Reliability and Validity of Four Arabic Language Tests: A comparison of performance of Qatari School-aged children with and without language impairment

Saleh Shaalan[1]

Abstract

This study describes the reliability and validity of four language tests: The Sentence Comprehension Test (SCT), the Expressive Language Test (ELT), the Sentence Repetition Test (SRT), and the Arabic Picture Vocabulary Test (APVT). These tests were administered to two groups of Qatari Arabic-speaking children: A typically developing group (n=81 to 88) aged 4;6-9;4 years old and a chronologically age-matched group with specific language impairment (SLI) (n=26). The results of the four language tests showed high levels of reliability and validity and support the usefulness of these tools to diagnose children with SLI, whose performance on the tests was mostly consistent with findings in other languages.

Keywords: *Qatari Arabic, child language development, Specific Language Impairment*

Theoretical background

Specific language impairment (SLI) is one of the most common developmental language disorders with an estimate prevalence rate of 5.7% (Tomblin et al., 1997). It is defined by the presence of significant receptive and/or expressive language impairments in the absence of cognitive, sensorimotor, and social-emotional deficits (Bishop, 1997; Leonard, 1998). Children with SLI are considered very heterogeneous as they come with varying individual linguistic abilities (Bishop, 1997; 2004; Leonard, 1998). However, they typically show significant deficits in some important aspects of language development. For example, children with SLI showed more difficulties with sentence comprehension skills when compared to typically developing (TD) peers (Bishop, 1979; Deevy& Leonard, 2004; Montgomery, 1995; Montgomery & Evans, 2009; Norbury, Bishop, & Briscoe, 2002; van der

---

[1]The New England Center for Children-Abu Dhabi
sshaalan@neccuae.org

Lely, 1998). Moreover, Children with SLI displayed significant deficits in many aspects of their expressive language skills, such as tense and agreement (Rice, Wexler & Hershberger, 1998), noun and verb morphology (Rice 2003; Rice et al, 1998), regular plurals (Oetting& Rice, 1993). Furthermore, these children's performance significantly lagged behind their peers' performance on sentence repetition skills (Archibald &Joanisse, 2009; Conti-Ramsden, Botting, &Faragher, 2001). Word learning and vocabulary development in children with SLI are another area where these children showed below average performance when compared to their TD peers (Rice, Cleave, &Oetting, 2010).

Researchers and clinicians working with children employ a variety of language tests to help them identify this heterogeneous population. Some of the most commonly used comprehensive language tests that are used with English- speaking children are the Clinical Evaluation of Language Fundamentals-4 (CELF-4, Semel, Wiig, & Secord, 2003) and the Preschool Language Scale- 5 (PLS-5; Zimmerman, Steiner & Pond, 2011) that assess different aspects of expressive and receptive language skills, such as sentence comprehension, sentence repetition, phonological awareness, and comprehension and production of different syntactic and morpho-syntactic structures. Moreover, most clinicians and researchers use vocabulary tests, such as the Peabody Picture Vocabulary test-4 (PPVT-4, Dunn & Dunn, 2007) as part of their test battery.

Challenges of working with children with SLI in Arabic

Despite being one of the most common childhood language disorders, SLI has not been studied thoroughly in Qatari Arabic or other varieties of Arabic, and clinicians and researchers working with this population have many challenges. The first challenge faced by any researcher studying this population is the lack of standardised tests, criterion-referenced measures or any formal tools for diagnosing children with SLI. There are no published tests for QA or any of the other varieties of Gulf-Arabic (GA) and there are no systematic and comprehensive investigations of language acquisition in this population. As far as the author knows, the only published study on language acquisition was an investigation of the development of tense and agreement of three toddlers in Kuwaiti GA (Aljenaie, 2001). So far, there is only one study of SLI in any variety of Gulf Arabic, namely, the Shaalan (2010) study of SLI in Qatari Arabic. Abdalla (2002) looked at morphosyntactic deficits in pre-schoolers with SLI acquiring Hijazi-Arabic, a variety of Arabic that is close but different from Qatari Arabic. Abdalla (2002) based her diagnosis of children with SLI on MLU as well as adaptations of English tests and clinical judgments of speech-language therapists.

Sentence Comprehension in Arabic

With regard to availability of language tests, there has been only one systematic attempt to create a comprehension test for typically developing children in Saudi Arabia (Al-Akeel,

1998), though the test has not been published yet. The test was developed to assess language comprehension skills of Saudi children aged 3;0-6;0 years old and was meant to be used with children using different regional dialects of Saudi Arabia. The test was designed to assess children's understanding of 24 morphosyntactic structures that were selected from three sources: spontaneous language samples of typically developing children interacting with their fathers; morphosyntactic structures that the author added himself based on his linguistic knowledge of Arabic and some morphosyntactic structures were modified from existing English language tests.

This study attempts at designing and assessing the reliability and validity of a test battery that comprises the following: The Sentence Comprehension Test (SCT), the Expressive Language Test (ELT), the Sentence Repetition Test (SRT), and the Arabic Picture Vocabulary Test. (APVT). There were other supplementary tests and screeners that were used as part of a large project aiming at identifying children with SLI, but will not be described here. These supplementary tools included two nonverbal IQ tests as well as two screening tests for oral-motor functioning and articulation skills (see Shaalan, 2010). The distribution of test scores will be examined as well as the overall performance of the TD and the SLI groups on these tests. The main utility of these tests is to see if children with SLI will consistently lag behind their TD peers on these four tests, a finding that has been reported in other languages.

Method

Participants

The number of typically developing (TD) children who participated in the four tests was between 81 to 88 with their ages ranging between 4;6 and 9;4. The Sentence Comprehension Test (SCT) was administered to 88 TD children, while in the Expressive Language Test (ELT) and the Sentence Repetition Test (SRT), the number of TD children was 86in both tests. The Arabic Picture Vocabulary Test (APVT) was administered to 81 TD children. The number of children in the SLI group was 26 across the four tests.

Selection criteria for children with SLI

The criteria adapted for establishing the cut-off points for typical vs. atypical language performance (in children with SLI) were largely based on Tomblin et al., (1997). These include having within normal range performance on one of two nonverbal IQ tests and the absence of any motor, neurological, or socio-emotional deficits. Children were included in the group of children with SLI if they had a score of – 2.0 standard deviations (SD) or more on one out of the four language tests, or -1.5 SD or more on two tests. These were stricter than those used by Tomblin et al. (1997) due to the smaller number of participants in this study. Due to lack of normative data in typical and atypical language acquisition in children acquiring QatariArabic, and due to lack of tests that could be used with typically

and atypically developing children, the project had to start by collecting normative data from typically developing children before conducting the experiments with children with SLI and the age and language controls.

The targeted age groups for children with SLI in the larger project were ages 6;0 – 8;11 years old (Shaalan, 2010). Therefore, the project started by conducting the full battery of tests with children in these age brackets to identify 'norms' for these four language tests. Following testing of at least 20 children in these age groups, means, standard deviations and z-scores were separately calculated for each group and for each language test. Cut-off scores of -1.5 and -.2.0 standard deviations were established for each group. This was followed by adding children below the age 6;0 as these were needed to act as language controls for other experiments in the same project (see Shaalan, 2010). Based on the criteria developed for the first three age brackets, children with SLI, who ranged in age between 6;0 and 8:11 years old at the time of initial testing were diagnosed based on comparing their performance with the normative sample for their age brackets. More children were added to all group bands, depending on availability and time constraints. Due to difficulties with scheduling and access to schools, the total number of typically developing children in each age bracket ranged between 19 and 24, falling below the initial target of 30 TD children in each age group.

Children with SLI met the selection criteria mentioned earlier as they all scored -1.5 SD or more on two out of the four language tests or -2.0 SD on one test. They all had within normal scores on one of the two nonverbal IQ tests used throughout the project, namely the Test of Nonverbal Intelligence (TONI-3) (Brown, Sherbenou, & Johnsen, 1997)for children aged 6 years and above or the Block Design and Picture Completion subtests of the Wechsler Preschool and Primary Scale of Intelligence (WPPSI-III) (Wechsler, 2002). All children with SLI passed a hearing screening at 20 dB for frequencies between 500-2000 Hz that was conducted by the author, who is a licensed speech-language pathologist. Moreover, they had uneventful developmental history with no sensory, motor, or social-emotional problems. Screening tools were used to rule out oral-motor dysfunction, childhood apraxia of speech, or severe articulation disorders. No other informal measures of spontaneous speech were conducted due to time constraints. However, all children were engaged in a five-minute conversation before administering the tests to put them at ease and get an initial idea of their language levels. Two of the children with SLI had been previously diagnosed with developmental language disorders when they were aged 4;0, the rest of them were not diagnosed with any language disorders. However, most of them came with concerns about their academic performance. These concerns were expressed by class teachers and social workers.

