

ASSESSMENT PROTOCOLS IN STUDYING CHILDREN'S SPEECH ACQUISITION

Pham Thi Ben^{*1} and Pham Thi Hang²

¹*Faculty of Special Education, Hanoi National University of Education*

²*Vietnam National Institute of Educational Sciences*

Abstract. Acquisition of children's speech has been conducted for a long time, across languages, and by different research designs and measurement protocols. Cross-sectional is one of the most popular research designs in studying children's speech acquisition. This paper reviews 33 studies including 29 international and four Vietnamese studies of children's speech acquisition to describe similarities as well as differences and limitations of measurement protocols used in the studies. Children's speech acquisition measurement protocol comprised different components including participant recruitment, instruments, procedure, examiners, recording, reliability, analysis, and reporting. These components need to be critically considered when studying children's speech acquisition.

Keywords: Assessment protocol, cross-sectional design, children, speech acquisition.

1. Introduction

Studying acquisition of children's speech has attracted researchers for a long time, across languages and countries. Methodological aspects of research are known to influence the data reported. Thus, it is essential to evaluate methodological considerations as they can lead to misleading generalisations when comparing data across studies [1].

Four main research designs have been used to elicit data regarding children's speech acquisition, namely diary studies, longitudinal studies, comparative studies, and cross-sectional studies [2]. The advantages and popularity for the use in studying children's speech acquisition for each design are varied. Both diary and longitudinal studies require studying children over a long period of time. Typically, diary studies include one child or two children whose speech production is recorded in a diary to examine any changes in the individual child's speech acquisition over time [3], [4]. Longitudinal studies typically investigate children to determine developmental tendencies or trajectories in a certain aspect of speech acquisition as well as to show variability within the child over time. In longitudinal studies, individual differences can be drawn precisely because data are obtained from a relatively small number of children. For example, Lleo and Prinz [5] investigated the developmental sequence of word shape acquisition of 9 children. Another example was the study undertaken by McLeod, van Doorn, and Reed [6] which explored consonant cluster development in 16 children aged two years. Másdóttir and Stokes [7] examined the growth of consonant inventories over a year by 28 typically-developing 2-year-old Icelandic-speaking children. Longitudinal data from 14 Cypriot-Greek-speaking typically-developing toddlers were obtained through four assessment sessions over a year to examine the acquisition of singleton consonants [8]. Kim, Ballard, and McCann [9] investigated

Received October 15, 2020. Revised November 14, 2020. Accepted December 7, 2020.

Contact Pham Thi Ben, e-mail address: ben.phamthi@hnue.edu.vn

age-related changes over a year of speech accuracy and error production in 16 preschool Korean-English bilingual children in New Zealand. Each of these studies included relatively small numbers of children.

Comparative studies and cross-sectional studies typically are based on larger numbers of children. The sample children are examined at one time. Comparative studies can provide data for comparison between groups, cross-linguistic or linguistic factors. For example, a comparative analysis was presented across languages about children's acquisition of consonant clusters containing /s/ in English, Hebrew, Dutch and Norwegian [10] or children's acquisition of consonants in English and Xhosa [11]. Comparative studies provided data about the relationships between neighbourhood density and speech accuracy of typically-developing children [12] or the relationships between language ability and phonological errors of typically-developing children [13]. As such, "comparative studies enhance our understanding of the nuances of children's speech acquisition" (p.190) [14].

