Diagnostic accuracy of language sample measures with Persian-speaking preschool children

Yalda Kazemi, Thomas Klee & Helen Stringer

Abstract

Purpose: This study examined the diagnostic accuracy of selected language sample measures (LSMs) with Persian-speaking children.

Method: A pre-accuracy study followed by phase I and II studies are reported. Twenty-four Persian-speaking children, aged 42 to 54 months, with primary language impairment (PLI) were compared to 27 age-matched children without PLI on a set of measures derived from play-based, conversational language samples.

Results: Correlations between age and LSMs were not statistically significant in either group of children. However, a majority of LSMs differentiated children with and without PLI at the group level (phase I), while three of the measures exhibited good diagnostic accuracy at the level of the individual (phase II).

Conclusions: General LSMs are promising for distinguishing between children with and without PLI. Persian-specific measures are mainly helpful in identifying children without language impairment whilst their ability to identify children with PLI is poor.

Key Words: language sample analysis, diagnostic accuracy, primary language impairment, pre-school children, Persian

Diagnostic Accuracy of Language Sample Measures with Persian-Speaking Preschool Children

This study examines the accuracy of language sample measures for diagnosing Iranian Persian-speaking preschool children with language difficulties not resulting from significant cognitive, sensory-motor or social deficits, a condition referred to here as primary language impairment (PLI; Law, Boyle, Harris, Harkness, & Nye, 2000; Boyle, McCartney, O'Hare, & Forbes, 2009), though the term 'specific language impairment' (SLI) is also used. Although operational definitions of PLI/SLI are usually reported in research studies, including the cut-off level used on a standardised test, the clinical diagnosis of the condition can be problematic (Betz, Eickhoff, & Sullivan, 2013), particularly in languages other than English, which may not have culturally-appropriate clinical assessment measures with known diagnostic accuracy. The motivation for conducting this study, therefore, was to provide initial data on the diagnostic accuracy of language sample measures for Persian-speaking preschool children. Before reporting three analyses based on Sackett and Haynes' (2002) framework for diagnostic research, we begin with a brief overview of the linguistic features of Persian relevant to this study, followed by child language assessments available to Iranian speech therapists and what is known about their diagnostic accuracy.

Persian grammar

Modern Persian is an Indo-Iranian language spoken by over 78 million people in Iran, Afghanistan (Dari Persian) and other countries (www.ethnologue.com). In Persian, a pro-drop language having a canonical word order of Subject-Object-Verb (SOV), morphemes can be categorised into five types:

1. lexical morphemes, an open class of vocabulary items that include nouns (e.g., کار *kar* → *work);* verbs (e.g., نویس (verb root) in می/نویس/م (inflected verb) *nevis* → *to write*); adjectives (e.g., خوب *xub* → *good*); adverbs; and prepositions (Meshkato-Dini, 2008);

2. functional morphemes, consisting of closed sets of item such as pronouns, conjunctions and direct object markers (Meshkato-Dini, 2008);
3. derivational morphemes/affixes, which play a role in novel word formation and outnumber inflectional morphemes. They are closer to the word root in proximity and change the grammatical category or subcategory of the derived word (Kalbasi, 2008; Meshkato-Dini, 2008) (e.g., درد (noun) + ناک (adjective marker): *dærd* + *nak* → pain + ful (adjective)).
4. inflectional morphemes/affixes, which are bound morphemes marking plurality and verb tense and aspect, which occur as prefixes or suffixes;
5. clitics, a class of morphemes which are not independently used but attach to antecedent or subsequent words. Unlike affixes, they are not part of the structure of the word, so the syntactic role of the attached word is not assigned to them. (See also Appendix A).

Morphologically, verbs in standard Persian contain syntactic and morphological components, including the verb root, a negation morpheme, passive and causative morphemes, auxiliary verbs, and verb particles in compound verbs (e.g. *baz kardan,* meaning 'open', is a compound verb in which *baz* is a particle that does not take inflections). Verb may be marked for tense, aspect, mood, person, number and transitivity. Only those verb roots with tense are able to be inflected. Inflections are added to the end of past or present stems as bound suffixes to form person and number agreement (Mohammad Ebrahimi Jahromi & Haghshenas, 2004). A further brief overview of Persian may be found in Samadi and Perkins (2012: 169-173). The following example illustrates some of these properties for the sentence which in English means, *I was going to good classes last year:*

Written Persian script (right to left): (من) پارسال به کلاسهای خوبی می رفتم.