Most children were recruited from two kindergartens and four primary schools in Doha, the capital of Qatar, and some were recruited through personal acquaintances. Most participants came from what can be described as middle-class families and Qatari Arabic was the language spoken at home. However, most of these children had some exposure to

English, which is taught at kindergarten level in Qatar and is widely spoken in the community due to the large number of expatriates in Qatar.

Table 1 summarizes the characteristics of the two groups of children who participated in the Sentence Comprehension Test (SCT). See tables 1 and 2 in Appendix 1 for characteristics of children who completed the Expressive Language Test (ELT), the Sentence Repetition Test (SRT), and the Arabic Picture Vocabulary Test (APVT).

Table1
Summary of the characteristics of participants in the Sentence Comprehension Test (SCT)

Age Groups	TD	SLI
Age Band 1: 4;6 - 5;11 years		
Number of participants (Male:Female)	24 (13:11)	5 (2:3)
Mean age in months (years)	64.0 (5;3)	62.6 (5;2)
Range in months (years)	54-71 (4;6-5;11)	58-70 (4;10-5;10)
Age Band 2: 6;0 - 6;11 years		
Number of participants	23 (15:8)	8 (7:1)
Mean age in months (years)	77.6 (6;6)	78.9 (6;7)
Range in months (years)	72-83 (6;0-6;11)	73-83 (6;1-6;11)
Age Band 3: 7;0 - 7;11 years		
Number of participants	22 (14:8)	5 (4:1)
Mean age in months (years)	90.6 (7;6)	88.8 (7;5)
Range in months (years)	84-99 (7;0-7;11)	85-94 (7;1-7;10)
Age Band 4: 8;0 - 9;4 years		
Number of participants	19 (13:6)	8 (5:3)
Mean age in months (years)	103.1 (8;7)	103.0 (8;7)
Range in months (years)	96-112 (8;0-9;4)	99-107 (8;3-8;11)
Total Number of participants	88 (54:33)	26 (18:8)
Mean age in months (years)	82.6 (6;10)	85.1 (7;1)
Range in months (years)	54-112 (4;6-9;4)	58-107 (4;10-8;11)

Materials and procedure

General procedures

The time it took to administer all the tests together ranged between 45-60 minutes, depending on child's age, participation, and whether he or she asked for a break. Children aged between 4;0 and 5;0 were given frequent breaks and most of the time testing was done in two 30-minute sessions. Most children enjoyed testing and were praised for their performance but no comments were made about their accuracy. At the end of each session, each child received some stickers as a reward. Testing usually started with a short chat with the child to establish rapport. This was followed by less verbally demanding tasks, such as

the nonverbal IQ test, which was followed by the Sentence Comprehension Test (SCT). Then the Expressive Language Test (ELT) and the Sentence Repetition Test (SRT) were conducted. The Arabic Picture Vocabulary Test (APVT) followed these two expressive tasks and the session ended after running the articulation and oral-motor screenings. Children's responses were scored on individual record forms. All children were required to attempt all test items and no basal or ceiling items were set, due to lack of normative data for typical and atypical language development in QA speaking children, except for the APVT where some ceiling was adopted due to large number of items in that test.

The Sentence Comprehension test (SCT)

The Sentence Comprehension test consisted of 40 items that examine the comprehension of different syntactic, morphological, and morphosyntactic structures in Qatari Arabic. Table 1 in Appendix 2 lists all the different linguistic structures used in the SCT. In the SCT, each student was required to listen to a sentence produced by the examiner and point to the right answer from an array of four different pictures on each sheet. An artist from the Gulf region drew some of the pictures, while others were taken from some English tests, such as CELF-3 (Semel, Wiig, & Secord, 1996) and were carefully examined to ensure they were culturally appropriate for this population.

During testing, the children were presented with two trial items and were given instructions in QA equivalent to the following in English: "We are going to look at this book and I will show you some pictures. I want you to point to the picture I am talking about. For example: "show me 'the girl is sleeping'". Instructions were repeated if necessary, and there were two trial items to familiarise children with the procedures. All children understood instructions and responded to all items. Self-corrections were accepted and the second answer was considered the final one. Children were given 0 for incorrect answers and 1 for correct answers. The score was written on the score sheet for the SCT. The highest possible raw score was 40/40. Children were praised for their compliance and not for the accuracy of their answers.

The Expressive Language Test (ELT). The Expressive Language Test (ELT) measured theproduction of various morphosyntactic structures commonly used by Qatari Arabic speaking children. It consisted of 68 items and children were required to answer all items. The distributions of all ELT items are listed in Table 2 in Appendix 2. These linguistic structures were chosen based on structures seen in language samples of TD children, the investigator's native knowledge of Gulf Arabic, his experience as a speech language pathologist, and previous research on Kuwaiti Arabic (e.g., Aljenaie, 2001) or varieties that are close to Gulf Arabic, such as those spoken in Saudi Arabia (e.g., Abdalla, 2002; Al-Akeel, 1998). Some English language tests, such as the Clinical Evaluation of Language Fundamentals-CELF3 (Semel, Wiig, & Secord, 1996) and Preschool Language Scale-PLS4 (Zimmerman, Steiner, & Pond, 1992) were consulted and some structures that appear in Arabic were used while ensuring their ecological validity (e.g., superlatives). Other clinicians working with Gulf Arabic children in Qatar were consulted about appropriate

structures to be used with this population and their input was incorporated in the choice of items used in the test.

The testing started with two practice items and the instructions were as follows (in Arabic): "Together, we will look at some pictures. I will show you some pictures and I will say something and I want you to complete what I say: For example, (showing the child a picture of one strawberry): "Here we have a strawberry (farawla), and here (pointing to the picture of three strawberries in the second page) we have three… (Child is expected to say 'farawlaat' (strawberries))". Children would get a score of 1 for a correct answer or 0 for an incorrect one. In this test, single repetition was allowed and a specific prompting procedure was permitted. When a child did not reply, her/his score was considered as 'no response' (NR) and she/he would get a score of 0.

The Sentence Repetition Test (SRT)

The Sentence Repetition (SR) test consisted of 41 sentences that were arranged in a least-to-most difficult order. Sentences increased in length and grammatical complexity as the child progressed through the test. Table 3 in Appendix 2 shows the distribution of the SRT items. The instruction was the equivalent of the following in Arabic: "You will hear some sentences and I will say each one once only. I want you to repeat them exactly the way I say them". This was followed with two practice items. Most children did not have problems understanding the instructions. In a few cases, a third example was needed. The scoring method used was adapted from the one used in the Clinical Evaluation of Language Fundamental-3 (CELF-3). Therefore, children would get 3 points if they repeated the whole sentence with no errors, 2 points when there was one error, 1 point when there were two errors and 0 if they produced three or more errors or when they provided no response. Error was defined as any change in the sentence that is not of articulatory nature. No repetition of any sentence was allowed. There was no basal or ceiling and children were required to attempt repeating all sentences.

The Arabic Picture Vocabulary Test (APVT)

The Arabic Picture Vocabulary Test (APVT) consisted of 132 items arranged in terms of difficulty into 10 groups with 12 items per group (see content validity for more information on how items were arranged). This is the method used in the development of the British Picture Vocabulary Test (BPVT) (Dunn, Dunn, Whetton, & Burley, 1997). A booklet was made that has 134 pages (2 pages for practice items and 132 for test items). Each page depicted 4 pictures that were mostly taken from non-copyrighted material (e.g., free clip arts). All answers were transferred to a record form.

Children were given the following instructions (in Arabic). "Together we will see a picture book. I will name one of the pictures and I want you to point to the picture I am talking about. Let's try a couple of pages…". This was followed by two practice items ('shoe' and 'fish'). Children were presented with four pictures and they were required to

point to the correct response. None of the children had any difficulties with the instructions. Because of the large number of stimuli, a ceiling criterion was employed in order to reduce fatigue, especially among younger children. The ceiling criterion used was a minimum number of eight errors in one group before stopping the test. Children were encouraged to continue if they seemed to like the test even when ceiling was established. Fourteen TD children continued until the last item despite reaching a ceiling at a previous item. Only 3 children with SLI reached the last item, with two of them reaching a ceiling at a previous one. All responses were recorded on a score sheet and children received (1) for a fully correct answer and (0) for a fully incorrect one. The total raw score was computed by subtracting the number of errors the child made from the last ceiling item. For example, a child who stopped at item number 60 and had total errors of 14 would have a raw score of 46.

