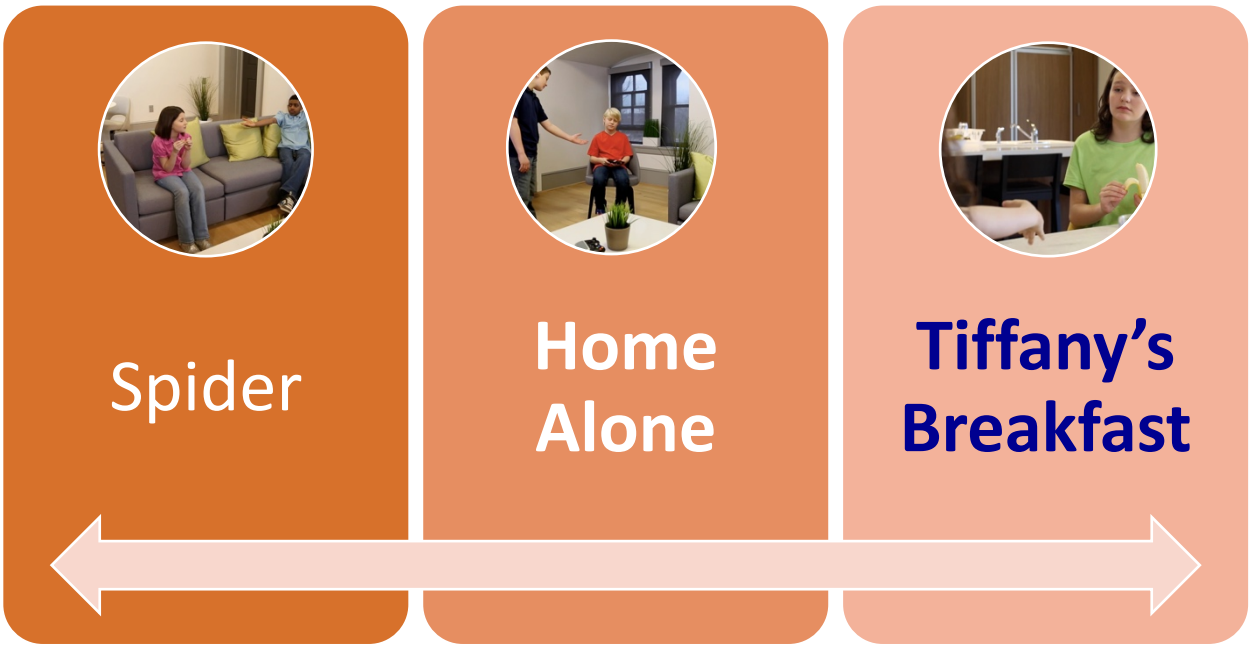
American Sign Language: Expressive Skills Test

Test Manual

Charlotte Enns, Kyra Zimmer, Cheryle Broszeit & Sarah Rabu



Acknowledgements

- We would like to acknowledge the team that developed the *British Sign Language Production Test* on which the *ASL Expressive Skills Test* is based – Rosalind Herman, Nicola Grove, Sallie Holmes, Gary Morgan, Hilary Sutherland, and Bencie Woll.
    - Special thanks to Rosalind Herman for her guidance and consultation throughout the process; and Nicola Grove, the author of the original Spider Story that inspired the other video stories.

- We greatly appreciate the work of the video production team involved in creating the three high quality video stories:
    - Patrick Boudreault – production manager
    - Zilvinis Paludnevicius - videographer
    - Rick Zimmer – production assistant
    - Gallaudet University – filming location
    - Special thanks to the six actors and their parents

- We thank Patrick Boudreault, Gallaudet University, and Sandra Pereira, University of Manitoba, for their careful work on the statistical analysis and psychometrics of the test.

- Our most heartfelt thanks go out to the many children across Canada and the United States who participated in our pilot and normative testing – thanks also to the families and schools for facilitating their involvement.