Cross-sectional studies are an important source of typical speech acquisition data for developmental norms, which typically are generated by examining a large number of children of the same age, sex, and socio-economic status. Sample children are examined at one specific point in time and by the same testing and analysis protocols which typically are single-word picture naming tasks [15], [16]. Information about the means, standard deviations, and average age of acquisition of speech sounds (e.g., consonants and vowels) are provided by scoring their speech accuracy and analysing their mismatches with adult pronunciation. Although there is a critique that techniques of data collection are less naturalistic and data overlook individual variability, the cross-sectional design is a predominant method for studying speech acquisition because it is an efficient way to sample all speech sounds of a language in a variety of contexts. Many cross-sectional studies have been undertaken on speech acquisition of children who speak English (e.g., Arlt & Goodban, 1976; Chirlian & Sharpley, 1982; Dodd et al., 2003; Dodd & McIntosh, 2010; Kilminster & Laird, 1978; Monaghan, 2014; Pearson, Velleman, Bryant, & Charko, 2009; Poole, 1934; Prather, Hedrick, & Kern, 1975; Smit et al., 1990; Templin, 1957; Wellman et al., 1931). Cross-sectional studies have also been used to investigate speech acquisition of children who speak languages other than English such as Arabic (Amayreh & Dyson, 1998; Ayyad, Bernhardt, & Stemberger, 2016), Cantonese (Cheung, 1990; So & Dodd, 1995; To, Cheung, & McLeod, 2013), Danish (Clausen & Fox-Boyer, 2017), French (MacLeod, Sutton, Trudeau, & Thordardottir, 2011), Haitian-Creole (Archer, Champion, Tyrone, & Walters, 2018), Malay (Phoon et al., 2014), Putonghua (Hua & Dodd, 2000), and Turkish (Topbaş, 2006; Yalcinkaya, Muluk, & Budak, 2010). In addition, cross-sectional studies have been designed to examine speech acquisition of children who are bilingual or multilingual (Kim, Ballard, & McCann, 2016). Cross-sectional design also has been used in four studies of speech acquisition of Vietnamese-speaking children (Luru, 1996; Nguyen, 2011; Nguyen & Pham, 2014; Pham & McLeod, 2019).

Cross-sectional design collects speech sample from a large sample size. Each sample participant is assessed at a point of time by an assessor. To have a large sample size for a cross-sectional study, there may be a team of assessors participated in conducting direct speech assessments to collect speech samples from different sample participants. In order to be consistent among assessors in assessing sample participants, a testing protocol is required to create for a cross-sectional study on children's speech acquisition. While McLeod and Crowe [16] provide guidelines to report data on children's consonant acquisition and McLeod and Verdon [17] provide recommendations on instruments designed for collecting speech samples of children, protocols on conducting speech assessment have been not addressed yet.

This paper summarises information about children's speech assessment measurements from previous studies of speech acquisition across languages including Vietnamese speech acquisition studies used cross-sectional design. Similarities as well as differences and limitations regarding methodological aspects between the studies of speech acquisition across languages are described in order to provide assessment protocols that will be used in studies on children's speech acquisition.

2. Content

2.1. Methods

According to McLeod, a list of over 250 studies that have been written about children's speech acquisition around the world introduces on a website at <http://www.csu.edu.au/research/multilingual-speech/speech-acq-studies> [18]. Many studies in the list are written in languages other than English so the authors of this paper was not able to access these studies. Many of the speech acquisition studies in the list have included only one or two participants, so these data are not able to be generalized. This comprehensive list was created as a result of a cross-linguistic review conducted by McLeod and Crowe that examined consonant acquisition from 60 articles across 27 languages [16]. In addition, a systematic review were conducted by McLeod and Verdon (2014) to evaluate instruments used to elicit children's speech across languages. Based on the references of McLeod (2016), McLeod and Crowe (2018) and McLeod and Verdon (2014) on speech acquisition studies, the authors examined titles and abstracts to determine international studies used for this paper. Different from studies of McLeod and Crowe (2018) and McLeod and Verdon (2014), this paper focused on reviewing of speech acquisition measurement protocols. 29 international studies were selected when they met the following inclusion criteria: (a) use cross-sectional design, (b) examine more than 10 participants, and (c) publish in English regardless of which countries and languages were investigated. The 29 studies described speech acquisition in 13 languages (i.e., Arabic, Cantonese, Danish, English, French, Hebrew, Haitian-Creole, Malay, Putonghua, Setswana, Swahili, Turkish, and isiXhosa) in which 13 studies were about English.

