on the development of signed language recognition technology. Older work of signed language recognition was based on a two-stage linguistic inspired framework where the first stage of recognition extracts high level features similar to those used in linguistic annotation and the second stage then recognizes the temporal order of these features. Later, Bowden discussed how this framework has progressed over the years to include the Hamburg Notation System, a form-based annotation system for signed languages, and modern sensors such as the Microsoft Kinect. In particular, he discussed in more detail some of the aspects of recognition including hand shape, movement and more recent pattern mining techniques for second stage temporal recognition. He then turned to faces and talked about facial feature tracking, e.g., expression and non-verbal facial communication in the context of speech including aspects of lip reading. Finally, he talked about employing mouthing’s in signed language recognition to bolster recognition performance.

Presentation 5: Automatic signed language generation
In the first part of their presentation, Sarah Ebling (Switzerland) and Robert Smith (Ireland) talked about the advantages and disadvantages of different approaches to automatic signed language generation, namely, (1) hand animation, (2) animation based on motion capturing, and (3) full synthesis. None of the three approaches offers the perfect solution. Hand animation and animation based on motion capturing offer high quality but are at the same time very time-consuming and not flexible, since they consist of finite sets of signed utterances. In contrast, the full synthesis approach offers more flexibility, as it does not consist of an inventory of pre-determined signed utterances. The disadvantage of this approach is the lower quality of the signing.

In the second part, Ebling and Smith reported on sign language animation acceptance and comprehension studies with respect to the three approaches mentioned above. The hand-animated avatars received the most positive feedback in these studies. The fully synthesized avatars were shown to be very stiff and robot-like. Avatars have so far not been used in signed language assessment but could offer children a more “game-like” experience. Here, the next step would be to conduct a study that assesses the comprehension and acceptance of signing avatars in a signed languages test.
1.2 Summary from the sub-workshops

On the second day, three parallel sub-workshops were offered. Each sub-workshop was provided twice, i.e., each workshop participant could visit two different sub-workshops. The sub-workshops dealt with the following topics:

1. Methodological issues in signed language test development
2. Computer-/mobile-assisted language testing (CALT) for signed language assessment and consequences of signed language testing for the learners

Summary of sub-workshop 1: Methodological issues in signed language test development

The following topics were discussed that will also inform the guidelines and have been used for the research agenda:

1. Qualifications of and training for test administrators and raters (for example, high proficiency in a signed language, test administration).
2. Qualifications of people involved in signed language test development: deaf people, practitioners, researchers, media design people, programmers etc.
3. Apply different tasks/items to assess the signed language development of L1 and L2 learners, such as vocabulary tests, judgment of sentence correctness, narrative tasks.
4. Experiment with task and response types that have not yet been used in signed language tests, such as cloze procedure, gap filling, or combining sentences.
5. Testing should be more tightly integrated with the curriculum (relating to the testing of deaf children).

Summary of sub-workshop 2: Computer-/mobile-assisted language testing (CALT) for signed language assessment

The following issues related to CALT, automatic signed language recognition, and generation were discussed in this sub-workshop and will also inform the guidelines and has been used for the research agenda:

1. Automatic signed language recognition and the use of avatars offer the possibility of anonymization for the test takers, both children and adults. Deaf communities are known to be rather small, therefore this possibility could increase, for example, intra-rater reliability.
2. Use of automatic signed language recognition systems for automatic scoring in tests of productive signing, i.e., the recognition system rates the production of the test taker. Still, the results would need to be checked by human raters due to the variability that is acceptable, for example, in a sentence repetition test.
3. Internet bandwidth poses a problem for the use of videos, maybe use vimeo or other services for minimizing the problem of different video formats.
4. General technical issues, such as server connectivity, video formats and file size, quality of webcams; provide good-practice examples in the guidelines.
5. Signed language tests should be developed by a team that contributes expertise from different areas, such as the technical side, but also expertise on signed languages and test construction.
6. Investigate to what extent interaction can be assessed using, for example eye-tracking technology.
7. Use of a signed language assessment portal that can be used across signed languages, by different user groups such as practitioners, researchers, test takers, test administrators. This portal should also include the possibility to provide feedback to the learner.
8. Test developers need to get a clear understanding of what is technically possible; this should inform the guidelines.
9. Gamification, especially for the testing of children.
Summary of sub-workshop 3: Consequences of signed language testing for learners