Transcription Notes:
We follow the convention of using upper case to indicate the English gloss for ASL signs, e.g. EAT, WANT. We also use a hyphen to indicate a single sign glossed with several English words, e.g., PICK-UP-CUP. Other abbreviations and codes used in the Score Forms are outlined in the Scoring Guidelines.

Confidentiality Note:
You, or some of the children you are testing, may recognize the children who are the actors in the test videos (stories). Please note that although we have their permission to appear in the videos, we would appreciate your confidentiality. Please do not acknowledge that you know these children, share their names or any identifying information about them with parents or anyone in your work environment or community.

Contents

List of Tables

1. Introduction

This test is the second in a series of language assessments in American Sign Language (ASL) produced by Northern Signs Research and the research team lead by Charlotte Enns at the University of Manitoba, Canada. The first assessment, the *ASL Receptive Skills Test (ASL-RST)* (Enns, Zimmer, Boudreault, Broszeit, & Rabu, 2013), measured children's understanding of ASL morphology and syntax. The purpose of the *ASL Expressive Skills Test (ASL-EST)* is to complement the *ASL-RST* by assessing children's expressive signed language, specifically their ability to produce a signed narrative with appropriate ASL grammatical features. Each individual child's score can be compared to the normative group and be given a percentile ranking for their age, indicating the child's level of expressive ASL development.

What is the *ASL Expressive Skills Test*?
The *ASL Expressive Skills Test* is a formal, norm-referenced test for children using ASL to measure narrative abilities and parts of ASL grammar. The test is based on an elicited language sample and it is a test you must be certified to use. Each of these points are explained in more detail in the following sections.

What is a "formal" language test?
Experienced teachers and ASL specialists often have a general idea of whether children are developing ASL appropriately, but a formal test provides a standardized way of determining age appropriate ASL abilities. A formal language test involves two key concepts – validity and reliability. When a test is valid, it means that the test actually measures what it says it measures, and therefore, the results are similar to other test results measuring similar abilities. In this case, the results of the *ASL Expressive Skills Test* should be similar to the results of other tests measuring children's ASL grammatical abilities. When a test is reliable, it means that the results are the same if a different person assesses the child, or if the child is tested a week later.

Another important aspect of a formal language test, is that it must be used consistently and in a specific way. The tester cannot add or modify the procedures to fit with the child's needs, as is often done in teaching situations. All children must be given the same instructions and follow the same procedures, so their test results can be compared to the normative group. Formal language tests are developed carefully with clear goals, and they are used with many children so that the test properties are known (know how the test works). In this way, norms for each age group are calculated and new scores can be compared to these established norms.

Who is the test for?
The *ASL Expressive Skills Test* is for any children using ASL. It is primarily for deaf children (aged 4 – 13 years) who are learning ASL as a first language, or any deaf children in this age group who use ASL, even with limited exposure to ASL or exposure to other signing systems. The test could also be used with older deaf children with learning disabilities or

delayed/disordered language, although a comparison to the normative group would not be possible. With these children, the test could be used to determine their visual language developmental needs and strengths. The test might also be used with hearing children who use ASL as their primary way of communicating.

Who can use the test?
Testers must be fluent in ASL, meaning that they are native or near native signers (deaf people) or Level 4 on the ASLPI. Hearing people can take the training workshop and help with scoring, but the test must be administered, i.e., the language sample elicited, by a deaf person. All testers must complete the training workshop and become certified as *ASL Expressive Skills Test* examiners

Who developed the test?
The *ASL Expressive Skills* Test is based on the *British Sign Language Production Test* (Herman, et. al., 2009). Our research team is very grateful for the work done by the BSL team to create the original format (specifically the Spider Story video) and scoring procedures for BSL grammar structures and narrative skills that we were able to build on and adapt to use in ASL.

Test developers include:
Charlotte Enns – professor and researcher in the Faculty of Education at the University of Manitoba (Winnipeg, Manitoba, Canada).