Regarding studies on speech acquisition of Vietnamese-speaking children, the authors collated studies from theses and journal articles from personal collection. Four studies described speech acquisition of Vietnamese-speaking children and conducted in Viet Nam were found. The three studies were written and published in Vietnamese (Luu, 1996; Nguyen, 2011; Nguyen & Pham, 2014) and a study was written and published in English (Pham & McLeod, 2019).

All four Vietnamese studies were reviewed in comparison with the 29 international studies; therefore, 33 studies were reviewed in total and all were listed after references. The methodological aspects of each study are summarised and described in the sub-session 2.2. below.

2.2. Results

2.2.1. Participant recruitment

Most of the international 29 studies applied criteria to recruit participants as follows:

- approximately equal numbers of girls and boys in each age group
- participants were monolingual
- participants were from a range of socio-economic backgrounds
- participants had no intellectual or hearing impairment identified via screening tests or reported by teachers, nurses, day care providers, or parents.
- participants had no oral motor difficulties

- and participants had no history of speech or language disorders or were not currently receiving speech or language intervention services reported by parents, teachers, nurses, or day care providers.

Some exceptions to these generalisations were found. Participants' in four studies lived in multilingual contexts and their first language was examined (Gangji, Pascoe, & Smouse, 2015; Mahura & Pascoe, 2016; Maphalala, Pascoe, & Smouse, 2014; To et al., 2013) and two studies included children who were at risk of or had reported speech and language difficulties in the sample (Ayyad et al., 2016; To et al., 2013). The McLeod and Crowe (2018) cross-linguistic review recommended that studies reporting age of acquisition of consonants should report participants' demographic information including: location (e.g., country), language/dialect spoken, spoken language status (e.g., monolingual or first language), sex distribution, socio-economic status, and developmental status including hearing, speech, and language [16].

The studies of Vietnamese children's speech acquisition also applied criteria to recruit participants who were monolingual Vietnamese-speaking children, from a range of socio-economic backgrounds, and were typically developing as reported by teachers and parents. However, sex distribution in each age group was not identified in the three studies (Lru, 1996; Nguyen, 2011; Nguyen & Pham, 2014). Criteria to recruit participants were clearly defined in the study of Pham & McLeod (2019). Children who met the inclusion criteria for participating when their parents and teachers reported that: (1) they spoke Northern Vietnamese as their first language; (2) they were from the Kinh (Vietnamese) ethnic group; (3) their mother or father was from Viet Nam, not from another country; (4) their parents and teachers were not concerned about their development (e.g., cognition, behaviour, motor, speech, language, and communication); and (5) they did not have a disability that could affect their speech production, for example, a cleft lip or palate, cerebral palsy, a history of ongoing hearing loss, wearing hearing aids, or a cochlear implant. In addition, although children were identified by their parents and teachers as typically developing, they were excluded from participating when (1) they did not pass or complete the hearing screening test, or (2) they did not pass or complete the oromotor assessment.

2.2.2. Participants

Across the 29 studies, there were a total of 11,948 participants, ranging between 12 and 1,756 participants ($M = 412.00$, $SD = 480.30$). The participants typically were described in 6 monthly age groups. Most studies considered children in the age range of 2;0 - 6;0 (year; month); however, the youngest children studied were 1;0 and the oldest were 12;4.

Across Vietnamese studies, there were a total of 1,165 participants, (each included 32, 303, 635 and 195 participants respectively, $M = 291$, $SD = 246.59$, age range = 2;0-6;3). The participants were described in 12 monthly age groups (Lru, 1996), 6 monthly age groups (Nguyen & Pham, 2014; Pham & McLeod, 2019) and both 6 monthly and 12 monthly age groups (Nguyen, 2011). In the study of Nguyen and Pham (2014), participants included both children with typical development and children with pronunciation difficulties ("có khó khăn về phát âm", p. 13). However, there was inadequate information about the purposes of using this sample and how to identify pronunciation difficulties in the sample, except a description that the participants were living in an orphan house and two out of the five had a repaired cleft palate.

