American Sign Language Comprehension Test: A Tool for Sign Language Researchers

Peter C. Hauser*,1,2, Raylene Paludneviciene2,3, Wanda Riddle3, Kim B. Kurz1, Karen Emmorey2,4, and Jessica Contreras1,2

1National Technical Institute for the Deaf, Rochester Institute of Technology; 2NSF Science of Learning Center on Visual Language and Visual Learning; 3Gallaudet University; and 4San Diego State University

*Correspondence should be sent to Peter C. Hauser, National Technical Institute for the Deaf, Rochester Institute of Technology, 52 Lomb Memorial Drive, Rochester, NY 14623 (e-mail: peter.hauser@rit.edu).

Abstract

The American Sign Language Comprehension Test (ASL-CT) is a 30-item multiple-choice test that measures ASL receptive skills and is administered through a website. This article describes the development and psychometric properties of the test based on a sample of 80 college students including deaf native signers, hearing native signers, deaf non-native signers, and hearing ASL students. The results revealed that the ASL-CT has good internal reliability (α = 0.834). Discriminant validity was established by demonstrating that deaf native signers performed significantly better than deaf non-native signers and hearing native signers. Concurrent validity was established by demonstrating that test results positively correlated with another measure of ASL ability (r = .715) and that hearing ASL students’ performance positively correlated with the level of ASL courses they were taking (r = .726). Researchers can use the ASL-CT to characterize an individual’s ASL comprehension skills, to establish a minimal skill level as an inclusion criterion for a study, to group study participants by ASL skill (e.g., proficient vs. nonproficient), or to provide a measure of ASL skill as a dependent variable.

Sign language research has grown exponentially since investigations in the 1960s and 1970s demonstrated that linguistic phenomena were not limited to the spoken modality (Klima & Bellugi, 1979; Stokoe, 1960). For over four decades, sign language research has contributed significantly to our understanding of language, human cognition, and socialization. Many of these sign language studies were conducted on the primary visual language used by deaf people in North America, American Sign Language (ASL). The number of ASL signers has been growing, and ASL is now the third most frequently offered college modern language course in the United States (Goldberg, Looney, & Lusin, 2015). There has been a great need for sign language researchers to identify the effects of sign language fluency on neuroplasticity, development, bilingualism, et cetera. However, ASL researchers face a major challenge in this endeavor because there are no readily available standard measures of ASL competency.

Efforts to develop a test of ASL competency are not new. Most of these tests were developed in individual sign language researchers’ laboratories for use in their own research. This includes tests that measure deaf children’s ASL development as their first language (L1), proficiency tests for deaf adults who learned ASL as their L1, and proficiency tests for hearing signers who learn ASL as their second language (L2) (see Singleton & Supalla, 2011 for reviews of ASL tests). There are many challenges for developing tests for ASL proficiency because it is not possible to simply translate English-based tests into ASL. The development of ASL assessments must take into consideration our current understanding of the structure of ASL based on linguistic research and current practices in psychological testing (see Paludneviciene, Hauser, Daggett, & Kurz, 2012 for a discussion on the specific challenges in sign language test development).

The emergence of ASL tests has made it possible for researchers to set cut-off criteria for study inclusion based on...
ASL fluency for adults (e.g., Morford, Wilkinson, Villwock, Piñar, & Kroll, 2011), to report their participants’ sign language fluency in neuroimaging studies (Corina, Layer, Hauser, & Hirshorn, 2013; Hirshorn, Dye, Hauser, Supella, & Bavelier, 2014), and to begin to identify and characterize ASL disorders (Quinto-Pozos et al., 2013). ASL fluency has also been used as a continuous or dependent variable in reading research (Myers et al., 2010; Twitchell, Morford, & Hauser, 2015) and bilingual studies (Freel et al., 2011; Morford et al., 2011).

Existing ASL tests are not available for distribution because they involve highly trained interviewers and raters (Hauser, Paludneviciene, Supalla, & Bavelier, 2008; Hoffmeister, 1999; Maller, Singleton, Supalla, & Wix, 1999; Supalla, Hauser, & Bavelier, 2014; Supalla et al., 1995). The transferability of the assessment system to different laboratories would require standardized rater training and replication of the tests’ psychometric properties to confirm equivalence. The goal here was to develop a brief, psychometrically sound ASL receptive test—the American Sign Language Comprehension Test (ASL-CT)—that can provide a score reflecting proficiency in ASL immediately upon completion of the test without relying on raters.