Romanised transcription: (mæn)-parsal-be-kelas/ha/e-xub/i-mi\ræft/æm

Literal English translation:	(I) last year to classes good went
Clause level:	(Subject) – Adverbial – Adverbial – Verb
Phrase level:	(pronoun)-adverb-preposition-noun/plural marker/clitic-adjective/clitic-progressive verb marker\verb/person verb marker

Clinical language assessment with Persian-speaking children

Preschool children in Iran suspected of having language difficulties are diagnosed by qualified speech therapists on the basis of clinical judgement after gathering information from case histories and informal assessments of phonology, language, pragmatics and other developmental areas. Only limited clinical assessment measures designed for Persian-speaking children currently exist (Kazemi, Stringer, & Klee, (in press)). For example, the *Test of Language Development–Primary* (Newcomer & Hammill, 1997), originally developed for use with American English-speaking children, was translated into Persian and standardized on 1,235 children with unimpaired language (Hasanzadeh & Minaei, 2002). The *MacArthur-Bates Communicative Development Inventory* was culturally adapted and translated into Persian (Kazemi, Nematzadeh, Hajian, Heidari, Daneshpajouh, & Mirmoeini, 2008). Other assessments have been developed to evaluate children's phonological awareness (Soleimani & Dastjerdi Kazemi, 2005), mean length of utterance (Oryadi-Zanjani, Ghorbani & Keikha, 2006; Kazemi et al., 2012), sentence repetition (Hasanati, Agharasouli, Mahmoudi Bakhtiyari, & Kamali, 2011; Sayyahi, Soleymani, Mahmoudi Bakhtiyari, & Jalaie, 2011), oral-motor development and speech intelligibility (Kazemi & Derakhshandeh, 2007; Heydari, Torabinezhad, Agharasouli, & Hoseyni, 2011).

Another rich source of information regarding children's language abilities can be found in samples of conversational speech (Leadholm & Miller, 1992; Miller, 1981). Samadi and Perkins (1998, 2012) reported one of the first studies of Persian child language

development based on a longitudinal investigation of three Persian-speaking preschool children living in England. In doing so, they adapted Crystal, Fletcher and Garman's (1976) *Language Assessment, Remediation and Screening Procedure* (LARSP) to Persian (P-LARSP). The P-LARSP profile chart sets out a sequence of grammatical development at the clause, phrase and word levels for Persian-speaking preschool children. More recently, Foroodi-Nejad (2011) compared nine Persian-speaking children with SLI to 16 children with age-matched typically-developing language (TDL). Group differences were reported in the proportion of clitics correctly used, as well as the case marker *ra* and the progressive verb prefix *mi*. Children with SLI demonstrated poorer control of these structures than children with TDL.

While the psychometric properties of some of these clinical assessments have been reported, little is known about their ability to differentiate children with and without language impairment. Consequently, Iranian speech therapists and those serving Persian-speaking children elsewhere cannot be confident in their diagnoses of children suspected of having language impairment. As is the case elsewhere, there is at present no gold standard for diagnosing childhood language impairment in this population. Given the lack of information regarding the diagnostic accuracy of the measures, Iranian speech therapists rely mainly on clinical judgement for making diagnoses.

Purpose

The current study examined linguistic features of the spoken language of Persian-speaking preschool children with and without primary language impairment (PLI). The goal was to provide the first data regarding the diagnostic accuracy of a set of language sample measures and in doing so, begin to build an evidence base for clinical assessment of Persian-speaking children. The study had three aims: (1) to examine the statistical relationship between age and a set of language sample measures (LSMs); (2) to determine which LSMs

differentiated groups of children with and without PLI; and (3) to estimate the diagnostic accuracy of LSMs in these groups.

Method

The study was approved by the research ethics committees of Newcastle University (UK) and Isfahan University of Medical Sciences (Iran).

Participants with typically-developing language

A total of 57 children (31 boys), 42 to 54 months of age, initially thought to have TDL, were randomly selected from 10 randomly selected nurseries out of 338 registered nurseries in and around Isfahan. Each child's age and medical history, including hearing, neurological, mental and physical health, were checked using medical records kept by nurseries and by teacher report. The information was then corroborated by parents. Parents and teachers were also asked to report any concerns regarding children's speech, language and communication abilities. When concerns were raised, children were assessed using a routine speech and language assessment by the first author, a qualified Iranian speech therapist. The assessment consisted of observing the child in a free-play setting and recording a language sample between the examiner and the child. Children's hearing was screened using the whispered voice test (Pirozzo, Papinczak, & Glasziou, 2003) and speech intelligibility, sentence comprehension, sentence formulation and vocabulary were informally assessed during the language sample in order to exclude children suspected of having speech or language difficulties based on the examiner's professional judgement. This reflects current clinical practice of speech therapists working in Iran, since no culturally-appropriate, norm-referenced, preschool language tests or measures existed when the data were collected. Two of the 57 children presented with language that was not age-appropriate, as judged by the first author, and were consequently excluded. Twenty-eight children were not seen due to time constraints. The TDL group consisted of 27 children (mean age in months = 48.9; *SD* = 3.6).