2.2.3. Instruments

The speech instruments used in majority of the 29 studies were single-word picture naming tasks that were either standardised tests, for instance, the Hong Kong Cantonese Articulation Test (HKCAT, Cheung, Ng, & To, 2006) used in To et al. (2013), or were developed for the purpose of the studies (e.g., Mahura & Pascoe, 2016). By using a set of target words to elicit the single-word productions, this mode of data elicitation has several advantages such as enabling

assessment of all phonemes in the language, and that children produce the same set of words. The single-word tasks that were described in 25 out of 29 studies included between 28 and 123 words ($M = 62.12$, $SD = 23.50$). Most studies indicated that the majority of words were nouns, high frequency, easy to picture, and known by most of participants in the age range (Amayreh & Dyson, 1998; Gangji et al., 2015; Hua & Dodd, 2000; Phoon et al., 2014; Smit et al., 1990; So & Dodd, 1995; Topbaş, 2006). In parallel to using a single-word task, some studies also used connected speech tasks that were either a story retell using a sequence of pictures (Hua & Dodd, 2000; So & Dodd, 1995), a spontaneous conversation related to the pictures (Topbaş, 2006), or sentence completion (Ben-David, 2006). Yalcinkaya et al. (2010) also used connected speech tasks (spontaneous or imitated) with their youngest participants who were under 2 years old.

Within the Vietnamese studies, Luru (1996) elicited spontaneous connected speech that included 98,738 utterances collected during schooltime. Single-word picture naming tasks were used in the other studies, including 284 words (Nguyen, 2011), and 50 words (Nguyen & Pham, 2014). Pham and McLeod (2019) used the Vietnamese Speech Assment (VSA, Pham, Le, & McLeod, 2016) to elicit children's speech. The VSA comprised 77 words that covered most of the potential Vietnamese speech sounds and all Vietnamese syllable shapes. The words were picturable and monosyllables. A stimulus book including 77 colourful drawings as well as instructions in a 4-step prompt hierarchy for elicitation of each target word, a stimulus book including four colourful drawings for elicitation of connected speech samples, and a recording sheet. The VSA has been conceptualised and partly operationalised.

2.2.4. Procedure

Most of the 29 studies reported the testing administration procedure as follows. Each participant was assessed:

- individually at a familiar place (e.g., at home, nursery, or kindergarten)
- in a quiet room with or without the presence of the parents or familiar adults (e.g., teachers)
- for an average of 15-20 minutes to complete all direct assessment tests (range = 5 minutes - 1 hour).

Each participant typically was instructed to respond to the single word tasks using a hierarchy to elicit as many spontaneous productions as possible in several steps: (1) firstly a spontaneous response was encouraged by asking the question such as "What is this?"; (2) if there was no spontaneous response, additional prompts and questions were provided; and (3) lastly the request for imitation was given to the participant (e.g., To et al., 2013). However, in two studies participants were instructed to imitate the examiner as the first prompt (Arlt & Goodban, 1976; Yalcinkaya et al., 2010).

Within the Vietnamese studies, Luru (1996) reported the procedure of collecting spontaneous connected speech. The authors had undertaken group observation of 60 participants in all school activities over a total of 5 days, for 6 hours each day in order to listen and record online what participants said. Connected speech collected from two individual participants was recorded in a diary during everyday activities at home. The two other studies also reported the procedure of test administration in providing instruction for participants by questions and prompts (Nguyen, 2011; Nguyen & Pham, 2014). However, the protocol of testing and instructed questions and prompts were not described and no examples were given. Thus, it was unclear whether words were elicited spontaneously or via imitation, and whether the assessments were administered across participants using the same testing procedure or not. In Pham and McLeod (2019)'s study, each participant was assessed individually in a designated room away from the classroom in the participants' preschools. Some parents or teachers were present with some participants during direct assessment sessions but they were advised to be

nonparticipating observers. A sound level meter was used to ensure that the room had acceptable background noise levels for the tests. The signal to noise ratio was measured. Before each assessment session, rapport with the child was established and child assent was gained. The speech assessment session took approximately 10-20 minutes for the 4- to 5-year-old children and 20-30 minutes for the 2- to 3-year-old children to complete. The examiner administered the speech assessment session consistently across participants by following the same testing procedure of a 4-step prompt hierarchy that was designed for the VSA. The examiner also recorded the cueing level used for each response to elicit each target word.