This test was also developed to be appropriate for hearing and deaf L1 and L2 signers (college-age adults), which would make it possible to compare the ASL comprehension skills of these four groups in the same experiment. We hope that such a tool will enable more sign language researchers to consider the effects of ASL competency on different variables when conducting basic science, translational science, or applied science research. This article describes how the ASL-CT items were developed and selected, and the psychometric soundness of the test (including reliability and validity). The discussion section also explains how ASL-CT raw scores can be converted to standard scores for studies that wish to document how well participants comprehend ASL compared to deaf native signers.

Method

Participants

In the first phase of the psychometric examination of the ASL-CT, we chose to focus on college students for two reasons: (a) this population is included in many psycholinguistic, neurolinguistic, and sociolinguistic research studies, and (b) the homogeneity of education and age minimizes the inclusion of confounding variables, which makes it possible to have a smaller sample size without sacrificing statistical power. Eighty college students (M_{age} = 20.6; SD_{age} = 1.7) were recruited for this study. Forty participants had deaf parents (20 deaf native signers and 20 hearing native signers), and 40 participants had hearing parents (20 deaf non-native signers and 20 hearing ASL learners). The deaf non-native signers were schooled in a variety of educational settings, and the mean age of ASL exposure was 6.6 years old (SD = 6.78). Hearing ASL learners were college students recruited from ASL courses at Rochester Institute of Technology (ASL 1: n = 3; ASL 2: n = 2; ASL 3: n = 5; ASL 5: n = 3; ASL 6: n = 5; two did not report their ASL course level). All four groups were gender balanced (50% female). There were no significant differences between the deaf and hearing participants’ mean ages, F(3, 76) = 0.039, p = .943, no significant differences between the native and non-native signers’ mean ages, F(3, 76) = .355, p = .553, and no significant interactions, F(3, 76) = 1.930, p = .169.

Procedures

The development of the ASL-CT involved (a) item development; (b) administration of beta items to all participants; (c) analyses of individual items for inclusion in the final version of the test (Item Selection in Results section); and (d) psychometric analyses of the final 30 items of the test. First, three of the test developers, all deaf native signers, met frequently throughout an academic year to develop possible items for this test. Their goal was for the ASL-CT items to include grammatical aspects of ASL such as phonology (sensitivity to small variations in form), vocabulary, role shift, and depicting verbs/constructions. Depicting verbs “visually” show an action that occurs in mental space (Dudis, 2004, 2008; Liddell, 2003). For example, ASL has complex verbs that depict a physical action using one hand to represent an agent causing the event and another hand to represent an entity undergoing a change. The team reviewed videos of natural conversations as a starting point to identify possible depicting constructions. In their decision making process, they kept in mind that test development required three foils (incorrect responses) that would be similar to the target response but with subtle changes in sign phonology, vocabulary choice, classifier morphology, facial expression, or use of space. Items for the pilot ASL-CT varied widely in difficulty in order to avoid floor and ceiling effects. Possible items were then recorded by the team and reviewed, modified, and rerecorded.

One of the team members was the model for the first set of 65 items and the foils. Each item was presented on a website with the prompt on the top and four choices on the bottom (see Figure 1).

![Example Item 1](image1.png)

Example Item 1

![Example Item 2](image2.png)

Example Item 2

Figure 1. The figure shows two sample ASL-CT items. These are examples only and are not included in the actual test. Participants used their computer mouse to click on the correct response. The response choice letters are for illustration only and are not presented in the test. For example 1), the correct choice is (A), and the foils correspond to classifier signs that differ from the target sign in final movement (B), final location and movement (C), or hand configuration (D). For example 2), the correct video is (C) which shows the correct handshape and movement for using the pictured hose, and the foils differ in (A) handshape (showing the wrong orientation for using the hose), (B) movement (showing the hose moving in a small circle), or (D) handshape (the sign WATCH-MOVIE which differs only in handshape with the correct classifier sign).
The prompt and choices were sometimes videos of the team members signing, videos of an activity, or a line drawing. For example, some items presented a line drawing (e.g., a dog laying in a corner) or a video of an event (e.g., water running out of a hose) as the prompt and four signed descriptions as response choices (see Figure 1). Some of the items presented a signed description of a scenario as the prompt (e.g., describing a volcano exploding) and either (a) four line drawings as the response choices (see Figure 1), (b) four videos of actions (with no signing) as response choices, or (c) four signed comments related to the prompt as choices. Participants saw one item at a time, and the next item was shown after a response was given. Participant responses were recorded on a secure server and were available to be downloaded as an excel spreadsheet.