The consequences for children as L1 learners are different (e.g., start with an intervention) than for adult L2 learners of a signed language (e.g., passing/failing a final exam). The following issues were considered as important, separately for L1 and L2 learners:

Consequences for L1 learners:
1. Consequences of an intervention for a deaf child, or placement in a different class/unit within a school
2. Very few signed language tests are available worldwide, most of them do not account for the different sub-groups of deaf signed language learners.
3. The need for signed language tests to form an established part of schools for the deaf that work bilingually
4. Availability of L1 tests and information about parents would affect educational policies.
5. Lack of access to a language (spoken and/or signed) has an impact on deaf children’s lives.

Consequences for L2 learners:
1. Impact of consequences for adult L2 learners, i.e., what are the stakes, for example, for teachers whose job may depend on the test results or a promotion.
2. Higher level of language proficiency according to the Common European Framework of Reference is often lacking, both for teaching and learning, but also for assessment.
3. Self-assessment should be used more often to support signed language learning.
4. The importance of providing feedback on how signed language proficiency can be improved to the learners; importance of gaining feedback on signed language competencies in, for example, signed language interpreting students from the local deaf community
5. Development of tests that tap into the linguistic skills and world knowledge of the language learners, i.e., the question of whether we are testing language or other skills/abilities.
6. Develop course materials and assessment procedures for hearing parents of deaf children who start learning a signed language.

The results of this sub-workshop will contribute to an ongoing debate on consequences of signed language assessment, which has not yet been addressed in the literature.

1.3 Outputs: Results of final discussion

It was decided at the workshop to set up an international steering group that will contribute to the guidelines. Eight workshop participants are interested in joining this steering group, consisting of deaf and hearing experts, with background in testing children or adults. The interested persons will meet via Skype in January 2016.

In what follows, the aims of the Exploratory Workshop will be revisited and it will be explained how they have been achieved. Minutes were taken during the final discussion, which will feed into the guidelines and will be distributed in January to all workshop participants. The guideline will be made available as laid out in the proposal and be submitted to a journal for publication.

**Aim 1**

To **develop and disseminate methodological guidelines** for the development of signed language tests by combining cutting-edge research and expertise of an interdisciplinary team of both experienced and starting deaf and hearing researchers involved in signed language test development.

During the final discussion it was decided that, among others, the following topics needed to be included in the guidelines:
1) Check existing guidelines as to whether they can be applied to signed language test development, for example, the “Assessment Guidelines” of the European Forum of Sign Language Interpreters (efsli, 2013)\(^2\) and different guidelines from the International Testing Commission\(^3\). Also check the ethical guidelines from the Sign Language Linguistics Society\(^4\).

2) The importance of having deaf people involved in teams developing tests (from development to dissemination), e.g., deaf native signers with academic training or deaf people that can be trained.

3) Having people with different areas of expertise involved in the development of a test, such as linguists, psychologists, high-level interpreters, media design experts, computer programmers.

4) The guidelines should be available in English and International Sign.

5) Include results from corpus projects to inform linguistic variation and change of signed languages.

6) The test format should match the research questions and the aims of a test’s target group (e.g., size of population, language, heterogeneity, usage).

7) Technical issues: provide good-practice examples for video formats etc.

8) Ethical issues, e.g.,
   a) Approval of test taker (or test taker’s legal representative)
   b) Data have to be encrypted when they are put on a web-based signed language portal (e.g., “SSL certificate”).