Participants with primary language impairment

Children suspected of having PLI were referred by qualified speech therapists holding Master's degrees in speech-language pathology and who worked in one of six university-affiliated clinics or in other clinical settings. The children were 36 to 60 months of age and had expressive language beyond the single word stage and speech that was intelligible enough to be orthographically transcribed. They were either on a waiting list for speech-language services or had been visited maximum of three times and all were assessed with the same assessment battery described above for children in the TDL group. Of 110 children initially referred, 24 (18 boys) qualified for the study and participated (mean age in months = 47.3; *SD* = 4.2). Children were excluded who were outside the target age range (*n* = 33), unable to be contacted (*n* = 25), unwilling to participate (*n* = 5), bilingual (*n* = 2), did not attend (*n* = 2), no longer needed therapy or discharged from therapy (*n* = 4), and those with unintelligible speech or at the single-word stage (*n* = 15).

The mean age of participants in each group did not differ statistically, $t(49) = 1.47$, $p = 0.14$. Moreover, the groups did not differ with respect to gender, prematurity, history of otitis media, or family history of language-related problems. However, because it was not part of the selection criteria, the education level of parents of children with TDL was significantly higher than that of parents of PLI children, both for fathers, $t(46) = 3.15$, $p = 0.003$ and for mothers, $t(47) = 2.76$, $p = 0.008$. The mean education level of fathers was 11.3 years (SD 4.0) and 14.3 years (SD 2.6) respectively for children in the PLI and TDL groups and similar for mothers (PLI group: 11.6 years (SD 3.6); TDL group: 14.0 years (SD 2.6)).

Study protocol

After recruiting study participants, each child's mother completed a questionnaire and was interviewed about their child's health and development, after which they were invited to play with their child for a minimum of 20 minutes using a standard set of toys, including a

doll house with two dolls, a bus, a broken car, and animals on a jungle play mat. Having mothers serve as interlocutor during the language sample was thought to maximize children's language output (Haynes, Purcell, & Haynes, 1979; Eisenberg, Fersko, & Lundgren, 2001; Bornstein, Painter, & Park, 2002) and provided a familiar and comfortable environment for the child. Free-play with toys was selected because of its potential for eliciting complex linguistic constructions (Klein, Moses, & Jean-Baptiste, 2010). Mothers were the only conversational partner and were asked to use open-ended questions and follow the child's lead with respect to topic selection. They were also encouraged to talk to their child about events related to the toys that were provided rather than asking the child to count or sing. Sessions were audio-recorded using a digital voice recorder in three speech therapy centres.

Transcription

All utterances occurring during the first 20 minutes of free play were transcribed using Romanized orthography, similar to the example presented in the section on Persian grammar above. Transcription was done using the *Systematic Analysis of Language Transcripts* software (SALT; Miller & Iglesias, 2012). The software was modified by Ann Nockerts, SALT software engineer, to accommodate the linguistic features of Persian (e.g., word-initial affixes). Decisions about utterance segmentation, word roots, and morpheme boundaries followed definitions in the literature on Persian grammar and semantics (Mahootian, 1997; Nilipour & Raghibdoust, 2001; Karimi, 2003; Kalbasi, 2008; Meshkato-Dini, 2008). A written protocol developed for this study set out a range of transcription conventions, including utterance and morpheme segmentation criteria (Kazemi, 2013). All complete and intelligible utterances occurring in the transcript were included for analysis, while routinized utterances and single morpheme utterances (e.g., *no, yes*) were excluded. Single-word utterances consisting of more than one more morpheme were included. Each transcript was also coded for the type of errors observed. Transcription was done by listening

to each utterance not more than three times if there was any doubt about what was said. Transcription was done blindly with respect to the child's age, gender and clinical group.

Transcriber reliability

Ten language samples were transcribed twice, four months apart, by the first author and several measures were computed by SALT, including mean length of utterance in words (MLUw) and morphemes (MLUm), the number of different words (NDW) and the total number of words (NTW). Intra-transcriber reliability, estimated using Spearman's rho, ranged from 0.88 to 0.99 across measures (all $p < 0.001$). It was not possible to examine inter-transcriber reliability since a second person with the necessary skills was unavailable.

Language sample measures

SALT was used to calculate 33 measures from the language samples. One set consisted of general LSMs, including measures of utterance length, lexical diversity and intelligibility. Another set consisted of Persian-specific LSMs, including counts of various inflections. Table 1 provides a complete list of measures.

Insert table 1 about here

Some measures were chosen on the basis of their reported utility in other languages (Bedore & Leonard, 1998; Conti-Ramsden, 2003; Klee, Stokes, Wong, Fletcher, & Gavin, 2004; Simon-Cereijido & Gutierrez-Clellen, 2007; Klee, Gavin, & Stokes, 2007; Heilmann, Miller, & Nockerts, 2010; Wong, Klee, Stokes, Fletcher, & Leonard, 2010; Thordardottir et al., 2011; Eisenberg & Guo, 2013; Gladfelter & Leonard, 2013). Linguistic markers relevant to Persian, including case marker *ra*, object clitic, subject-verb agreement, and present tense marker *mi* were included because they had been shown to differentiate Persian-speaking children with and without PLI (Foroodi-Nejad, 2011; Maleki Shahmahmood, Soleymani, & Faghihzade, 2011).