Results and discussion

Across the four language tests, the SCT, the ELT, the SRT, and the APVT, children with SLI consistently lagged behind their typically developing (TD) peers. In the following, we report the results in terms of distribution of test scores, one way ANOVAs, and reliability and validity of the four tests. A comparison of children's performance across tests is also examined.

Distribution of test scores:

One of the important psychometric properties of a test is the distribution of test scores. The following figures show the distribution of the scores of all typically developing children on the SCT, ELT, and SRT. They depict a broadly normal distribution of these scores across the whole sample of TD children.
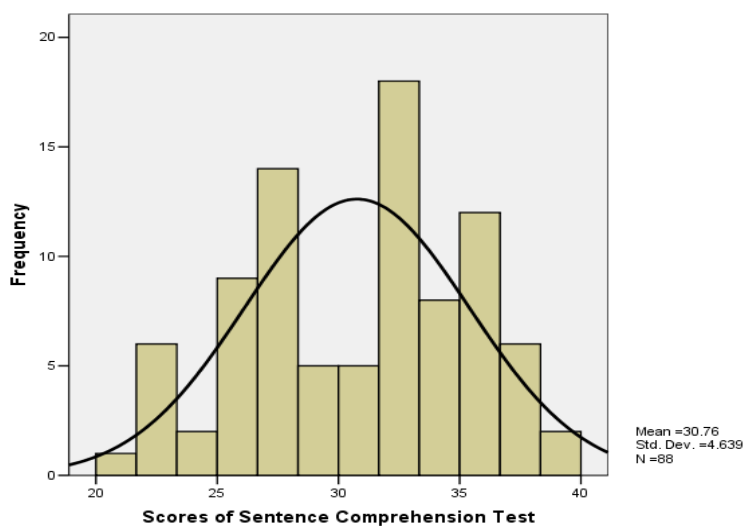
Figure 1. Distribution of typically developing children on scores of the Sentence Comprehension Test (SCT), n=88
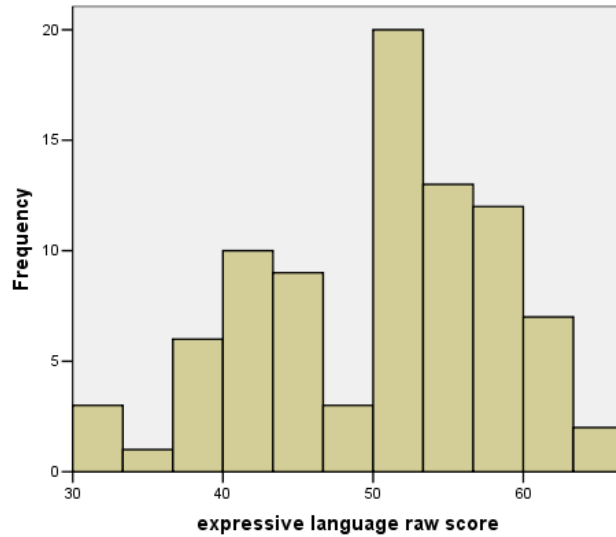

Figure 2. Distribution of typically developing children on scores of the Expressive Language Test (ELT), *n*=86
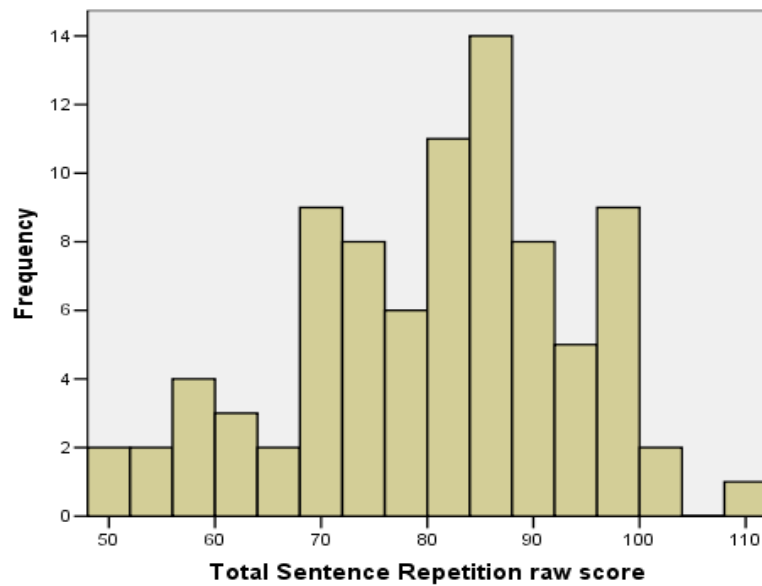

Figure 3. Distribution of typically developing children on scores of the Sentence Repetition (SR), *n*=86

Figure 4 shows the distribution of the scores of all typically developing children on the Arabic Picture Vocabulary Test. It depicts a broadly normal distribution of the test scores.
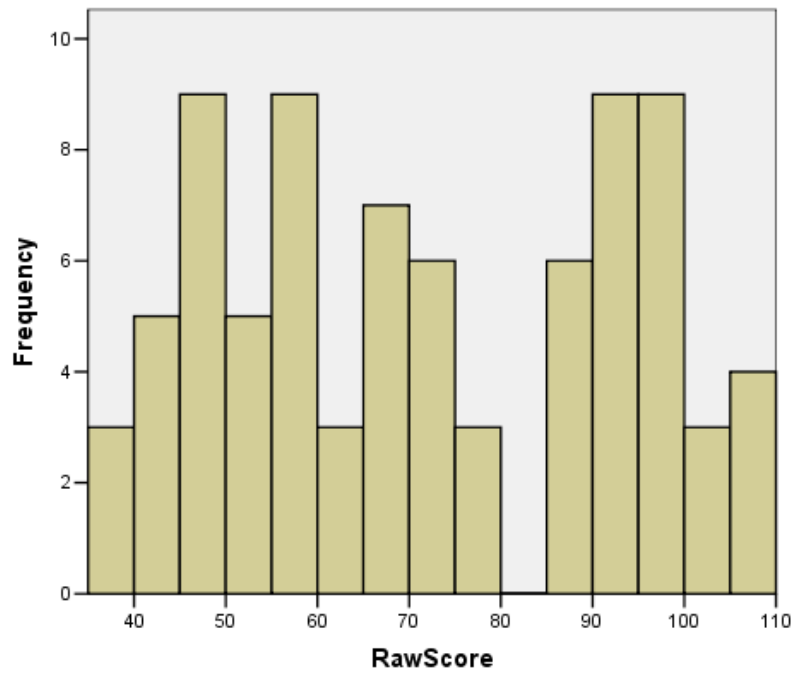
Figure 4.Distribution of typically developing children on scores of the Arabic Picture Vocabulary Test (APVT), n=81

The Shapiro-Wilk test of normality was conducted and it was significant for two groups (5 and 6 year olds). However, it was not significant for the 7 and 8-year-old groups. Figures 5 and 6 depict the distribution of the scores of the 5 and 6-year-old groups on the APVT. They both showed some mild negative skewness.