9) Rely on most recent research (e.g., linguistic theory) when developing a test.

10) Maximize test usability.

11) Community approval (face validity).

### Aim 2

To identify key issues in signed language test development that can help to form the basis for an **interdisciplinary and international research agenda**. These key issues will cluster around the following two domains: (a) methodological issues in signed language test development and (b) selected aspects of CALT (delivery formats, response type, task type, media density, scoring mechanism) applied to signed languages.

Workshop participants discussed the following topics to be included in the international research agenda:

1. Usability studies on different types of response and task formats
2. Use of a web-based portal for signed language tests across countries that can serve different purposes (e.g., application, research).
3. Acceptance studies of avatar-generated test items
4. Use of different types of tests, such as more objective (e.g., multiple-choice) or subjective tests (e.g., interview form) and their impact on test performance
5. Use of self-assessment in adult learners and how self-assessment enhances the learning of a signed language.
6. Approach testing of signed interaction

### Aim 3

To **discuss consequences of signed language testing for the learners** (e.g., to start signed language intervention with deaf children; consequences of test results for adults L2 learners in an interpreter training program).

---


\(^3\) [https://www.intestcom.org/page/5](https://www.intestcom.org/page/5)

The consequences of signed language testing for the learners have been discussed and documented in the minutes of the workshop. The minutes will be made available to all workshop participants in January. Some workshop participants were interested in following up on this topic in the form of a publication.

### Aim 4

To strengthen international research collaboration and networks.

There were many opportunities for the workshop participants to interact with each other and strengthen collaboration and networks.

### 1.4 Milestones after the workshop

All but one milestones were achieved (see box below). The milestone that is still pending achievement concerns the creation of guidelines. The guidelines are expected to be ready at the end of February 2016. They will be sent to the SNSF.

#### Milestones

- September 18, 2015: Submission of results of sub-workshops to workshop organizer (by workshop chairs).
- September 18, 2015: Setting up a core group to finalize the guidelines and research agenda.
- October 16, 2015: Finalizing guidelines and research agenda and sending out feedback to workshop participants (allowing two weeks time).
- October 30, 2015: Incorporating feedback of workshop participants in final version of the guidelines and research agenda, preparing for dissemination.
- November 20, 2015: Dissemination of final documents.

### 2. Future collaboration

During the workshop, participants had the possibility to network and discuss future collaborations. The following collaborations emerged:

1. A usability study on avatar-generated test tasks.
2. Linking dynamic assessment to the learning of adult learners of a signed language.
3. Applying an interview-like test to children.

### 3. Interpretation

The interpretation English/International Sign contributed greatly to the accessibility of this event for the deaf workshop participants and secured the interaction between them and their hearing peers, which benefited both sides.
4. Appendix: Final program

Day 1: September 4, 2015

<table>
<thead>
<tr>
<th>Time</th>
<th>Title</th>
<th>Presenter(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>9.00-9.30</td>
<td>Welcome note</td>
<td>Tobias Haug</td>
</tr>
<tr>
<td>9.30-10.30</td>
<td>Key methodological issues in signed language test development for L1 and L2 learners (aim 1 &amp; 2)</td>
<td>Robert Hoffmeister (L1) &amp; Peter Hauser (L2)</td>
</tr>
<tr>
<td>10.30-11.00:</td>
<td>Coffee break</td>
<td></td>
</tr>
<tr>
<td>11.00-12.00</td>
<td>What are differences and similarities in signed language test design for L1 and L2 learners?</td>
<td>Charlotte Enns (L1) &amp; Eveline Boers (L2)</td>
</tr>
<tr>
<td>12.00-13.30:</td>
<td>Lunch</td>
<td></td>
</tr>
<tr>
<td>13.30-14.30</td>
<td>Applying a framework of computer-/mobile-assisted language testing (CALT) for spoken languages to signed language assessment (aims 2 &amp; 4)</td>
<td>Tobias Haug &amp; Wolfgang Mann</td>
</tr>
<tr>
<td>14.30-15.30</td>
<td>Automatic signed language recognition (aim 2)</td>
<td>Richard Bowden</td>
</tr>
<tr>
<td>15.30-16.00:</td>
<td>Coffee break</td>
<td></td>
</tr>
<tr>
<td>16.00-17.00</td>
<td>Automatic signed language generation (aim 2)</td>
<td>Sarah Ebling &amp; Robert Smith</td>
</tr>
<tr>
<td>17.00-17.15h</td>
<td>Summary of the day</td>
<td>Jörg Keller</td>
</tr>
<tr>
<td></td>
<td>Dinner, self organized</td>
<td></td>
</tr>
</tbody>
</table>