Study framework

First, we examined the correlations between LSMs and the age of the children to determine whether any measures showed developmental change over the narrow age range of the study (pre-accuracy phase). In Phase I of the study, following Sackett and Haynes' (2002) framework for diagnostic research, we then examined whether children with TDL and PLI differed at the group level on any of the measures, using tests of mean differences (Mann-Whitney *U* test or *t*-test). The measures that did were then examined in Phase II by evaluating the diagnostic accuracy of each measure at the level of the individual, first by constructing receiver operating characteristic (ROC) curves to determine each measure's optimal diagnostic threshold and then by calculating each measure's sensitivity, specificity and likelihood ratios for positive and negative test results (Dollaghan, 2007).

Using clinical judgment to diagnose PLI is common in studies conducted in countries where child language research is still in a nascent stage and where it is not yet possible to define a reference standard (e.g., Dollaghan & Horner, 2011; Thordardottir et al., 2011). Therefore, the clinical judgment of the first author served as the reference standard, or outcome variable, against the measures were compared.

Results

Only three measures showed statistical associations with age: the frequency of plural marker */ha* ($r = 0.543$, $p = 0.003$), the frequency of utterances having wrong word order ($r = 0.450$, $p = 0.01$) and the number of utterances with the missing case marker *ra* ($r = 0.472$, $p = 0.02$).

Phase I

Table 2 presents group summaries for each of the 33 LSMs. The mean scores of the TDL and PLI groups statistically differed for all general and Persian-specific LSMs except three: total number of complete and intelligible utterances (TNU), percentage of intelligible utterances (Intelligibility %), and total number of prefixes. Language samples of children in

the PLI group contained significantly more one-word utterances (TNOU) than those of children in the TDL group, while the mean scores of all other general and Persian-specific LSMs were significantly higher for the TDL group.

Insert table 2 about here

Turning to the measures of errors listed in Table 2, significantly more errors were exhibited in the language samples of those in the PLI group than in the TDL group, apart from two, the number of missing conjunctions and missing */ha* (plural noun marker) was not statistically different in the groups.

Phase II

The original set of 33 LSMs, listed in Table 2, was reduced to 12 measures, listed in Table 3 by excluding those with areas under the curve (AUCs) less than 0.8 (Haynes, et al., 2006 p. 351). Table 3 summarises each measure's sensitivity (Sn), specificity (Sp), positive likelihood ratio (LR+), negative likelihood ratio (LR-), and 95% confidence intervals (CI), based on cut-off values determined from ROC curves and AUC. Sensitivity and specificity were calculated to estimate the accuracy of each measure in correctly identifying children with (Sn) and without (Sp) PLI. LRs were calculated to estimate how many times more likely a particular outcome, positive or negative, occurred in children with PLI than in those without PLI. These metrics provide estimates of the diagnostic accuracy of each LSM.

Insert table 3 about here

Measures demonstrating LR+ point estimates of 10 or greater were: MLUw-exc (45.92), total errors (41.44), MLUm-exc (36.96), semantic errors (24.75), and Wrong Responses (16.87). Dollaghan (2004; p. 397) suggests that 'when a measure's LR+ value is 10 or more, an individual scoring within the affected range on the measure is very likely to have the condition; when LR+ is 20 or more, such an individual is virtually certain to have

the target condition'. LR+ point estimates for the remaining measures ranged from 3.21 to 6.10.

Four measures demonstrated LR- point estimates of less than 0.20: semantic errors (0.09), MLUm and MLUw (0.011), and MLUw-exc (0.18). Dollaghan (2004; p. 397) suggests that 'when a measure has an LR- value lower than 0.20, an individual scoring within the unaffected range is very likely to be free of the target condition (in this case, PLI); when LR- is 0.10 or lower such an individual is virtually certain to be free of the condition'. The rest of the measures ranged from 0.21 to 0.39

Discussion

This study provides new evidence regarding the clinical utility of language sample measures in diagnosing PLI in Iranian Persian-speaking children. The fact that only a few of the LSMs showed statistically significant correlations with age was not unexpected, given the narrow 12-month age range of the children in this study. In the Phase I study, 27 of the 33 language sample measures differentiated children with PLI and TDL at the group level (Table 2). In the Phase II study, 12 of these measures demonstrated reasonable diagnostic accuracy at the level of the individual (Table 3). Three of them, grammaticality, MLUw-exc and semantic errors, demonstrated promising AUC values, sensitivity, specificity, and likelihood ratios.