Participants were recruited via flyers posted on the campus of Gallaudet University in Washington DC and the National Technical Institute for the Deaf (NTID), as well as through email announcements to students. After the first contact, the research team asked participants via email if they had any history of visual difficulties, motor difficulties, neurological disorders, or learning disorders, and no participants reported having any of these conditions. The participants were recruited from NTID and Gallaudet because these schools are attended by a large number of signers who were raised in different regions across North America. All participants consented, were tested individually, and were paid for their participation. The same sign model for the test stimuli presented instructions in ASL. English voice over audio was also available to participants who clicked a radio button requesting it.

Participants were tested individually and were told that they would be given an ASL comprehension test and that the purpose of the study was to further develop the test and to document the test's reliability and validity. Participants were instructed to click on one of the four response choices on the bottom of the screen that best matched the prompt video, picture, or line drawing at the top of the screen. Participants were allowed to view the videos as often as they liked, but they were not able to go back after they had selected their answer. The instructions included one sample item and two practice items followed by explanations of the correct choice. Participants’ responses to individual items and the total number of correct responses were recorded. Participation took ~45 min to complete the 65-item test, and they were compensated $20 for their time.

A subsample of 49 participants (14 deaf native signers; 14 deaf non-native signers; 12 hearing native signers; 9 hearing ASL students) were also administered the American Sign Language Sentence Reproduction Test (ASL-SRT; Hauser et al., 2008; Supalla et al., 2014). The ASL-SRT is a 20-item test of ASL fluency that required participants to watch ASL sentences on video—one at a time—and repeat back the sentence in ASL. The ASL-SRT has high inter-rater reliability (R = .83) and internal consistency (Chronbach alpha = .875). Its discriminant validity was established by demonstrating that deaf native signers perform better than non-native signers (p < .001) and deaf adults perform better than deaf children (p < .05) (Hauser et al., 2008). A deaf native signer trained on the ASL-SRT scoring procedure scored all of the responses. Following the standard test procedure, any reproductions with phonological, lexical, or syntax errors are scored as zero, and the test score is the total number of correct reproductions.

Standardized English comprehension tests, including reading tests, that are used by educational and psychology researchers include a normative sample in order to provide a normative index of English comprehension. Such samples include native English speakers only, because the goal of these tests is to measure competency in English. The same approach was utilized in the development of this test, as deaf native signers are considered the “Gold Standard” when it comes to ASL comprehension. In this study, hearing native signers were grouped separately from the deaf native signers because the two groups have different language experiences. The “24/7 effect” can be found with deaf native signers, due to the frequency of use of ASL and the tendency of deaf native signers to have a larger social network of ASL users compared to those who are not deaf (see Paludneviciene et al., 2012, for discussion).

Results

Item Selection

Our goal was to select the best 30 items for the final version of the test. The first step in the item selection was a review of the deaf native signers’ performance. Items that were most difficult for deaf native signers were reviewed, and it was determined that items that involved reversals of locations in signing space between the target and foils were particularly problematic and therefore were removed. For these items, the target response required a 180° rotation of the location(s) in signing space in order to match the model signer’s viewpoint; however, participants could interpret these items from a “shared space” perspective in which the model is understood as describing a jointly viewed scene and thus such a mental transformation is not needed (see Emmorey & Tversky, 2002). In addition, items that appeared confusing (e.g., items for which there was more than one possible correct answer) were also removed. Further, any items in which the foils were either too easy (i.e., all participants chose the target item) or too hard (i.e., a large proportion of native signers chose one or more foils rather than the target item) were removed. In the remaining items, two out of the three foils for all test items were selected by at least one participant, indicating that these foils were effective. Second, an analysis of variance was computed for each item illustrating the probability of detecting group differences between the four groups of participants. Items where all participants performed similarly were omitted, including items with both ceiling and floor effects.

The Chronbach alpha was then computed for the remaining items (N = 57) using all 80 participants’ data. Chronbach’s alpha is a function of the number of test items in a test, the average covariance between item-pairs, and the variance of the total score. In other words, the Alpha represents the average correlation of the items with each other. If the items were not clear and were thus not good estimates of ASL comprehension skills, there should be a wider variation between participants in their answers indicating more “noise” or errors in the test as a measurement of ASL competency. Such a situation would have a negative impact on the Chronbach’s alpha and would yield a low Alpha. The change in Chronbach’s alpha if an item was omitted was also computed. The 30 items with the highest Chronbach’s alpha then were selected for the final version of the test.
Characteristics of the Final 30-Item Version of the ASL-CT