Kyra Zimmer – research assistant at the University of Manitoba; instructor in the Deaf Studies Program and the ASL/English Interpretation Program at Red River College (Winnipeg, Manitoba, Canada).

Cheryle Broszeit - research assistant at the University of Manitoba; instructor in the Deaf Studies Program and the ASL/English Interpretation Program at Red River College (Winnipeg, Manitoba, Canada).

Sarah Rabu - research assistant at the University of Manitoba; ASL specialist at the Manitoba School for the Deaf (Winnipeg, Manitoba, Canada).

Important contributions:
Patrick Boudreault – researcher and associate professor in the Interpretation and Translation Department, Gallaudet University (Washington, DC, United States).

Zilvinis Paludnevicius – independent filmmaker and digital media specialist at Gallaudet University (Washington, DC, United States).

Sandra Pereira – statistical consultant in the Department of Statistics, University of Manitoba (Winnipeg, Manitoba, Canada).

2. Test Development

As previously mentioned, the *ASL Expressive Skills Test* is an adaptation of the *BSL Production Test*. The adaptation process included the following steps:

- Consultation with ASL linguists and Deaf teachers to determine the feasibility of test adaptation (would the test task elicit appropriate ASL structures/narrative?)
- Pilot testing with Deaf adults (fluent in ASL) to determine the target structure expectations for the story re-telling task, and then revising the scoring procedures accordingly
- Creation of three videos (based on original BSL Spider Story) to provide parallel versions of the test for comparison over time
- Conducting pilot testing with children (n=47) using all three stories/versions of the test to ensure that scores could be compared across the different stories
- Revising scoring procedures based on pilot testing; changes, compared to BSL, included similar Grammar items, more emphasis on Role Shift, and combining Narrative Content and Structure
- Collecting normative data on a representative sample (n=215) of deaf children across Canada and the United States

The reason for developing three different stories/versions of the test (parallel elicitation videos) was to prevent children from becoming overly familiar with one story through the process of longitudinal testing, or pre- and post-testing procedures for research purposes. In addition, it was an opportunity to improve the technical quality of the video, remove visual distractions (plain furniture, walls, dishes, etc.), and increase the racial and gender variations of characters. (For more details regarding the creation of the three video stories, please refer to Boudreault, Zimmer, & Enns, 2015).

Normative Sample

All the testing was completed in schools for the deaf in Canada and the United States, by members of the research team. All children were deaf and had been exposed to ASL before the age of 3 years. In order to ensure that children had average nonverbal intelligence, the *Test of Nonverbal Intelligence* (TONI-4) (Brown, Sherbenou, & Johnsen, 2010) was administered to most children, unless they were too young (less than 5 years of age) or the school had conducted similar testing within six months. Children with scores less than one standard deviation below the mean, including those with diagnosed disabilities (e.g., autism), were excluded.

We collected data from a total of 215 deaf children between the ages of 3.5 years to 13.9 years of age (please see Table 2 for details). The sample was not exactly gender balanced, with 125 girls and 90 boys, but we found no statistical evidence for differences in test performance based on gender (ANOVA Sig=0.933, greater than 0.05). Also, most of the children (n=144) had at least one deaf parent, with the remaining 71 children having non-deaf parents. Again, no statistical evidence was found for differences in test performance

between children with deaf parents or non-deaf parents (Asymp. Sig=0.659, greater than 0.05). It is important to keep in mind that we had a selective sample, and only children with early exposure to ASL (before the age of 3 years) were included. We recognize that controlling for early exposure to ASL and nonverbal IQ resulted in a sample that may not be representative of the general population of deaf children, as considerable variability exists in terms of deaf children's cognitive and language abilities. However, for this test we wanted to have a normative sample that represents what achievements are possible when children have early and rich full access to language. In this way the children that you are testing will be compared to a normative group that is acquiring ASL age-appropriately, therefore, if you find that children are delayed in comparison to the normative sample you can provide the necessary support to develop their skills to their full potential.