The diagnostic values of grammaticality in the current study are very close to those reported previously (Eisenberg, Guo, & Germezia, 2012; Eisenberg & Guo, 2013), in demonstrating good sensitivity and fair specificity with a low LR-. Amongst the general grammatical LSMs, MLUw-exc showed the highest AUC, with fair sensitivity and good specificity with a high LR+. Similar measures, such as *mean syntactic length* (Klee & Fitzgerald, 1985) and *MLU2*; MLU calculated by excluding repetitions, single-word responses to yes/no questions, and elliptical question responses (Johnston, 2001; Acarlar &

Johnston, 2006), have also been used in research with English- and Turkish-speaking children.

As seen in Table 3, semantic errors demonstrated the best diagnostic accuracy of all semantic LSMs. One explanation for the large number of semantic errors observed in the language samples of children with PLI might be attributed to their problem with multi-argument structures and their vulnerability to processing verb structures having more than two arguments, due to children's limited processing capacity (see Simon-Cereijido & Gutierrez-Clellen, 2007 for a review). The verb structure of Persian has at least eight syntactic complements (or arguments). This, coupled with more than 4000 Persian verbs, means that the number of combinations of complements is quite large (Rasooli, Moloodi, Kouhestani, & Minaei-Bidgoli, 2011), a feature that is likely to affect the language processing of children with PLI.

None of the Persian-specific measures were shown to have acceptable diagnostic accuracy, which could be due to the fact that the former measures, individually, represent a single semantic or morphological phenomenon rather than encompassing several aspects of grammar. Iranian speech therapists, consequently, should be cautious in using these measures as quantitative indicators of PLI based on the results of this study. Nevertheless, these measures might still be used to describe weaknesses in expressive grammar and semantics.

Although highly sensitive measures are desirable, they should also have a low false positive rate to prevent 'time and economic burden as well as a potential psychological burden for parents' (Eisenberg & Guo, 2013: 28). In search of a language sample measure that could be used cross-linguistically (Klee et al., 2007), measures of grammaticality might be considered due to their promising results in at least two languages examined so far, English (Eisenberg & Guo, 2013) and Persian. A disadvantage with regard to other diagnostic measures of grammaticality, however, is their imprecision (i.e., wide confidence interval),

particularly for specificity. It also possesses low LR- which provides some assurance to clinicians in that scores in the normal range are characteristic of unimpaired children. The only measure of average utterance length with reasonable sensitivity and specificity was MLUw-exc, but its wide confidence interval would suggest that clinicians should be cautious in interpreting its results when identifying children with PLI. This is also supported by a high LR+, indicating some certainty in assigning atypical scores to children in the impaired range.

Future directions

First, replication of Phase II with a larger sample would help better estimate the point estimates and precision of the measures. The results of this study, however, can be applied with some confidence by clinicians working with Persian-speaking children. Secondly, following Sackett and Haynes' framework (2002), the next phase of diagnostic research, Phase III, would involve a diagnostic accuracy study of children for whom it is reasonable to suspect language difficulties, such as those referred for clinical assessment. The outcome of such a study would build on the findings reported here by having greater clinically applicability since the study design would more closely mirror a real clinical setting.

The time has come for tailoring clinical assessment tools to Persian-speaking children in Iran and eventually replacing clinical tools originally designed with English-speaking children in mind. There is a need for high-quality, evidence-based clinical measures. This study is a first attempt at what is sure to become a long-term clinical research endeavour that will require many contributors at both the clinical and research levels.

Acknowledgements

We thank the children and their parents who took part in this study and Kowsar Rehabilitation Centre, managed by Shahryar Shariat and Tila Shokrani, for providing the setting for data collection. Thanks also to Ann Nockerts and Jon Miller for modifying the SALT computer program in order to accommodate the morphological structure of Persian, and Mahboubeh Nakhshab and Leila Ghasisin for their help with handling data. We also thank Paul Fletcher and Ghada Khattab for commenting on the first author's PhD thesis on which this study based, and Nick Riches for commenting on an earlier draft of the manuscript.

Declaration of Interest

The authors have no conflicts of interest to declare.

Appendix A

Persian clitics (Kalbasi, 2008; Meshkato-Dini, 2008)

Type	Transcription	Persian example	Example transcription	English meaning
Present inflectional forms of *to be*. They are also called *enclitic verbs* or *copulative verbs.*	*/æm, /i, /æst, /im, /id, /ænd*	خوبید. معلمیم.	xub/*id*. moælem/*im*.	You *are* good. We *are* teachers.
Dependent personal pronoun (Inseparable pronouns): a. accusative b. possessive c. Prepositional Phrase (PP) complement	*/æm, /æt, /æʃ, /eman, /etan, /eʃan*	بردش. دستم برایشان	bord/*eʃ* dæst/*æm* bæray/*eʃan*	(Somebody) took *him/her*. *My* hand For *them*
/e-Ezafeh (addition) or genitive sign: a. Nominal or possessive b. Adjective	/e	کتابِ پارسا کتابِ بزرگ	ketab/*e* Parsa ketab/*e* bozorg	Parsa's book Large book
Indefinite noun marker	*/i*	دختری	doxtær/*i*	*A* girl
Definite noun marker, only in colloquial Persian	*/e*	اسبه	æsb/*e*	*That* horse