Figure 5.Distribution of the scores of the TD 5 year olds on the APVT

Figure 6.*Distribution of the scores of the TD 6-year-old children on APVT*

Group Comparisons, t-tests and One Way ANOVAs

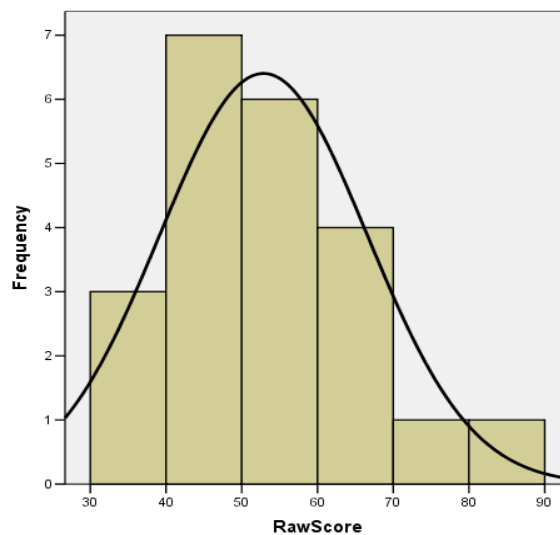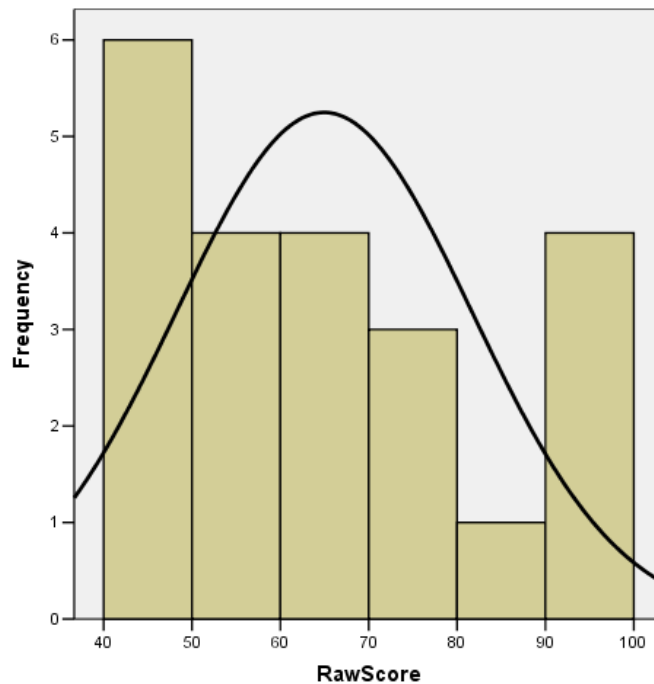**The Sentence Comprehension Test (SCT)**

Table 2 and figure 7 show the performance of the TD and SLI groups on the SCT. A t-test was performed to compare the overall means of the two groups. It showed that the TD group was significantly better than the SLI group on the SCT $t(112)=4.6$, $p<.001$. Children with SLI obtained a mean score equivalent approximately to their TD peers who were 2 years younger, as depicted in Figure 7. One way ANOVA of the scores of the four TD age bands (Age Band 1: 4;6-5;11 years; Age Band 2: 6;0-6;11; Age Band 3: 7;0-7;11 ; Age Band 4: 8;0-9;4.) showed there was a significant difference among their performances, $F(3,84)=31.8$, $p<.001$. Post-hoc tests with Bonferroni correction showed that all age groups of TD children were significantly different from each other except the 7 and 8-year-old groups. The 5 year old group was significantly different from the 6 year old group $t(45)=-2.89$, $p=.02$, the 6 year old group had significantly lower scores when compared to the 7 year old group $t(43)= -4.0$, $p<.001$. However, there was no significant difference between the 7 and 8 year old groups $t(39)= -1.73$, $p=.54$. This shows that the test was sensitive to age factors in typically developing children, especially for younger children from 4:6 to 7 years old. These differences ceased to be significant in children aged between 7 and 8 years old, because the test became less challenging at this age.

Table 2
Means (and standard deviations) for performance on the Sentence Comprehension Test (SCT)

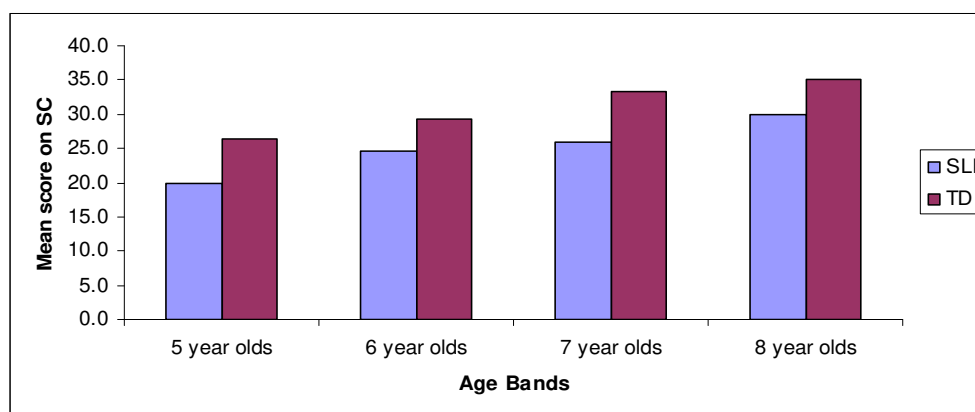| Age Groups	Typically Developing Children	SLI
4;6-5;11 years		
Number of participants (Male:Female)	24 (13:11)	5 (2:3)
Mean Raw Score of SCT (SD)	26.4 (3.65)	19.80 (4.65)
Range of SC scores	20-33	15-26
6;0-6;11 years		
Number of participants	23 (15:8)	8 (7:1)
Mean Raw Score of SCT (SD)	29.3 (3.38)	24.63(4.56)
Range of SC scores	24-37	18-31
7;0-7;11 years		
Number of participants	22 (14:8)	5 (4:1)
Mean Raw Score of SCT (SD)	33.3 (3.41)	26.00(4.52)
Range of SCT scores	27-38	20-31
8;0-9;4 years		
Number of participants	19 (13:6)	8 (5:3)
Mean Score of SCT (SD)	35.1 (4.05)	30.00 (5.19)
Range of SCT scores	32-39	21-35
Total Number of children	88 (54:33)	26 (18:8)
Mean Raw Score of SCT (SD)	30.8 (4.64)	25.62 (5.78)
Range of SCT scores	20-39	15-35

Figure 7.Comparison between children with SLI and their typically developing (TD) peers on the Sentence Comprehension Test (SCT)

ANOVA of the SCT scores of the four groups of children with SLI showed a significant effect of age group $F(3,22)=4.8$, $p=.01$. Multiple comparisons with Bonferroni correction showed that only the 5 and 8 year old groups were significantly different from each other, $t(11)= - 10.2$, $p<.01$. No comparisons among the other age groups were significantly different from each other. These null findings may be explained by a combination of small sample sizes and lack of developmental effects, whereby severity level might have influenced performance more than chronological age.

The Expressive Language Test (ELT)

Table 3 summarises the results of all children on the Expressive Language test. It shows that children with SLI were consistently lagging behind their TD peers, and that the performance of the TD groups improved consistently with age.

Table3 and Figure 8 show that the oldest group of children with SLI (8 years old) had a score that was near the score achieved by the youngest TD group (4-5 years old), indicating that production of various syntactic and morphological structures could constitute a major area of deficits in QA children with SLI.

Table *3*
Results of all participants on the Expressive Language Test (ELT)

Age Groups	Typically Developing Children	SLI
Age Band 1: 4;6 - 5;11 years		
Number of participants (Male: Female)	24 (13:11)	5 (2:3)
Mean Raw Score of ELT (SD)	42.6 (7.8)	23.6 (3.0)
Range of EL scores	30-55	20-27
Age Band 2: 6;0 - 6;11 years		
Number of participants	23 (15:8)	8 (7:1)
Mean Raw Score of ELT (SD)	50.0 (6.6)	27.4 (13.6)
Range of EL scores	37-61	10-43
Age Band 3: 7;0 - 7;11 years		
Number of participants	21 (13:8)	5 (4:1)
Mean Raw Score of ELT (SD)	52.8 (5.4)	35.2 (4.3)
Range of EL scores	44-62	29-39
Age Band 4: 8;0 - 9;4 years		
Number of participants	18 (12:6)	8 (5:3)
Mean Raw Score of ELT SD)	57.2 (4.0)	45.9 (9.1)
Range of EL scores	51-66	32-59
Total Number of children	86 (53:33)	26 (18:8)

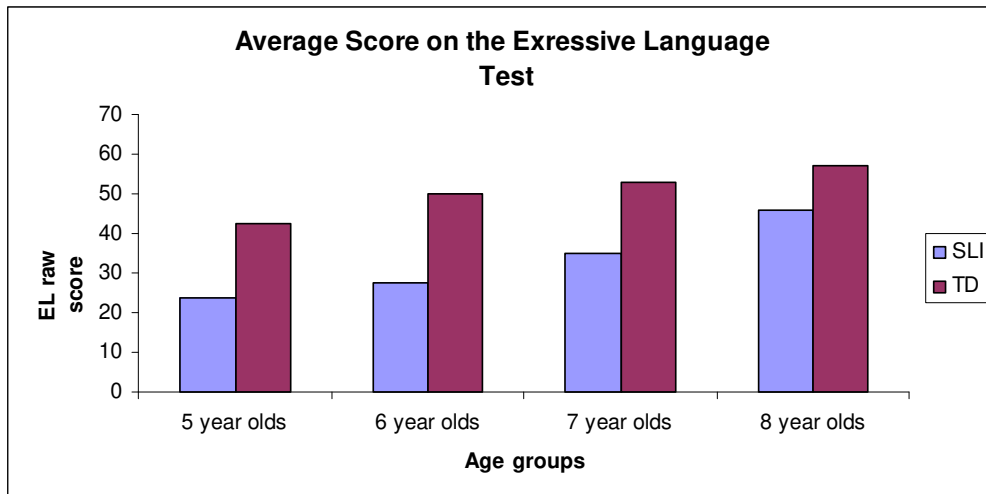|		
Mean Raw Score of ELT (SD)	50.1 (8.1)	33.8 (12.7)
Range of EL scores	30-66	10-59

Figure 8.Comparison of the overall Expressive Language raw scores by children with SLI and typically developing (TD) children across different age groups

T-test showed that the group of TD children was significantly better than the SLI group on the EL test t (31.4)=6.6, p<.001. One-way ANOVA of the scores of the four TD age groups, showed a significant group effect, F (3.82)=20.4, p<.001, indicating the presence of a developmental factor. Multiple comparisons with Bonferroni effect showed that the 5-year-old group scored significantly lower than all three older groups. The 5 year old group was significantly different from the 6 year old group t (45) =-7.41, p=.001. However, the 6 year olds' scores were not significantly different from the 7 year old group, but they had significantly lower scores when they were compared with the 8 year old group t(39)= -7.12, p=.003. There was no significant difference between the 7 and 8-year-old TD groups.