Table 2: Description of Normative Sample

Age (years)	Gender		Parents		Ethnicity			
	Male	Female	Deaf	Non-Deaf	White	Hispanic	Black	Other
3.5-4.9 (n=25)	8	15	14	11	14	11	0	0
5.0-5.9 (n=23)	10	13	15	8	15	7	1	0
6.0-6.9 (n=23)	9	14	14	9	15	3	5	0
7.0-8.9 (n=50)	26	24	36	14	31	10	3	6
9.0+ (n=94)	38	56	65	29	63	13	14	4
Totals	**125**	**90**	**144**	**71**	**138**	**44**	**23**	**10**

Test Validity

In order to check if the test was measuring what it was designed to measure (children's ASL abilities), scores (based on the categories of average, above average, below average) from the ASL Expressive Skills Test were compared with the same children's scores on the ASL Receptive Skills Test using a Pearson's correlation. A highly significant correlation (0.91, $p<0.01$) was found, suggesting good concurrent test validity.

Test Reliability

The reliability of the ASL Expressive Skills Test was investigated using inter-scorer comparisons and test-retest analyses with the three versions of the test.

Inter-scorer reliability was assessed by having 10% (30 videos) of the data independently scored by two different trained testers and comparing the results. Statistical analysis using Pearson's correlation resulted in a highly significant correlation of 0.87 ($p<0.01$), indicating inter-scorer reliability was very good.

The test-retest reliability was based on the 47 children participating in the pilot testing who each completed two versions of the test (retold two different stories) within the same testing session. Parametric statistics (ANOVA) were used to compare the children's scores between all combinations of the three stories, and no statistical evidence was found for any differences (Sig=0.288, greater than 0.05). A Test of Homogeneity of Variances was used to validate the assumption of the homogeneity for ANOVA and had similar results (Sig=0.234, greater than 0.05). These analyses indicate strong test-retest reliability within participants and across all three versions of the test.

Test Scores According to Age

An investigation of the relationship between children's ages and total test scores revealed a weak linear relationship (R-Square Linear=0.619). These findings were supported with both Pearson and Spearman's rho coefficients and the results were 0.787 and 0.767 respectively, which were slightly better and greater than 0.70. These statistical analyses point out the considerable variability that occurred in children's scores at each age level. For this reason, it was decided that using mean scores, and the range of scores around the mean, at each age level would provide a more effective comparison.

Group means were calculated for five age ranges (see Tables 3 – 7, pp. 25 - 26), to include the Total score, and the sub-sections of Grammar, Role Shift, Questions, and Narrative. The younger children are grouped by one-year age ranges (4-, 5-, and 6-year olds), whereas the older children are grouped in two-year (7-8 year olds) or multi-year (9 years and older) ranges. The reason for this is because rapid language development occurs in these younger years, but development levels out as children get older. Also, grouping the ages together allows for larger numbers of participants in each group which provides a more reliable basis for the means and percentiles. As noted previously, there is considerable variability within the groups, particularly for the younger ages, due to the small group sizes. Even one child responding differently than the rest of the group can have a big influence on the group as a whole. For this reason, mean scores for the younger age groups should be interpreted with caution.

3. Test Administration

The *ASL Expressive Skills Test* uses a stimulus (video) to encourage children to re-tell a story, which is recorded for later analysis and scoring. More specifically, the child watches a short three minutes video with no signing or talking. The characters in the video tell the story through gestures and facial expressions. When the child finishes watching the video, they retell the story to a deaf person who is not familiar with the video. It is important for the child to believe that the listener has not seen the video so that they tell a more detailed story. Following the child's retelling of the story, which is videotaped for later analysis, the child must also answer three comprehension questions. The child's responses to these questions are also video recorded. There are three different video stimuli (stories) so that the test can be repeated without the child becoming too familiar with the story. The three videos can be used interchangeably because they are parallel in terms of the story structure, number of events, two characters, and the types of ASL grammatical features that are elicited.