References

Acarlar, F., Johnston, J. R. (2006). Computer-based analysis of Turkish child language: clinical and research applications. *Journal of Multilingual Communication Disorders, 4*(2), 78-94. doi: 10.1080/14769670600704821

Bedore, L. M., & Leonard, L. B. (1998). Specific language impairment and grammatical morphology: a discriminant function analysis. *Journal of Speech, Language, and Hearing Research, 41*, 1185–1192. doi:10.1044/jslhr.4105.1185

Betz, S. K., Eickhoff, J. R., & Sullivan, S. F. (2013). Factors influencing the selection of standardized tests for the diagnosis of specific language impairment. *Language, Speech, and Hearing Services in Schools, 44*(2), 133-146. doi: 10.1044/0161-1461(2012/12-0093)

Bornstein, M. H., Painter, K. M., & Park, J. (2002). Naturalistic language sampling in typically developing children. *Journal of Child Language, 29*, 687-699. doi:10.1017/S030500090200524X

Boyle, J. M., McCartney, E., O'Hare, A., & Forbes, J. (2009). Direct versus indirect and individual versus group modes of language therapy for children with primary language impairment: principal outcomes from a randomized controlled trial and economic evaluation. *International Journal of Language & Communication Disorders, 44*, 826-846. doi: 10.1080/13682820802371848

Conti-Ramsden, G. (2003). Processing and linguistic markers in young children with specific language impairment (SLI). *Journal of Speech, Language, and Hearing Research, 46*, 1029-1037. doi: 10.1044/1092-4388(2003/082)

Crystal, D., Fletcher, P., & Garman, M. (1976). *The grammatical analysis of language disability*. London: Edward Arnold.

Dollaghan, C. A. (2004). Taxometric analyses of specific language impairment in 3- and 4-year-old children. *Journal of Speech Language and Hearing Research, 47*(2), 464-475. doi: 10.1044/1092-4388(2004/037)

Dollaghan, C. A. (2007). *The handbook for evidence-based practice in communication disorders*. Baltimore: Paul H. Brooks Publishing Company.

Dollaghan, C. A., & Horner, E. A. (2011). Bilingual language assessment: a meta-analysis of diagnostic accuracy. *Journal of Speech, Language, and Hearing Research, 54*, 1077-1088. doi: 10.1044/1092-4388(2010/10-0093)

Eisenberg, S. L., Fersko, T. M., & Lundgren, C. (2001). The use of MLU for identifying language impairment in preschool children: a review. *American Journal of Speech-Language Pathology, 10*, 323-342. doi:10.1044/1058-0360(2001/028)

Eisenberg, S. L., & Guo, L.-Y. (2013). Differentiating children with and without language impairment based on grammaticality. *Language, Speech, and Hearing Services in Schools, 44*, 20-31. doi: 10.1044/0161-1461(2012/11-0089)

Eisenberg, S. L., Guo, L. Y., & Germezia, M. (2012). How grammatical are 3-year-olds? *Language, Speech, and Hearing Services in Schools, 43*(1), 36-52. doi: 10.1044/0161-1461(2011/10-0093)

Foroodi-Nejad, F. (2011). *Towards the identification of linguistic characteristics of specific language impairment in Persian* (Doctoral dissertation). Retrieved from University of Alberta Library. http://hdl.handle.net/10048/2391

Gladfelter, A., & Leonard, L. B. (2013). Alternative tense and agreement morpheme measures for assessing grammatical deficits during the preschool period. *Journal of*

Speech, Language, and Hearing Research, 56, 542-552. doi: 10.1044/1092-4388(2012/12-0100)

Hasanati, F., Agharasouli, Z., Mahmoudi Bakhtiyari, B., & Kamali, M. (2011). آزمون تکرار جمله برای بررسی ساختار دستور زبان کودکان فارسی زبان [Sentence repetition test for measurement of grammatical development in Farsi speaking children]. *Audiology, 20,* 73-81. Retrieved from http://aud.tums.ac.ir

Hasanzadeh, S., & Minaei, A. (2002). انطباق و هنجاریابی آزمون رشد زبان 3 برای کودکان فارسی زبان تهرانی [Adaptating and norming Test of Language Development – Primary:3 for Persian-speaking children in Tehran]. *Research on Exceptional Children, 1*(2), 119-134. Retrieved from http://www.ensani.ir

Haynes, R. B., Sackett, D. L., Guyatt, G., & Tugwell, P. (2006). *Clinical epidemiology: how to do clinical practice research* (3rd ed.). Philadelphia: Lippincott Williams and Wilkins.