Day 2: September 5, 2015

<table>
<thead>
<tr>
<th>Time</th>
<th>Title</th>
<th>Presenter(s) / Workshop facilitator(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>9.00-9.30</td>
<td>Information from the organizer and information about the sub-workshops</td>
<td>Tobias Haug</td>
</tr>
<tr>
<td></td>
<td>Parallel sub-workshops, first round</td>
<td></td>
</tr>
<tr>
<td>9.30-11.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11.00-11.30:</td>
<td>Coffee break</td>
<td></td>
</tr>
<tr>
<td>11.30-12.30</td>
<td>Continuation of sub-workshops 1-3</td>
<td></td>
</tr>
<tr>
<td>12.30-14.00:</td>
<td>Lunch</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Parallel sub-workshops, second round</td>
<td></td>
</tr>
<tr>
<td>Time</td>
<td>Title</td>
<td>Presenter(s)</td>
</tr>
<tr>
<td>------------</td>
<td>----------------------------------------------------------------------</td>
<td>--------------------------------------------------</td>
</tr>
<tr>
<td>14.00-15.30</td>
<td>Sub-Workshop 1: Methodological issues in signed language test development</td>
<td>Annieck van den Broek &amp; Evelien Boers (with interpreters)</td>
</tr>
<tr>
<td></td>
<td>Sub-Workshop 2: Computer-/mobile-assisted language testing (CALT) for signed language assessment</td>
<td>Jon Henner &amp; Patrick Boudreault (in International Sign)</td>
</tr>
<tr>
<td></td>
<td>Sub-Workshop 3: Consequences of signed language testing for the learners</td>
<td>Geoff Poor &amp; Rosalind Herman (with interpreters)</td>
</tr>
<tr>
<td>15.30-16.00</td>
<td>Coffee break</td>
<td></td>
</tr>
<tr>
<td>16.00-17.00</td>
<td>Continuation with sub-workshops 1-3</td>
<td></td>
</tr>
<tr>
<td>17.00-17.30</td>
<td>Summary of the day/wrap-up</td>
<td>Tobias Haug &amp; Wolfgang Mann</td>
</tr>
<tr>
<td>17.30-ca. 20.00</td>
<td>Apéro riche (&quot;dinner&quot;)</td>
<td></td>
</tr>
</tbody>
</table>

**Day 3: September 6, 2015**

<table>
<thead>
<tr>
<th>Time</th>
<th>Title</th>
<th>Presenter(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>9.00-9.15</td>
<td>Information from the organizer</td>
<td>Tobias Haug</td>
</tr>
<tr>
<td>9.15-9.45</td>
<td>Summary of sub-workshops 1</td>
<td>Sub-workshop participant(s)</td>
</tr>
<tr>
<td>9.45-10.15</td>
<td>Summary of sub-workshop 2</td>
<td>Sub-workshop participant(s)</td>
</tr>
<tr>
<td>10.15-10.45: Coffee break</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10.45-11.15</td>
<td>Summary of sub-workshop 3</td>
<td>Sub-workshop participant(s)</td>
</tr>
<tr>
<td>11.00-13.00</td>
<td>Final discussion and decisions</td>
<td></td>
</tr>
</tbody>
</table>