The purpose of the test is to assess narrative abilities, which includes a variety of cognitive skills, like memory, sequencing, and cause-effect, as well as language skills, like ASL grammar and story structure. The test does not assess specific vocabulary, which means that it easily adjusts to regional variations of signs. For example, some children referred to the "crepe" as a "tortilla", which is fine and has no impact on their score as long as they use the correct grammatical structures when referring to the item.

Before the Test

- Make sure you have parental consent to video record the child

  Your school may already have a protocol or procedure in place for gaining parental consent, but it is important when you are video recording children that you provide parents with an explanation in writing (letter or email) and obtain signed consent.

- Book an appropriate room for the assessment

  The room should be a private space (without other people or activities) and free from distractions. Try to make sure activities outside the room are limited and avoid interruptions by putting a notice on the door. There should be enough space in the room to arrange chairs, table and equipment in the right places for comfortable viewing (eye level) and video recording. We suggest that the children not sit on chairs with wheels as this tends to encourage distracting movements during signing. Lighting is important for video recording so there should be enough light, but no backlight or reflections on the signer. Always check that computer and video equipment is available and working before you begin.

- Inform and prepare the child

  Provide the child with a clear explanation of the following information:

  - You are interested in finding out how good they are at telling a story (try to avoid saying it is a "test")
  - They will watch a short story on video
  - They need to remember it so they can tell it to you after, and that *you will not watch it with them* (or tell it to another person that will not watch it, if that is the case)
  - They will be recorded on video when they tell the story
  - They will need to remove gum/food from their mouth, and put down any toys or objects so they can sign clearly on the video recording

Suggestions for Eliciting the Best Narrative Sample

- Ensure you are (or you recruit) an appropriate listener

  Who the child is telling the story to may influence the quality of the narrative sample. For example, some children, especially younger ones, will be more comfortable and share more details of the story with a familiar person rather than someone they do not know. Although children may feel comfortable telling the story to another child (peer), this may result in more casual communication and they may not use aspects of formal grammar and story structure. The most important consideration is that the child tells the story to a deaf person to clearly set the expectation to interact in ASL. So, the ideal listener is a familiar, deaf adult with experience administering the test.

- Make the child feel comfortable

  Begin with some general conversation to engage the child and establish a connection. Do not tell them it is a "test". Be warm, encouraging, patient and show interest while the child is telling the story.

- Encourage the child to be specific, but keep prompts general (don't give too much information)

  Some suggestions for prompts include: Start at the beginning; Tell me exactly what happened; What else do you remember?

  Do not help the child with their answers or suggest there are right or wrong answers.

- Remember that being a tester is different from being teacher

  It is important to emphasize that when you are in the tester role, you must resist your natural teacher urge to support, scaffold and help the child to do better. The test must be given in the same standardized way to each child, so that the results can be compared to the normative group.

- Accept the child's best re-telling of the story

  Sometimes children re-tell the story in response to the questions. If this is just a repetition of what they have already told you, then it can be ignored. However, sometimes, especially with younger or shy children, the questions can prompt a better re-telling, so this is the narrative sample you should use for scoring.

After the Story

The child needs to answer three questions after they re-tell the story. Explain that they will watch the questions on the video, and then tell you the answers. Show them each question only once, and wait patiently for their response. Video record the child's answers, and in fact you should keep the video recording going throughout. DO NOT present the questions live, except with children younger than five years of age. If you do present the questions live, try to repeat them exactly how they are signed on the video. If children are reluctant to answer the questions, you can provide some general encouragement, e.g., MORE? ANYTHING-ELSE?