2.2.5. Examiners

Across the 29 studies, examiners typically were certified speech-language pathologists or graduate students in speech-language pathology with training in phonetic transcription. One study used examiners who were special educators with at least two years of training in phonetic transcription and were supervised by a speech-language pathologist during assessment sessions (Chirlian & Sharpley, 1982). In many studies the examiners were trained to undertake the data collection and phonetic transcription relevant to the study (e.g., Chirlian & Sharpley, 1982; Dodd et al., 2003; Hua & Dodd, 2000; Topbaş, 2006). The examiners were usually reported to be native speakers of the language and dialect that was tested. For example, MacLeod et al. (2011) used examiners who were Quebec French speakers, and Amayreh and Dyson (1998) used examiners who were native Jordanian Arabic speakers, while Ayyad et al. (2016) used examiners who were native Kuwaiti Arabic speakers.

Within the Vietnamese studies, examiners in one study were the author who was a researcher in early childhood education (Luu, 1996). The two other studies did not report who were examiners; however, preschool teachers and undergraduate students in primary education were acknowledged as testing contributors for the data (Nguyen, 2011; Nguyen & Pham, 2014). The author was an examiner for all of the direct assessments conducted in Pham and McLeod (2019) study. The author was a special educator and a native speaker of Northern Vietnamese, had extensive experience using the International Phonetic Alphabet to transcribe Vietnamese, had many years' experience working with children (Smit, 1986), and had been trained to conduct speech assessment.

2.2.6. Recording of participants' productions

All of the 29 studies recorded the participants' productions online in real time and 19 out of 29 studies additionally used audio or/and video recordings to check transcription at a later time. In most of the studies ($n = 25$), the participants' responses were transcribed (at the phoneme or word level) using the International Phonetic Alphabet. A few studies ($n = 3$) used a binary judgment of a correct or incorrect production based on orthography.

Three of the Vietnamese studies recorded the participants' productions online in real time and scored as a binary judgment of a correct or incorrect production based on orthography. Audio or video recording was not used, transcriptions thus were not checked. In one study, the participants' responses were recorded online in real time and whole words were transcribed on the recording sheet by using the International Phonetic Alphabet to enable consideration of the accuracy of production and dialectal variants. Children's responses were digitally audio and video recorded with the permission of parents and the preschool. The audio recordings were made by using a Zoom H1 audio recorder, were saved in .wav format, and were used to re-transcribe the responses of all participants after the assessment sessions by the author. The video recordings were made by using a Sony HDR-PJ790VE video camera that was set up to provide a front view of the children's faces. The video recordings were used to check when there was a discrepancy between the online and audio transcriptions. The audio and video recordings were

used for double checking with the online transcriptions to increase accuracy of transcriptions for the analysis (Pham & McLeod, 2019).

2.2.7. Reliability

Most of the 29 studies ($n = 26$) presented data regarding the reliability of transcription (phonetic or phonemic) of children's speech. The exceptions were the three studies undertaken over 40 years ago: Poole (1934), Templin (1957), and Kilminster and Laird (1978). Inter-judge reliability of approximately 10% of the total samples was reported and ranged between 72.5% and 100% across two or more examiners. Intra-judge reliability was reported in six studies (Ben-David, 2006; Dodd & McIntosh, 2010; Gangji et al., 2015; Mahura & Pascoe, 2016; Maphalala et al., 2014; So & Dodd, 1995) and ranged from 84.78% to 98%. Reliability data showed a higher agreement for experienced examiners and for examiners who were trained in phonetic transcription (e.g., Hua & Dodd, 2000; So & Dodd, 1995). Reliability data showed higher agreement using both online transcription that was double checked using audio or video recordings than only using online transcription in real time (Hua & Dodd, 2000).