Haynes, W. O., Purcell, E., & Haynes, M. D. (1979). A pragmatic aspect of language sampling. *Language, Speech, and Hearing Services in Schools, 10,* 104-110. doi:10.1044/0161-1461.1002.104

Heydari, S., Torabinezhad, F., Agharasouli, Z., & Hoseyni, F. (2011). ساخت آزمون اندازه گیری وضوح گفتار برای کودکان بهنجار 3 تا 5 ساله [Development of speech intelligibility measurement test for 3 to 5 years old normal children]. *Audiology, 20,* 47-53. Retrieved from http://aud.tums.ac.ir

Heilmann, J. J., Miller, J. F., & Nockerts, A. (2010). Using language sample databases. *Language, Speech, and Hearing Services in Schools, 41,* 84-95. doi: 10.1044/0161-1461(2009/08-0075)

Johnston, J. R. (2001). An alternate MLU calculation: magnitude and variability of effects. *Journal of Speech, Language, and Hearing Research, 44*, 156-164. doi:10.1044/1092-4388(2001/014)

Kalbasi, I. (2008). *ساخت اشتقاقی واژه در زبان فارسی* [The derivational structure of word in Modern Persian]. Tehran: Institute for Humanities and Cultural Studies.

Karimi, S. (2003). *Word order and scrambling*. Cornwall: Wiley-Blackwell.

Kazemi, Y. (2013). *Clinical Assessment of Persian-speaking Children with Language Impairment in Iran: Exploring the Potential of Language Sample Measures* (Unpublished doctoral dissertation). Newcastle University, Newcastle upon Tyne.

Kazemi, Y., & Derakhshandeh, F. (2007). تعیین ضریب پایایی دستورالعمل کنترل حرکتی دهان/گفتار رابینز-کلی در جامعه کودکان فارسی زبان اصفهانی و ارائه برخی نمرات هنجار آنها [Exploring the normal range of scores of Persian-speaking children in Robbins-Klee oral/speech motor control protocol]. *Journal of Rehabilitation, 8*(3), 50-56. Retrieved from http://rehabilitationj.uswr.ac.ir

Kazemi, Y., Nematzadeh, S., Hajian, T., Heidari, M., Daneshpajouh, T., & Mirmoeini, M. (2008). روایی و پایایی ترجمه فارسی فرم کلمات و ژستها از فهرست تکامل برقراری ارتباط مک آرتور- بیتز [Finding the validity and reliability coefficient of MacArthur-Bates Communicative Development Inventory (CDI-Infant form) in Persian]. *Journal of Research in Rehabilitation Sciences, 3*(4), 45-51. Retrieved from http://jrrs.mui.ac.ir

Kazemi, Y., Stringer, H., Klee, T. (in press). Study of child language development and disorders in Iran: a systematic review of the literature. *Journal of Research in Medical Sciences*.

Kazemi, Y., Taheri, A., Kianfar, F., Shafiei, M., Daneshi, L., Moradi, A., . . . Zarei, M. (2012). بررسی میانگین طول گفته در کودکان 2/5 تا 5/5 ساله فارسی زبان اصفهانی [Mean Length of

Utterance (MLU) in typically-developing 2:6-5:6 year-old Farsi-speaking children in Isfahan]. *Journal of Research in Rehabilitation Sciences, 8*(5), 1-10. Retrieved from http://jrrs.mui.ac.ir

Klee, T., & Fitzgerald, M. D. (1985). The relation between grammatical development and mean length of utterance in morphemes. *Journal of Child Language, 12*(2), 251-269. doi: http://dx.doi.org/10.1017/S0305000900006437

Klee, T., Gavin, W. J., & Stokes, S. F. (2007). Utterance length and lexical diversity in American- and British-English speaking children: what is the evidence for a clinical marker of SLI? In R. Paul (Ed.), *Language disorders from a developmental perspective: essays in honour of Robin S. Chapman* (pp. 103-140). Mahwah, NJ: Lawrence Erlbaum.

Klee, T., Stokes, S. F., Wong, A. M. Y., Fletcher, P., & Gavin, W. J. (2004). Utterance length and lexical diversity in Cantonese-speaking children with and without specific language impairment. *Journal of Speech, Language and Hearing Research, 47,* 1396-1410. doi: 10.1044/1092-4388(2004/104)

Klein, H. B., Moses, N., & Jean-Baptiste, R. (2010). Influence of context on the production of complex sentences by typically developing children. *Language, Speech, and Hearing Services in Schools, 41*, 289-302. doi: 10.1044/0161-1461(2009/08-0080)

Law, J., Boyle, J., Harris, F., Harkness, A., & Nye, C. (2000). The feasibility of universal screening for primary speech and language delay: findings from a systematic review of the literature. *Developmental Medicine & Child Neurology*, *42*, 190-200.

Leadholm, B. J., & Miller, J. F. (1992). *Language sample analysis: the Wisconsin guide.* Madison, Wisconsin: Wisconsin Department of Public Instruction.

Mahootian S. (1997). *Persian*. New York: Routledge.