Useful Tips

- Be familiar with the assessment materials – read the manual and watch the videos
- Try a practice run – give the test to a colleague so you are familiar with the instructions, materials and equipment
- Have someone who knows the child (a teacher or classroom assistant) help with the assessment (this is especially important with younger children)
- Give the assessment with two adults if possible or needed:
  - First person stays with the child watching the video
  - Second person serves as listener (child tells the story to them)

4. Analysis of Children's Narratives

The *ASL Expressive Skills Test* involves analyzing and scoring the child's video recorded story re-telling and responses to the three comprehension questions. The analysis is done by the trained tester after the session with the child. The tester will need to view the video repeatedly to accurately analyze and score all the test components, therefore, the analysis process can take 30 – 90 minutes (depending on the testers skills and experience) for each child. Some children's samples will be easier to analyze than others, and the length of time needed is also influenced by the tester's familiarity with the child's signing.

The Score Form is in an electronic Excel workbook format (included in the test USB) with a separate form for each of the three stories (Spider, Tiffany's Breakfast, and Home Alone). Use the appropriate form for the story that was administered with the child. Although there are slight differences in the analysis/scoring for each story, the total scores are the same so children's scores can be compared across different stories/versions of the test. Each form includes four tabs:

a) Cover Page
b) Grammar
c) Role Shift
d) Narrative

Blank copies of each tab (for the Spider Story) are provided in Appendix A.

Testers should begin by watching the entire story (child's sample), to become familiar with the child's signing; then view short clips, pause to check off items on the score form, and review again to confirm the analysis. Testers are not required to transcribe the child's signing. When numbers (0, 1, or 2) are entered into the "Child's Score" columns in the Excel Score Form, they will automatically be tallied for each section. It is recommended that testers double check the total score to make sure the automatic scoring is accurate.

Guidelines for analyzing the child's sample and completing all sections of the Score Forms (based on the test training) are provided:

a) Spider Story – Appendix B
b) Tiffany's Breakfast Story – Appendix C
c) Home Alone Story – Appendix D

Appendix A also provides a sample of a completed Score Form (for Home Alone Story).

Analysis and scoring guidelines must be followed consistently for valid and reliable test results. Any unusual language patterns or additional information that does not fit with the scoring guidelines should be noted on the form/cover page, and outlined in the test report.

5. Scoring the Test

Once you finish analyzing the child's story, you need to find out if their narrative skills are age-appropriate. The Score Form will automatically add up the scores you enter in each section and provide you with the total score (raw scores). The raw score is the total number of points that a child achieved for a section of the test:

Section of the Test	**Maximum Raw Score**
ASL Grammar	30
Role Shift	6
Questions	6
Narrative Content/Structure	18
Total	**60**

What Does the Raw Score Tell You?
The raw score simply tells you how many points the child achieved out of the possible maximum. The raw score is useful for comparing a child's first score with their later performance on the test. Also, the raw score is useful when testing children for whom there are no norms available, e.g., older children; children with learning difficulties/disorders. The raw score ***does not*** tell you how the child's score compares to other children of the same age – for this we need to compare the raw score to percentiles.

What is a Percentile?
Percentiles show you the range of scores achieved by the sample of children on whom the test was developed (the normative sample). Percentiles allow you to compare a child's score with other children their age.

Comparing Raw Scores to Percentiles
Due to the variation in deaf children's language acquisition there are a range of scores that are "normal" for each age. For this reason, the standard deviations of the mean are large, and the medians (50^{th} percentiles) are better estimates of age appropriate levels than the means. The *ASL-EST* groups the percentiles into five age groups: a) 4 year olds, b) 5 year olds, c) 6 year olds, d) 7 & 8 year olds, and e) 9 year olds and above. The first three groups (4, 5, and 6 year olds) cover only one year because there is significant growth and development in language and narratives abilities at these ages. The remaining two groups cover several years as development is leveling off (less variable) at these older ages. Each age group has raw scores that fit into the five percentile categories: 10%, 25%, 50%, 75%, and 90%.

The child's level of performance for their age can be determined by using the charts (**Tables 3 – 7**) to compare raw scores to the percentiles with the following steps:

1. Make sure you have completed the analysis of the child's videotaped story (Grammar, Role Shift, Questions, Narrative), entered all the points accurately, and checked that the automatically tallied raw scores on the Score Form are accurate.