The Sentence Repetition Test (SRT)

Results of all children on the Sentence Repetition Test are summarised in Table 4. It shows that children with SLI performed consistently worse than their TD peers across different age groups. The mean score of all children with SLI was less than the mean score of the youngest typically developing group. Table 4 shows that children with SLI had a performance comparable to TD peers who were two years younger than they were. Figure 9 compares the performance of all groups of children on the Sentence Repetition test.

T-test showed that the TD group was significantly better than the SLI group on the Sentence Repetition test t(33.3) =5.5, p<.001. One way ANOVA of the scores of the four age groups of TD children showed a significant group effect, F (3.82) =13.9, p<.001. Multiple

comparisons with Bonferroni correction showed that the 5 year old group scored significantly lower the 6 year old group t (45)= -5.56, p=.02 and the other older groups. The 6-year-old

Table 4
Results of all participants on the Sentence Repetition (SR) test

Age Groups	Typically Developing Children	SLI
Age Band 1: 4;6-5;11 years		
Number of participants (Male: Female)	24 (13:11)	5 (2:3)
Mean Raw Score of SRT (SD)	69.8 (13.9)	40.8 (7.8)
Range of SR scores	49-94	29-49
Age Band 2: 6;0-6;11 years		
Number of participants	23 (15:8)	8 (7:1)
Mean Raw Score of SRT (SD)	79.3 (10.6)	52.5 (17.6)
Range of SR scores	62-101	28-76
Age Band 3: 7;0-7;11 years		
Number of participants	21 (13:8)	5 (4:1)
Mean Raw Score of SRT (SD)	84.3 (8.4)	64.6 (9.6)
Range of SR scores	68-99	50-76
Age Band 4: 8;0-9;4 years		
Number of participants	18 (12:6)	8 (5:3)
Mean Raw Score of SRT (SD)	90.4 (9.9)	75.0 (12.4)
Range of SR scores	71-111	32-59
Total Number of children	86 (53:33)	26 (18:8)
Mean Raw Score of SRT (SD)	80.5 (13.0)	59.5 (17.9)
Range of SR scores	49-111	28-9

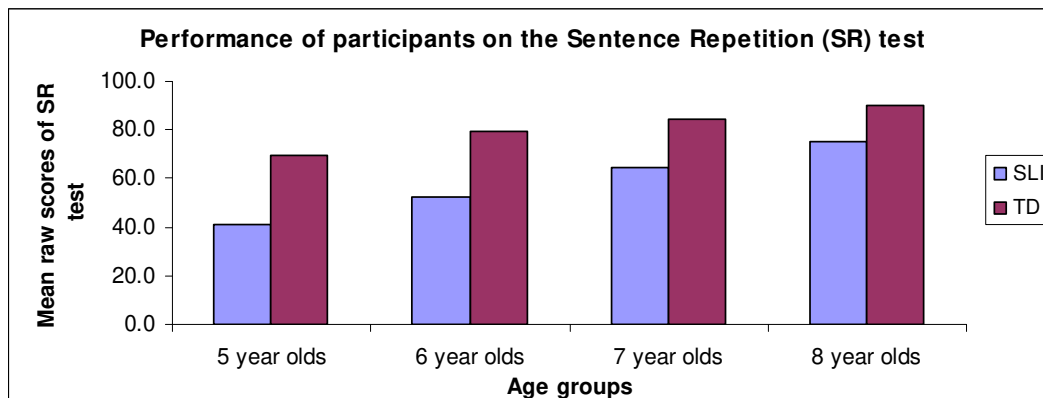Performance of participants on the Sentence Repetition (SR) test

Figure 9.Comparison of the overall Sentence Repetition Test (SRT) raw scores of children with SLI and typically developing (TD) children across different age groups

group had asignificantly lower score than the 8 year old group t (39)=3.39, p=.01. However, there were no significant differences between the 6 and 7-year-old groups on one hand and between the 7 and 8-year-old groups on the other hand. Overall, these results are consistent with developmental trends, where groups of TD older children perform better than younger groups.

The Arabic Picture Vocabulary Test (APVT)

Table 5 summarises the results of all children on the Arabic Picture Vocabulary Test (APVT). It shows that typically developing children scored significantly higher than children with SLI and across different age groups, indicating that children with SLI have limited receptive vocabulary compared to their TD peers. Table 5 and Figure 10 show that children with SLI have generally scores similar to those of TD peers who were 2 years younger than they were, a pattern that has been noticed in the SRT and the SCT.

Table 5
Summary of results of all participants on the Arabic Picture Vocabulary Test (APVT)

Age Groups	Typically Developing Children	SLI
Age Band 1: 4;6-5;11 years		
Number of participants (Male: Female)	22 (11:11)	5 (2:3)
Mean Raw Score of APVT (SD)	52.9 (13.7)	31.2 (7.8)
Range of APVT scores	37-89	21-42
Age Band 2: 6;0-6;11 years		
Number of participants	22 (14:8)	8 (7:1)
Mean Raw Score of APVT (SD)	65.0 (16.7)	36.0 (15.2)
Range of APVT scores	43-92	13-54
Age Band 3: 7;0-7;11 years		
Number of participants	19 (11:8)	5 (4:1)
Mean Raw Score of APVT (SD)	75.4 (17.7)	50.6 (14.1)
Range of APVT scores	48-101	37-72
Age Band 4: 8;0-9;4 years		
Number of participants	18 (12:6)	8 (5:3)
Mean Raw Score of APVT (SD)	98.0 (7.5)	65.3 (13.5)
Range of APVT scores	86-110	43-83
Total Number of children	81 (47:34)	26 (18:8)
Mean Raw Score of APVT (SD)	71.5 (21.8)	46.9 (18.8)

Range of APVT scores	37-110	13-83

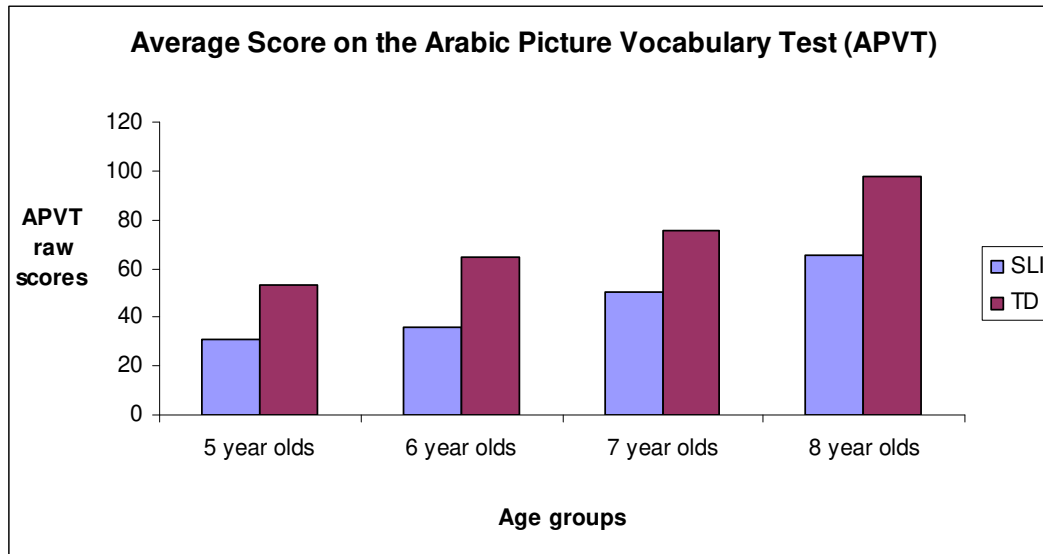**Average Score on the Arabic Picture Vocabulary Test (APVT)**