Reliability of transcription (intra-judge or inter-judge) was not described in three Vietnamese studies (Luu, 1996; Nguyen, 2011; Nguyen & Pham, 2014). In study from Pham and McLeod (2019), intra- and inter-judge point-to-point reliability was conducted for consonants, semivowels, vowels, and tones for the speech assessment sessions. Intra-judge reliability was checked by using the audio and video recordings. The author re-transcribed 10.3% of audio recordings to compare with the original online transcription to determine intra-judge reliability, reaching to an agreement of 95.1%. Inter-judge reliability was checked by the author and a Vietnamese speech-language pathologist. Online transcriptions were created independently then double checked via audio recordings after the session. The transcriptions were compared and inter-judge reliability reached to an agreement of 96.1% for 5.6% of the sample.

2.2.8. Analyses of speech acquisition

Four main analyses of speech acquisition data are typically considered in speech acquisition studies throughout the world (McLeod, 2010):

- (1) Comparison of the child's speech sound with the adult target (relational analysis)
 - Age of acquisition of speech sounds
 - Percentage of consonant correct
 - Common mismatches
 - Phonological processes
- (2) Description of the abilities of the child (without comparison with the adult target) (independent analysis)
 - Phonetic inventory
 - Syllable structure
- (3) Prosody (stress, intonation, and tones)
- (4) Intelligibility (oral communication competence).

All of the 29 studies collected data about children's production of consonants. Most studies considered consonants in word-initial and word-final position, while some studies also included within word consonants. Some studies also collected data about children's production of vowels (e.g., Dodd et al., 2003; Prather et al., 1975), consonant clusters (Pearson et al., 2009; Smit et al., 1990; Templin, 1957; Wellman et al., 1931), and tones (e.g., Cheung, 1990; Hua & Dodd, 2000; So & Dodd, 1995; To et al., 2013). Recall, this analysis focussed on 29 studies of consonant acquisition. Other studies have been undertaken that specifically focus on vowels (e.g., Donegan, 2002; Pollock & Berni, 2003), and consonant clusters (e.g., Kirk, 2008;

McLeod, van Doorn & Reed, 2001; Phoon, Maclagan, & Abdullah, 2015). Analyses of the samples included:

- age of acquisition (criteria = customary 50%, acquired 75%, and mastery 90%)
- percentage of consonants, vowels, phonemes, and tones correct (PCC, PVC, PPC, PTC)
- phonological processes/patterns
- inventories of phonemes
- inventories of syllables and word shapes
- whole word proximity (WWP)

Within the Vietnamese studies, all collected data about children's production of consonants (in syllable-initial and syllable-final positions), semivowels, vowels, and tones but not all possible Vietnamese phonemes were examined. Analysis of the samples described age of acquisition by listing speech sounds that “xuất hiện” (appeared) (Luu, 1996, n.p.) or were produced correctly in children's speech. No account was made of the impact of dialect on children productions. Nguyen (2011) used an analysis of the percentage of individual consonants, semivowels, vowels, and tones correct, while Nguyen and Pham (2014) used an analysis of the percentage of individual consonants, semivowels, vowels, and tones that were incorrect. However, the criteria used in the both studies to define age of acquisition were unclear. The description of substitution, omission, distortion, and addition (SODA) in speech production was used to describe mismatches. An analysis of phonological processes/patterns was not reported. Based on the age of acquisition for individual consonants, semivowels, vowels, and tones correct and incorrect gained from the three studies, it could be discerned that the tones were acquired earliest followed by vowels, consonants, and semivowels (Luu, 1996; Nguyen, 2011). Regarding initial consonants, the last late consonants to be acquired included /f, t^h, s, ʃ, z, k, ɣ, x/ (Nguyen & Pham, 2014). Not all of the Vietnamese consonants, semivowels, vowels, and tones have been assessed.