The ASL-CT has three practice items that are presented with feedback. Half of the test items present a line drawing (N = 11) or an event video (N = 4) as the prompt with four signed response choices. The other half of the items present a signed description as the prompt with (a) four signed response choices (N = 9), (b) four event video choices (N = 3), or (c) four line-drawing choices (N = 3). Items were randomized for a fixed presentation order across all participants. All items involve the use of depictive constructions in which at least one “classifier” sign is used. Twelve items involve role shift or constructed action (e.g., shifting from the perspective of one referent to another; using facial expression to illustrate the opinion of a person performing an action); twelve items require the participant to understand subtle variations in the form of a depictive (non-lexical) sign (e.g., variations in the orientation of a whole entity handshape or form variations that create minimal contrasts between depictive signs and lexical signs); eleven items require understanding the appropriate handling handshape or whole entity handshape; seven items require understanding the locative and referential use of signing space; and one item requires understanding list buoys (Liddell, 2003). These item characteristics are not mutually exclusive, for example, a test item might involve understanding both role shift and selecting the appropriate classifier handshape for a given action.

Reliability

Chronbach’s alpha was used to determine the internal consistency of the final 30 ASL-CT items. Chronbach’s alphas between 0.6 and 0.7 are viewed as acceptable (Kline, 2000). The 30-item ASL-CT has a mean alpha of .834, which is above the acceptable threshold and suggests that the test has good internal reliability.

Concurrent Validity

One way concurrent validity is established is when the results of a new test correlate well with another measure of the same construct (i.e., ASL) that has already been validated. The ASL-SRT was used here to establish the concurrent validity of the ASL-CT although the two tests differ in how ASL competency is measured. The ASL-CT is a receptive test while the ASL-SRT involves both receptive and expressive ASL skills. The ASL-CT was found to positively correlate with the ASL-SRT, r(47) = .715, p < .001, which demonstrates that both tests measure an individual’s ASL skills. Another way concurrent validity is established is when the results of a new test correlate well with an independent criterion of ASL skill, and here we used level of ASL classes (ASL 101–ASL 106). For the 18 hearing students who reported which ASL class they were currently enrolled in, their ASL-CT performance positively correlated with their ASL class level, r(16) = .726, p < .005 (see Table 1 for the mean performance and standard deviations for all of the participant groups on the ASL-CT).

Discriminant Validity

Another approach to establish validity is to demonstrate whether the test is able to discriminate between groups with different language experiences. Three groups of signers (deaf native signers, deaf non-native signers, and hearing native signers) were used to determine the discriminant validity of the ASL-CT. Deaf native signers are exposed to ASL from birth and achieve the same language milestones as hearing individuals learning a spoken language (e.g., Lillo-Martin, 2009; Newport & Meier, 1985). In contrast, deaf non-native signers do not have the same early language experience and do not have as many opportunities to use ASL as deaf native signers, especially at home among hearing parents. Many studies have found that deaf non-native signers perform less well on tests of ASL knowledge and processing compared to deaf native signers (e.g., Boudreau & Mayberry, 2006; Mayberry, Lock, & Kazmi, 2002; Newport, 1990). In addition, hearing native signers do not always exhibit the same level of fluency as deaf native signers, despite being exposed to ASL from birth (e.g., Hauser et al., 2008; Supalla et al., 2014). One explanation for this finding is the “24/7” effect (Paludneviciene et al., 2012). That is, deaf native signers use ASL all day in all situations while hearing native signers use spoken English for much more of their day. For deaf native signers, ASL is the dominant language, but English has been argued to be the dominant language for hearing native signers (Emmorey, Petrich, and Gollan, 2012).

An ANOVA was computed with ASL-CT accuracy as the dependent variable, and significant group differences were found F(2, 57) = 7.06, p < .005; η² = .199. Post-hoc analyses revealed that deaf native signers performed better than deaf non-native signers, t(38) = 3.603, p = .001, and hearing native signers, t(38) = 3.661, p = .001. The deaf non-native signers and the hearing native signers did not differ significantly from each other, t(38) = -2.267, p = .01. The discriminant validity of the ASL-CT, along with its reliability and concurrent validity, add to the overall construct validity of the test, indicating that the ASL-CT is an appropriate measure of ASL comprehension.

Discussion

The results of this study demonstrated that the ASL-CT has good reliability and validity for use with college-age deaf and hearing L1 and L2 signers. The test has an internal reliability alpha of .834, ASL-CT performance positively correlates with ASL-SRT performance (r = .715) and ASL class level (r = .726) establishing concurrent validity, and deaf native signers perform better than deaf non-native signers and hearing native signers (p < .001) establishing discriminant validity. These findings indicate that sign language researchers can use the ASL-CT in research studies involving college students when they need to document participants’ ASL comprehension skills.