Figure 10.Comparison of the scores of the Arabic Picture Vocabulary Test (APVT) by children with SLI and typically developing (TD) children across different age groups

The TD group scored significantly better than the SLI group on the APVT, $t(105)=5.2$, $p<.001$. One-way ANOVA of the scores of the four age groups of TD children showed a significant group effect, $F (3.77)=33.6$, $p<.001$. Multiple comparisons with Bonferroni correction showed that the 5-year-old group scored significantly lower than all three older groups. The difference between 5 and 6 year olds was significantly in favour of the latter $t(42)=12.10$, $p=.045$; the 7 year olds had higher scores than those obtained by the 5 year olds $t(39)=22.56$, $p>.001$, and so did the 8 year olds $t(38)=45.13$, $p<.001$. The difference between 6 and 7 year olds was not significant, but the difference between the 6 year olds and the 8 year old group was significant $(38)=33.04$, $p=.001$. Finally the 8 year olds were significantly better than the 7 year olds, $t(35)=22.48$, $p<.001$. Overall, there was a clear developmental pattern with each age group obtaining higher scores on the APVT than the age group preceding it, except for the difference between 6 and 7 year olds, which was not significant.

Reliability

Reliability refers to the ability of a test to yield consistent measures when used under identical conditions. It is usually divided into three types of measures: split-half analysis, Cronbach's alpha, and test-retest reliability.

Split-half Analysis

A split-half analysis was conducted to examine the correlation between the scores obtained from odd-numbered items with scores from even-numbered items for all four tests. This

was more appropriate than measuring the correlation between the first and second half of the test as items were arranged in terms of difficulty. Correlation coefficients of .70-.80 are considered acceptable (Field, 2005).

Cronbach's alpha

While the Split-half analysis groups the items into one way only (e.g., odd vs. even), Cronbach's alpha splits the data into two in every possible way and then computes the correlation between these items. Therefore, it is considered a better measurement of internal reliability. Cronbach's alpha is considered acceptable if it falls between .70 and .80 (Field, 2005).

The results of split half-analysis with Spearman-Brown coefficient and the Cronbach's alpha of the four tests are shown in Table 6 and they show mostly high levels of internal reliability.

Table 6.
Results of Split-half analysis and Cronbach's alpha

Test	Split-half analysis (Spearman-Brown coefficient)	*Cronbach's alpha.*
The Sentence Comprehension Test (SCT)	.89	.79
The Expressive Language Test (ELT)	.94	.93
The Sentence Repetition Test (SRT)	.96	.89
The Arabic Picture Vocabulary Test (APVT)	.71	.82

Tests re-test reliability

Test-retest reliability is used to measure the stability of the test when used with the same individual over time. To examine test-retest reliability, six children were retested one week after they took the tests for the first time. This group of children consisted of five male students and one female student, aged 75 to 107 months (6;3-8;11 years old). Five of them were typically developing and one was diagnosed with SLI. Results of test-retest reliability are shown in Table 7 and they show that all these tests were stable over time, with a Pearson correlation coefficients ranging between .95 to.97.

Table 7.
Results of test-retest measures of internal reliability

Test	Test-retest correlation coefficients
The Sentence Comprehension Test (SCT)	.95

The Expressive Language Test (ELT)	.95
The Sentence Repetition Test (SRT)	.97
The Arabic Picture Vocabulary Test (APVT)	.97

Validity

Validity refers to the extent to which a test measures what it intends to measure. Two types of validity are usually assessed: content validity and concurrent validity.

Content Validity

Content validity refers to what extent the test items are relevant and representative of the targeted constructs being assessed (Haynes, Richard, &Kubany, 1995). To ensure that the SCT possesses an appropriate level of content validity, all structures used in the test were chosen based on the same criteria adopted by Al-Akeel (1998) in his test of comprehension of morphosyntactic structures in Saudi Arabic. Therefore, the structures were selected based on these criteria: they appeared in the language samples of TD QA speaking children; they werechosen by the investigator based on his native knowledge of the language and his clinical experience as a speech-language therapist. Thirdly, some structures were carefully chosen from English language tests (such as CELF-3 (Semel, Wiig, & Secord, 1996) or PLS-3 (Zimmerman et al., 1992), provided they also appear in Gulf Arabic and are culturally appropriate.

For the ELT, most of the linguistic structures were included based on language samples taken from more than 35 Qatari Arabic speaking children, whose age ranged between 2;11 and 4;11 years old (Khater& Shaalan, 2007; Shaalan &Khater, 2006). Some linguistic structures were based on the investigator's knowledge of Gulf Arabic, as a native speaker, and on his experience as a speech-language therapist working with Qatari Arabic speaking children with and without language impairment. Finally, a group of Qatari Arabic speaking clinicians and linguists were asked to examine the structures in the ELT when preparing the tests. Their overall responses were positive and they gave some suggestions about elicitations that were incorporated in the test.

Content validityAPVT

While the SCT, ELT, and SRT were based on structures from language samples, clinician's feedback, investigator knowledge of his native language, and some adaptations of English material, the Arabic Picture Vocabulary test (APVT) had to be developed de novo. The process started by asking 24 adult speakers of Qatari Gulf- Arabic to rate 600 words in terms of familiarity. Each word received a rating from 1-5 (1= rarely heard or used, 5=very familiar and used very frequently). These words belonged to 20 different semantic categories (e.g., verbs, animals, occupations, adjectives (…) etc.) following the same practice used in the development of the British Picture Vocabulary Test (BPVT) (Dunn et al, 1997). Out of these 600 words, 132 words were chosen and organised into 11 groups of 12 words

per group ranked according to their difficulty, which was determined based on the familiarity rating of each item. The criteria for choosing these words were similar to those used in the British Picture Vocabulary Test (BPVT). Hence, all the words included were functional, easy to depict pictorially and common in everyday life, except for the advanced vocabulary where some were taken from Classic and Modern Standard Arabic.

Concurrent validity

Concurrent validity measures the correlation of the novel test with other tests taken by the same group of children at the same time (Anastasi & Urbina, 1997). Ideally, these tests should tap into the same skill, e.g., various vocabulary tests are expected to correlate with each other. However, due to lack of any standardised tests in Qatari Arabic, these four tests had to be compared with each other. Results of the Pearson Correlation revealed that the four tests were significantly correlated as shown in Table 8.

Table8
Correlations of the four language tests.

	SCT (*n*=114)	ELT (*n*=112)	SRT (*n*=112)	APVT (*n*=107)
SCT	-	.54**	.43**	.63**
ELT	.54**	-	.69**	.50**
SRT	.43**	.69**	-	.34**
APVT	.63**	.50**	.34**	-

**Correlation is significant at the 0.01. Level (2-tailed).
Note. SCT=The Sentence Comprehension Test, ELT=Expressive language test, SRT=Sentence Repetition test, APVT=the Arabic Picture Vocabulary Test.

Comparing performance across language tests

Table 9 and figure 11 summarizes the performance of TD and children with SLI on the four language tests. They show that children with SLI significantly lagged behind their TD peers on the four language tests and they had more difficulties with expressive language skills, as indicated with their mean score on the ELT. Their best performance, on the other hand, was on the Sentence Comprehension Test.