Pham and McLeod (2019) employed relational analyses with regards to degree of production accuracy and the percentage of children achieving the specified accuracy level (75% and 90%) in each age range. Relational analyses which compare the children's productions with the adult form and determine error patterns in children's speech were applied. The number of correct productions or acceptable responses which were considered the effect of dialectal variants were calculated for each participant. The raw scores of these correct productions or acceptable responses were formed into a number of speech accuracy variables: percentage of consonants correct (PCC), percentage of syllable-initial consonants correct (PICC), percentage of syllable-final consonants correct (PFCC), percentage of semivowels correct (PSVC), percentage of vowels correct (PVC), percentage of tones correct (PTC), percentage of phonemes correct (PPC), and percentage of word shapes correct (PWC). PCC was calculated by combining the number of correct productions or acceptable responses of syllable-initial consonants, syllable-final consonants, and semivowels and PCC was used to be consistent with previous international studies. The analyses of phonological processes/patterns and substitutions, omissions, distortion, and additions (SODA) were calculated to determine the number of occurrences. Descriptive statistics and frequency distribution on the sample population as well as the significance of relationships between independent variables (e.g., age, sex, and socioeconomic status) and dependent variables (e.g., PPC, PCC, PVC) by hierarchical regression analyses were applied. Criteria of 75% and 90% across all children and all opportunities that each phoneme was produced correctly was recorded and applied to determine the age of acquisition. Non-adult realisations and phonological processes/patterns were analysed from participants' speech productions. Non-adult realisations included any phoneme

substitution, deletion, distortion, or addition and were scored as incorrect. Phonological processes/patterns that occurred more than 10% were summarised for syllable-initial and syllable-final consonants, and semivowels. Independent analysis was not applied in this study to explore phonetic inventory and syllable structure.

As stated previously, cross-sectional design is the most common design used for studying children's speech acquisition in comparison to the diary, longitudinal and comparative designs. Cross-sectional design investigates on a large sample size; therefore, it enables to provide data about the means, standard deviations, and average age of acquisition of speech sounds (e.g., consonants and vowels) by scoring speech accuracy. Although there is a critique that techniques of data collection are less naturalistic and data overlook individual variability than that in the diary and longitudinal studies, the cross-sectional design is a predominant method for studying speech acquisition because it is an efficient way to sample all speech sounds of a language in a variety of contexts. Since each of the sample participants is examined at one time, cross-sectional design requires to use the same protocol of testing and analysing for all participants to ensure data from each participant are collected equally.

3. Conclusions

In summary, the information summarised from 29 cross-sectional studies across languages and four cross-sectional Vietnamese studies above has provided useful methodological guidance for studying speech acquisition of children across languages including Vietnamese-speaking children. The findings reported in the previous Vietnamese studies have provided valuable groundwork regarding Vietnamese children's speech acquisition but caution is required when interpreting these findings because of shortcomings with regards to participant recruitment, data collection methods, and analyses compared to the international standards. Therefore, all aspects of the speech assessment protocol including: participant selection, instruments, procedure, examiners, recording, reliability, analysis, and reporting should be critically considered when conducting studies of speech acquisition of children across languages and Vietnamese. These elements of the assessment protocol will be applied in future studies on Vietnamese-speaking children's speech acquisition to ensure the Vietnamese studies are align with the international studies and international clinical guidelines.

Acknowledgments: This research is funded by Vietnam National Foundation for Science and Technology Development (NAFOSTED) under grant number 503.02-2020.312.

REFERENCES

- [1] Jimenez, B. C. 1987. Acquisition of Spanish consonants in children aged 3–5 years, 7 months. *Language, Speech, and Hearing Services in Schools*, 18(4), 357-363. doi:10.1044/0161-1461.1804.357
- [2] McLeod, S. 2017. Speech sound acquisition. In J. E. Bernthal, N. W. Bankson, & P. Flipsen (Eds.), *Articulation and phonological disorders: Speech sound disorders in children* (8th ed., pp. 58-113). Boston, MA: Pearson Education.
- [3] Lleo, C., 1990. Homonymy and reduplication: On the extended availability of two strategies in phonological acquisition. *Journal of Child Language*, 17, 267-278. doi:10.1017/S0305000900013763
- [4] Lru, T. L., 1996. *The language development stages of children aged 1-6 (basing on the sample of children's language in the intercity of Hanoi)*. (Unpublished PhD thesis), University of Social Sciences and Humanities, VNU Hanoi.