2. Find the correct Table (on pages 25 and 26) based on the child's age:
   Table 3 – 4 year olds (3.5 – 4.95 years)
   Table 4 – 5 year olds (5.0 – 5.95 years)
   Table 5 – 6 year olds (6.0 – 6.95 years)
   Table 6 – 7 & 8 year olds (7.0 – 8.95 years)
   Table 7 – 9 year olds and above (9.0+ years)

3. Look under each of the sections (Total, Grammar, Role Shift, Questions, Narrative) listed along the top row of the table, and find where the child's raw score falls in the percentile categories (listed in the bottom row). If the child's raw score falls between two percentile categories, always use the higher percentile. In the example below, the 8-year old child's raw score of 20 in the Grammar section falls under the 75%; however, her raw score of 12 in the Narrative section falls under the 50%.
   Example:
   Table 6: Comparing Raw Scores and Percentiles (7 & 8 year olds)

Age Range 7.0 - 8.95 years (50 children)	**Total (60)**	**Grammar (30)**	**Role Shift (6)**	**Questions (6)**	**Narrative (18)**
Mean	34.69	15.66	2.58	3.92	12.52
Standard Error of Mean	3.09	0.91	0.29	0.21	0.50
Percentile					
10	15	6	0	2	6
25	27	11	0	3	11
50	37	15	3	4	13*
75	43	23*	4	5	15
90	49	24	6	6	16

Completing Cover Page

There are several sections on the Cover Page tab of the Score Form that must be completed by the examiner. These include:

a) Background Information – child's name, age, date of birth, date of test, examiner's name
b) Scoring (Scores for ASL Grammar; Scores of Narrative Skills) – these should tally automatically, but double check to make sure all numbers have transferred to the cover page accurately
c) Results in Percentiles – after comparing the raw scores to percentiles with Tables 3 – 7, enter the percentile results for all sections in the graph
d) Comments on testing – make notes on any observations related to the child's state, behavior, or performance (sick; attention; behaviours; distractions/interruptions; repetition of video, etc.)
e) Child's strengths and development needs – provide a summary of findings; areas for future teaching based on percentile rankings in each section and overall observations.

Interpreting Test Scores

In general, scores that fall under each of the percentiles can be interpreted in the following way in comparison to other children in the age group:

10th percentile – significantly delayed
25th percentile – moderately delayed
50th percentile – average
75th percentile – above average
90th percentile – significantly above average

It is important to consider the percentile rankings for all sections of the test, as children may demonstrate delays in ASL grammar, but average abilities in the area of narrative content and structure. Even if the child scores at the 50th percentile or above in all sections of the test, some areas (e.g., Role Shift or Grammar) may be significantly above average indicating a relative strength in that area. The comparison of percentile rankings can provide insight into future programming and instruction.

There are numerous reasons why a child may receive a low score that does not accurately represent their language and narrative abilities. The following factors regarding the child's performance on the test must be taken into consideration to ensure that the assessment of the child's abilities is appropriate:

- Not comfortable with tester or test environment
- Distracted or reluctant to participate
- Not understanding test task
- Sick, tired, not feeling well
- Aware that tester is familiar with the story
- Difficulties with memory, learning, vision (factors other than language).

Once these factors have been minimized or eliminated, then the low score may be attributed to a language delay or disorder, or limited or late exposure to ASL.

Also, a child may receive a high score on the test which over-estimates their language and narrative abilities in other settings. This may be due to the child being familiar with the video story (tested too frequently or informed by another student), or because they have good memory and sequencing skills. The child's age may also influence their performance on the test, as older children are expected to receive higher scores (the skills are age-related). In most cases, a high score indicates early, or native, exposure to ASL and strong development of key ASL grammar structures (spatial verbs, agreement verbs, and role shift).