Table 9
Descriptive statistics of the performance of typically developing children and those with SLI on various language tests

Tests	Typically Developing Children (*n*=86)	SLI (*n*=26)
1- The Sentence Comprehension test		
Mean Standard Score (SD)	99.4 (14.5)	70.5 (26.2)
Range	74-132	0-106

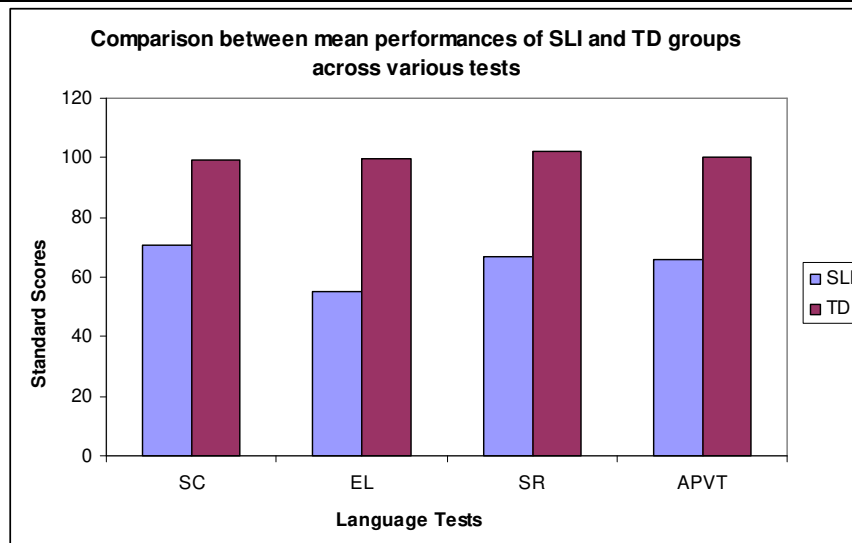| 2- The Expressive Language test		
Mean Standard Score (SD)	99.7 (14.3)	55.3 (24.8)
Range	76-133	6-107
3- The Sentence Repetition test		
Mean Standard Score (SD)	102 (16.8)	66.9 (20.3)
Range	69-177	30-111
4- The Arabic Picture Vocabulary Test		
Mean Standard Score (SD)	100.2 (12)	65.7 (25.6)
Range	77-128	33-102

*Figure 11.*Means of Standard scores of TD and SLI children on four language tests.

Note. SC=the Sentence Comprehension Test; EL=the Expressive Language Test; SR= the Sentence Repetition Test; APVT=the Arabic Picture Vocabulary Test.

Discussion

The four tests described in this study, the Sentence Comprehension Test (SCT), the Expressive Language Test (ELT), the Sentence Repetition Test (SRT), and the Arabic Picture Vocabulary Test (APVT) showed good psychometric properties. The distributions of the test scores were mostly normal and showed a developmental pattern with older children performing better than younger ones and TD children performing significantly better than children with SLI. All four tests showed high levels of reliability and validity, as shown by their good levels of internal consistency and content and concurrent validity. These results validate the use of these tests to identify children with specific language impairment.

Results showed that the pattern of performance of children with SLI was mostly similar to that observed in other languages. When the four language tests were compared,

Qatari Arabic speaking children with SLI had significant weakness in their expressive language abilities, as shown by their very low scores on the Expressive Language Test (ELT) in comparison to the other three tests. This has been consistently reported in studies of SLI in other languages (Bedore& Leonard, 1998; Bortolini, Casalini, & Leonard, 1997; Leonard, 1998; 2009; Moyle, Karasinski, Ellis Wismer, & Gorman, 2011), where the expressive language abilities of children with SLI typically lagged behind other language abilities, especially their receptive language skills. Children with SLI were found to have difficulties producing morphoysntactic structures (e.g., tense, subject verb agreement, relative clauses) that varied from language to another (see Leonard, 2009 for an overview). No detailed comparisons were possible between specific linguistic structures in QA and results reported in other languages due to the general nature of the tests.

Children with SLI's performance on the receptive vocabulary test, which was comparable to their performance on the Sentence Repetition Test might seem incongruent with many studies that found better performance on vocabulary than on sentence repetition, but there is evidence that some children with SLI were reported to have comparable results on these two tests in English (e.g., Leonard, 1998). Another possible explanation for the relatively poor performance on the receptive vocabulary test might be attributed to the root-and-pattern nature of Arabic. In Arabic, a semantic root undergoes many morphological and phonological transformations to generate various vocabulary items derived from that root. For example, the root'd-r-s' (study) is used to derive finite and infinite verbs 'daras' (he studied) and 'tadris' (she studies), the nouns 'madrasa' (school), 'dira:sa' (study), mudarris' (teacher)…etc. Therefore, children with SLI, who have been shown to have syntactic, morphological, morphosyntactic, and phonological deficits, might be less competent at using these roots to derive more vocabulary items out of them. However, it is difficult to have conclusive remarks based on these tests only, as they need further revising and should be used in bigger projects with larger number of participants.

Due to the nature of these tests, which were general tests that were designed to assess a wide range of skills, it was difficult to compare the TD and SLI groups on items due to limited exemplars from each linguistic structure. However, some items of interest for further follow up can be recommended. For example, on the SCT, it was observed that children with SLI had more difficulties with certain syntactic structures, such as relative clauses, negation, and passive.Among the patterns seen in the ELT was that children with SLI seemed to have more difficulties with different types of clitic pronouns and irregular plurals when compared to TD peers. On the other hand, children with SLI performed relatively well on verb inflections and regular feminine plurals on the ELT. However, the limited number of items and their unsystematic distribution does not allow any conclusions to be drawn and these preliminary notes warrant further investigations.

Conclusion

Qatari Arabic speaking children with SLI were identified based on a battery of language tests that were developed specifically for this purpose, due to lack formal and informal language assessment tools or language development norms. These tests were the Sentence Comprehension test, the Expressive Language test, the Sentence Repetition test and the Arabic Picture Vocabulary Test. The results showed good levels of reliability and validity; therefore, supporting the use of these tests in the identification of Qatari Arabic-speaking children with SLI. Moreover, the pattern of performance of typically and atypically developing Gulf Arabic speaking children on various language tests is consistent with findings reported in other languages, thus further validating the results of these tests and their ability to diagnose children with SLI. However, while children with SLI acquiring European languages tend to have relative strength in receptive vocabulary, Arabic speaking children with SLI showed poor performance on the receptive vocabulary test. This is probably due to the root-and-pattern nature of the language. Overall, Gulf-Arabic speaking children with SLI showed variable abilities on the four language tests used in the project, hence confirming the heterogeneous characteristics of SLI seen in other languages. However, since this is the first attempt at developing such tests, all these assessment tools warrant further revisions and should be administered with a larger number of participants.

References

Abdalla, F. (2002). Specific language impairment in Arabic-speaking children: deficits in morphosyntax. PhD dissertation, McGill University, Canada.

Al-Akeel, A. (1998). *The acquisition of Arabic language comprehension by Saudi children.* Unpublished PhD. thesis, University of Newcastle upon Tyne, UK.

Aljenaie, K. (2001). *The emergence of tense and agreement in Kuwaiti Arabic children.* Unpublished PhD dissertation, University of Reading.

Anastasi, A., & Urbina, S. (1997). Psychological testing (7th ed.). Upper Saddle River, NJ: Prentice Hall.

Archibald, L.M.D, &Joanisse, M.F. (2009). On the sensitivity and specificity of nonword repletion and sentence recall to language and memory impairments in children. *Journal of Speech, Language, and Hearing Research*, 52, 899-914.

Bedore, L., & Leonard, L. B. (1998). Specific language impairment and grammatical morphology: A discriminant function analysis. *Journal of Speech, Language, and Hearing Research*, 41, 1185-1192.

Bishop, D. V. M. (1979). Comprehension in developmental language disorders. *Developmental Medicine and Child Neurology, 21*, 225-238.

Bishop, D. V. M. (1997). *Uncommon understanding: Comprehension in specific language impairment*. Hove: Psychology Press.

Bishop, D. V. M. (2004). Specific language impairment: diagnostic dilemmas. In L. Verhoeven& H. van Balkom (Eds.), *Classification of developmental language disorders* (pp. 309-326). Mahweh, New Jersey: Lawrence Erlbaum.

Bortolini, U., Casalini, C., & Leonard, L. B. (1997). Grammatical deficits in Italian-speaking children with specific language impairment. *Journal of Speech, Language, and Hearing Research*, *40*, 809-820.

Brown, L., Sherbenou, R. J., & Johnsen, S. K. (1997). TONI-3: Test of nonverbal intelligence. *Wood Dale, IL: Stoelting Co.*

Conti-Ramsden, G., Botting, N., &Faragher, B. (2001). Psycholinguistic markers for specific language impairment. *The Journal of Child Psychology and Psychiatry*, *42*, 741-748.

Deevy, P., & Leonard, L. B. (2004). The comprehension of Wh-Questions in children with specific language impairment. *Journal of Speech Language and Hearing Research, 47*, 802-815.

Dunn, L. M. & Dunn, D. M. (2007). *The Peabody Picture Vocabulary Test: Fourth Edition*. Circle Pines, MN: AGS.

Dunn, L. M., Dunn, L. M., Whetton, C., & Burley, J. (1997). *The British Picture Vocabulary Scale: Second Edition*. Windsor: NFER-Nelson publishing Company Ltd.

Field, A.P. (2005) *Discovering statistics using SPSS (2ⁿᵈ Edition).* London: Sage.

Haynes, S. N., Richard, D. C. S., &Kubany, E. S. (1995). Content validity in psychological assessment: A functional approach to concepts and methods. *Psychological Assessment, 7*, 238-247.