- [5] Lleo, C., & Prinz, M., 1996. Consonant clusters in child phonology and the directionality of syllable structure assignment. *Journal of Child Language*, 23, 31-56. doi:10.1017/S0305000900010084
- [6] McLeod, S., van Doorn, J., & Reed, V. A., 2001. Consonant cluster development in two-year-olds: General trends and individual difference. *Journal of Speech, Language, Hearing Research*, 44, 1144-1171. doi:10.1044/1092-4388(2001/090)
- [7] Másdóttir, T., & Stokes, S. F., 2016. Influence of consonant frequency on Icelandic-speaking children's speech acquisition. *International Journal of Speech-Language Pathology*, 18(2), 111-121. doi:10.3109/17549507.2015.1060525
- [8] Petinou, K., & Theodorou, E., 2016. Early phonetic development in typically developing children: A longitudinal investigation from Cypriot-Greek child data. *Clinical Linguistics and Phonetics*, 30(1), 12-28. doi:10.3109/02699206.2015.1095244
- [9] Kim, J.-H., Ballard, E., & McCann, C., 2018. Age-related changes in segmental accuracy and error production in Korean-English bilingual children: Implications for clinical speech sound assessment procedures. *Speech, Language and Hearing*, 21(1), 41-54. doi:10.1080/2050571X.2017.1323700
- [10] Yavaş, M., Ben-David, A., Gerrits, E., Kristoffersen, K. E., & Simonsen, H. G., 2008. Sonority and cross-linguistic acquisition of initial s-clusters. *Clinical Linguistics and Phonetics*, 22(6), 421-441. doi:10.1080/02699200701875864.
- [11] Mowrer, D. E., & Burger, S., 1991. A comparative analysis of phonological acquisition of consonants in the speech of 2½-6-year-old Xhosa-and English-speaking children. *Clinical Linguistics and Phonetics*, 5(2), 139-164. doi:10.3109/02699209108985510
- [12] Sosa, A. V., & Stoel-Gammon, C., 2012. Lexical and phonological effects in early word production. *Journal of Speech, Language, and Hearing Research*, 55(2), 596-608. doi:10.1044/1092-4388(2011/10-0113)
- [13] Roulstone, S., Loader, S., Northstone, K., & Beveridge, M., 2002. The speech and language of children aged 25 months: Descriptive data from the Avon Longitudinal Study of Parents and Children. *Early Child Development and Care*, 172(3), 259-268.
- [14] McLeod, S., & Baker, E., 2017. *Children's speech: An evidence-based approach to assessment and intervention*. Boston, MA: Pearson Education.
- [15] Hua, Z., & David, A., 2008. Study design: Cross-sectional, longitudinal, case, and group. In W. Li & M. G. Moyer (Eds.), *The Blackwell guide to research methods in bilingualism and multilingualism* (pp. 88-107). Malden, MA: Blackwell.
- [16] McLeod, S., & Crowe, K., 2018. Children's consonant acquisition in 27 languages: A cross-linguistic review. *American Journal of Speech-Language Pathology*, 27, 1546-1571. https://doi.org/10.1044/2018_AJSLP-17-0100.
- [17] McLeod, S. & Verdon, S., 2014. A review of 30 speech assessments in 19 languages other than English. *American Journal of Speech-Language Pathology*, 23, 708-723. https://doi.org/10.1044/2014_AJSLP-13-0066.
- [18] McLeod, S., 2016. *Multilingual speech acquisition studies*. Bathurst, NSW, Australia: Charles Sturt University. Retrieved from <http://www.csu.edu.au/research/multilingual-speech/speech-acq-studies>.