In behavioral studies that involve participants responding to ASL stimuli, the ASL-CT could be used to control for variations in the participants’ ASL comprehension skills (as a co-variant or inclusion criteria). In studies testing hypotheses regarding possible effects of ASL comprehension, the ASL-CT could be used as an independent or grouping variable (high scorers vs. low scorers) or as a dependent variable. One benefit of the ASL-CT is that there was not a ceiling effect for the deaf native signers, hence making this test useful to those who wish to assess variation in the ASL comprehensive skills of native signers. In addition,

Table 1. Percent correct and mean and standard deviations of raw scores

<table>
<thead>
<tr>
<th></th>
<th>Percent correct (%)</th>
<th>Meanraw</th>
<th>SDraw</th>
</tr>
</thead>
<tbody>
<tr>
<td>Deaf native signers</td>
<td>86.67</td>
<td>26.00</td>
<td>2.03</td>
</tr>
<tr>
<td>Hearing native signers</td>
<td>72.00</td>
<td>21.60</td>
<td>4.98</td>
</tr>
<tr>
<td>Deaf non-native signers</td>
<td>70.50</td>
<td>21.15</td>
<td>5.67</td>
</tr>
<tr>
<td>Hearing ASL students</td>
<td>63.34</td>
<td>19.00</td>
<td>3.57</td>
</tr>
</tbody>
</table>
studies that include ASL signers will be able to report their sample’s mean ASL comprehension skill in the participant’s section, and this information will be important for others who wish to replicate the study.

We recommend that researchers use Standard scores when reporting ASL-CT results. Standard scores are commonly used when reporting the results of standardized psychoeducational tests such as reading, spoken English, and cognitive tests. Standard scores have a mean of 100 and a standard deviation of 15. Standardized psychoeducational tests have standard scores computed based on a published comparison group. Because English-based psychoeducational tests utilize a comparison group of native English speakers, excluding non-native speakers, we recommend the use of deaf native signers’ mean (26) and standard deviation (2) when computing standard scores. The ASL-CT standard scores can be computed by adding participant’s number of correct responses (raw scores) in the following equation: (participant’s raw score − 26)/2 × 15 + 100. The benefit of using the standard score is that sign language researchers do not need to collect a sample of 20 deaf native signers if they want to compare their participants’ ASL-CT results with this population. However, one limitation of the standard score approach is that this deaf native signers sample is small (n = 20), and therefore researchers should mention this limitation in their studies when they use the ASL-CT.

Caution is also needed when interpreting ASL-CT results with individuals who are not within the college-age range. The use of the ASL-CT for clinical, educational, or organizational purposes should be investigated in future studies (reliability, validity, and comparison groups). Further, the ASL-CT focuses primarily on knowledge of linguistic structures associated with depiction (e.g., classifier constructions, role shift, use of signing space), and therefore the test does not provide a targeted assessment of other linguistic domains (e.g., word order, question formulation, etc.). As with any language assessment, it is always best if multiple measures are used when the situation requires thorough documentation of language abilities.

Conclusion

The psychometric analyses indicate that the ASL-CT has good internal reliability and construct validity as a measure of ASL comprehension when used with college students. The advantage of the ASL-CT is that it is easy to use (no rater training required), takes ~20min to administer, has been validated psychometrically (unlike most lab-internal ASL assessment tests), and provides test results immediately upon completion of the test (unlike ASL tests that require scoring by raters). The availability of the ASL-CT is important because sign language researchers often work with the adult population of deaf and hearing signers (or sign learners) and need a measure to document participants’ ASL comprehension skills. Future studies are needed to investigate the psychometric appropriateness of the test when used with children, adolescents, and older adults. The authors hope that the test will also be useful for clinicians, educators, and organizational leaders who need to document the ASL skills of their clients, students, or employees.

Those who are interested in using the ASL-CT can contact the NSF Science of Learning Center on Visual Language and Visual Learning (VL2). Click on the Contact Us link at the bottom of the VL2 website (http://vl2.gallaudet.edu) and fill out the form (indicating your request to access the ASL-CT in the body of the email).

Conflicts of Interest

No conflicts of interest were reported.

Funding

NSF Science of Learning Center (SBE-0541953 and SBE-1041725); Gallaudet University Priority Grant.

Acknowledgments

We thank Paul Dudis for his contribution to the development of this test and Jonja Smith for creating illustrations for the test.

References