Maleki Shahmahmood, T., Soleymani, Z., & Faghihzade, S. (2011). بررسی عملکردهای زبانی در کودکان فارسی زبان مبتلا به آسیب ویژه زبانی [The study of language performances of Persian children with specific language impairment]. *Audiology, 20,* 11-21. Retrieved from http://aud.tums.ac.ir

Meshkato-Dini, M. (2008). *دستور زبان فارسی، واژگان و پیوندهای ساختی* [Persian Grammar, the lexical categories and merge] (3rd ed.). Tehran: SAMT.

Miller, J. F. (Ed.). (1981). *Assessing language production in children: experimental procedures.* Baltimore: University Park Press.

Miller, J. F., & Iglesias, A. (2012). *Systematic Analysis of Language Transcripts (SALT, version 8)*. Middleton, WI: SALT Software LLC.

Mohammad Ebrahimi Jahromi, Z., & Haghshenas, A. M. (2004). ساختار هسته گروه فعلی در زبان فارسی معیار [The structure of VP core in standard Persian]. *Journal of Letters and Language, 12*(15), 153-177. Retrieved from http://jll.uk.ac.ir

Newcomer, P. L, & Hammill, D. D. (1997). *TOLD-P:3: Test of Language Development,* 3rd ed. Austin, TX: Pro-Ed.

Nilipour, R., & Raghibdoust, S. (2001). Manifestations of aphasia in Persian. *Journal of Neurolinguistics, 14*, 209-230. doi: 10.1016/S0911-6044(01)00015-X

Oryadi-Zanjani, M. M., Ghorbani, R., & Keikha, F. (2006). هنجاريابي شاخص هاي تعداد كل واژه ها، ميانگين طول گفته و ميانگين طول 5 جمله طولاني در كودكان طبيعي فارسي زبان 2 تا 5 ساله شهر سمنان [Standardization of total numbers of word, mean length of utterance and mean length of 5 long sentences in normal Persian language children between 2 to 5 years old in

Semnan city]. *Koomesh, 7*(3), 177-182. Retrieved from http://koomeshjournal.semums.ac.ir

Pirozzo, S., Papinczak, T., & Glasziou, P. (2003). Whispered voice test for screening for hearing impairment in adults and children: systematic review. *BMJ, 327*, 967-971. doi: http://dx.doi.org/10.1136/bmj.327.7421.967

Rasooli, M. S., Moloodi, A., Kouhestani, M., & Minaei-Bidgoli, B. (2011). *A syntactic valency lexicon for Persian verbs: the first steps towards Persian dependency Treebank.* Paper presented at the 5th Language and Technology Conference (LTC): Human Language Technologies as a Challenge for Computer Science and Linguistics, Poznan, Poland. Paper retrieved from http://www.cs.columbia.edu/~rasooli/papers/VAL-1.pdf

Sackett, D. L., & Haynes, R. B. (2002). Evidence base of clinical diagnosis; the architecture of diagnostic research. *British Medical Journal, 324*, 539-541.

Samadi, H., & Perkins, M. R. (1998). P-LARSP: A developmental language profile for Persian. *Clinical Linguistics & Phonetics, 12*, 83-103. doi: doi:10.3109/02699209808985215

Samadi, H., & Perkins, M. (2012). Persian: devising the P-LARSP. In M. J. Ball, D. Crystal & P. Fletcher (Eds.), *Assessing grammar: the languages of LARSP*. Bristol: Multilingual Matters.

Sayyahi, F., Soleymani, Z., Mahmoudi Bakhtiyari, B., & Jalaie, S. (2011). تهیه و بررسی روایی و پایایی آزمون تکرار ناکلمه در کودکان چهار ساله [Providing a nonword repetition test in 4-year-old Persian children and determining its validity and reliability]. *Audiology, 20,* 47-53. Retrieved from http://aud.tums.ac.ir

Simon-Cereijido, G., & Gutierrez-Clellen, V. F. (2007). Spontaneous language markers of Spanish language impairment. *Applied Psycholinguistics, 28*, 317-339. doi: doi:10.1017/S0142716407070166

Soleimani, Z., & Dastjerdi Kazemi, M. (2005). تعیین روایی و اعتبار آزمون آگاهی واج شناختی [Study of validity and reliability of phonological awareness test]. *Journal of Psychology, 33*, 82-100. Retrieved from http://www.iranapsy.ir/en/journal/download

Thordardottir, E., Kehayia, E., Mazer, B., Lessard, N., Majnemer, A., Sutton, A., . . . Chilingaryan, G. (2011). Sensitivity and specificity of French language and processing measures for the identification of primary language impairment at age 5. *Journal of Speech, Language, and Hearing Research, 54*, 580-597. doi: 10.1044/1092-4388(2010/09-0196)

Wong, A. M. Y., Klee, T., Stokes, S. F., Fletcher, P., & Leonard, L. B. (2010). Differentiating Cantonese-speaking preschool children with and without SLI using MLU and lexical diversity (D). *Journal of Speech, Language, and Hearing Research, 53*, 794-799. doi: 10.1044/1092-4388(2009/08-0195)