Kauffman, N. R. (1995). The Kaufman Speech Praxis Test for Children: Detroit: Wayne State University Press.

Khater, M., & Shaalan, S. (2007). *Reporting norms for mean length of utterance (MLU) in words and morphemes for Qatari Speaking Children*. Paper presented at Linguistics in the Gulf Conference, University of Qatar, Doha, Qatar.

Leonard, L. B. (1998). *Children with specific language impairment*. Cambridge, Mass.: MIT Press.

Leonard, L. B. (2009). Cross-linguistic studies of child language disorders. In Schwartz, R.G. (Ed.), *Handbook of child language disorders* (pp.308-324). New York, NY: Psychology Press.

Montgomery, J. W. (1995). Sentence comprehension in children with specific language impairment: The role of phonological working memory. *Journal of Speech and Hearing Research*. *38* 187–199.

Montgomery, J.W., Evans, J.L. (2009). Complex sentence comprehension andworking memory in children with specific language impairment. *Journal of Speech, Language, and Hearing Research, 52*, 269-288.

Moyle, M.J, Karasinski, C., Ellis Wismer, S., & Gorman, B.K. (2011). Grammatical morphology in school-age children with and without language impairment: a discriminant function analysis. *Language, Speech, and Hearing Services in Schools, 42*, 550-560.

Norbury, C. F., Bishop, D. V., Briscoe, J. (2002). Does impaired grammatical comprehension provide evidence for an innate grammar module? *AppliedPsycholinguistics*. 23 247–268.

Oetting, N., & Rice, M.L. (1993). Plural acquisition in children with specific language impairment. *Journal of Speech, Language, and Hearing Research, 36*, 1236–11248.

Rice, M. L. (2003). A unified model of specific and general language delay: Grammatical tense as a clinical marker of unexpected variation. In Y. Levy & J. Schaeffer (Eds.), *Language competence across populations: Toward a definition of Specific Language Impairment* (pp. 63-95). Mahwah, NJ: Erlbaum.

Rice, M. L., Cleave, P. L., &Oetting, J. B. (2000). The use of syntactic cues in lexical acquisition by children with SLI. Specific Language Impairment. *Journal of Speech, Language, and Hearing Research, 43*, 582-594.

Rice, M. L., Wexler, K., & Hershberger, S. (1998). Tense over time: The longitudinal course of tense acquisition in children with specific language impairment. *Journal of Speech, Language, and Hearing Research, 41*, 1412–1431.

Semel, E., Wiig, E., & Secord, W. (1996). *Clinical evaluation of language fundamentals-(3rd ed.)*. San Antonio, TX: Psychological Corp.

Semel, E., Wiig, E., & Secord, W. (2003). *Clinical evaluation of language fundamentals-(4rd ed.)*. San Antonio, TX: Psychological Corp.

Shaalan, S. (2010). Investigating grammatical complexity in Gulf-Arabic speaking children with specific language impairment. PhD dissertation, University College London, UK.

Shaalan, S., &Khater, M. (2006). *A comparison of two measures of assessing spontaneous language samples in Arabic speaking children*. Poster presented at Child Language Seminar 2006, University of Newcastle upon-Tyne, U.K, 19-21 July 2006.

Tomblin, J. B., Records, N. L., Buckwalter, P., Zhang, X., Smith, E., & O'Brien, M. (1997). Prevalence of specific language impairment in kindergarten children. *Journal of Speech and Hearing Research, 40*, 1245-1260.

van der Lely, H. K. J. (1998). SLI in children: movement, economy and deficits in the computational syntactic system. *Language Acquisition, 72*, 161-192.

Wechsler, D. (2002). *WPPSI-III: Wechsler preschool and primary scale of intelligence, (3rd ed.)*: Psychological Corp.

Zimmerman, I. L., Steiner, V. G., & Pond, R. E. (1992). *Preschool language scale (PLS-3)*. San Antonio, TX: Psychological Corp.

Zimmerman, I. L., Steiner, V. G., & Pond, R. E. (2011). *Preschool language scales (5th ed.)*. Bloomington, MN: Pearson.

Appendix 1

Table 2
Characteristics of participants in the Arabic Picture Vocabulary Test (APVT)

Age Groups	Typically Developing Children	SLI
Age Band 1: 4;6 - 5;11 years		
Number of participants	22 (11:11)	5 (2:3)
Mean age in months (years)	64.1 (5;4)	62.6 (5;2)
Range in months (years)	54-71 (4;6-5;11)	58-70 (4;10-5;10
Age Band 2: 6;0 - 6;11 years		
Number of participants	22 (14:8)	8 (7:1)
Mean age in months (years)	77.7 (6;5)	78.9 (6;7)
Range in months (years)	72-83 (6;0-6;11)	73-83 (6;1-6;11)
Age Band 3: 7;0 - 7;11 years		
Number of participants	19 (10:9)	5 (4:1)
Mean age in months (years)	90.1 (7;6)	88.8 (7;5)
Range	84-95 (7;0-7;11)	85-94 (7;1-7;10)
Age Band 4: 8;0 - 9;4 years		
Number of participants	18 (12:6)	8 (5:3)
Mean age in months (years)	103.3 (8;7)	103.0 (8;7)
Range	96-112 (8;0-9;4)	99-107 (8;3-8;11)
Total No. of participants	81 (47:34)	26 (18:8)
Mean age in months (years)	82.6 (6;10)	85.1 (7;1)
Range in months (years)	54-112 (4;6-9;4)	58-107 (4;10-8;11)

Appendix 2
Table1
Distribution of items used in the Sentence Comprehension Test (SCT), n= 40

Category	Item Number	Total
Negative	14, 23	2
Modification	12,13, 24	3
Prepositional Phrase	2, 3,29,39	4
Indirect Object	8,21,31	3
Verb Phrase present	1,5,18,26	4
past	6,4,	2
future	16,40,34	3
Relative Clause	10,22, 25,28	4
Subordinate Clause	7,17,30,35,36,37	6
interrogative	11,38	2
Passive	20,33	2
Indirect Request	32	1
Coordinated sentence	9,27	2
Imperative	15	1
Topicalisation	19	1

Table2
Distribution of Items in the Expressive Language Test (ELT)

Linguistic Structure			Item Number
Possessive pronouns+Cl3[rd] Person Feminine singular			25
	3rd person plural		26
	3rd Person Masculine singular		27
	2nd person masculine Singular		28
	Feminine Singular		5,6
Subject Pronouns	Plural		15
Demonstrative Pronouns			61, 62
Reflexive Pronouns	Plural		10, 67
	Masculine Singular		68
Prepositions			1, 4
Possessive particle			9,29, 30
Plurals	Regular	Feminine	3,31, 33
		Masculine	32
	Irregular		34, 35, 36

Dual		Masculine	37, 38
		Feminine	39
Verb Markers Plural	Present	3rd Person Masculine	7, 40
		2nd person Feminine Singular	16, 41
		2nd Person Plural	42
		3rd Person Feminine Singular	2, 8
singular	Past	3rd Person masculine	11
		3rd person Plural	50
Singular		3rd person Feminine	51
singular		3rd person masculine	52
	Future		13
Construct State			43, 44
Derivation of Nouns			43, 44
Derivation of Adjectives			47, 48,49
Adjective	Plural		18,20
	Feminine		22,23
	Dual		17,19,
CliticPronouns			
	Dative Clitic 3rd Person MS		53, 54
	Object pronoun clitic 3rd FS		55,56
	Genitive (Possessive) clitic 2nd MS		57
	Genitive (Possessive) clitic 2nd MP		58
	Genitive (Possessive) clitic 3rd FS		5,6,59
	Genitive (Possessive) 3rd MS		60
	Object clitic 3rd FPl		12
	Object clitic 3rd MS		14
Comparative and Superlative			
	Comparative		63, 64
	superlative		65, 66
Negation			21, 4

Table3
Distribution of the items used in the Sentence Repetition test (n=41)

Category		Item number
Simple	Active	2,4, 10

	With noun modification	9, 16, 23, 40
	With negation	12, 15
	With coordination	6, 14, 25, 35
	Conjunction deletion	39
	Negation	21
Imperative		1, 8
Interrogative	What/where	3, 5 ,17
	With noun modification	7, 19
	With coordination	11
Complex	With relativisation	17, 24, 27, 28, 30, 33, 34, 41
	With subordination	13, 18 26, 36, 37
	Passive	
	negative	22
	with subordinate	38, 31,32
	clause	
	Topicalisation	20, 29