Phonological Development of Typically Developing Monolingual Saudi Hejazi Arabic-speaking children Aged 2;6–5;11

Deema Fawzi Turki

A thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy

The University of Sheffield
Faculty of Medicine, Dentistry and Health
Department of Human Communication Sciences

September 2022

Abstract

A reliable diagnosis of speech sound disorders (SSDs) in any language requires the availability of assessment tools following international linguistic and psychometric guidelines (Hua, 2006) and normative data on speech development for that language (Fabiano-Smith, 2019). To date, these criteria still need to be fulfilled for Saudi Hejazi Arabic (SHA). Normative data on phonological development that is theoretically and developmentally reliable is still missing for SHA-speaking children.

This thesis aimed to describe the typical phonetic and phonological development in monolingual SHA-speaking pre-schoolers. Speech samples were collected cross-sectionally using a newly designed, linguistically controlled picture naming test (SHAPA) and a phone–imitation (stimulability) task from 235 SHA-speaking children in Saudi Arabia, aged 2;6–5;11 (7x6–month age bands). SHAPA includes 151 words to elicit urban Hejazi Arabic (UHA) and Modern Standard Arabic (MSA) phoneme inventory. All samples were phonetically transcribed using Phon (Rose & MacWhinney, 2014).

SHAPA was evaluated regarding its construction criteria, items' familiarities and suitability as a clinical assessment tool. The speech data were analysed regarding 1) the age of SHA-speaking children acquire and master the Arabic (UHA and MSA) phonetic system (PCC, PCC-R, PVC, PCCC, C and CC structures' age of acquisition), and 2) the occurrence (Tokens and Types) of phonological patterns typically for SHA-speaking children, and those phonological variants that occur in children's speech but not frequently enough to be counted as phonological patterns (InfVar). Two cut-off criteria (≥4 and ≥6) were applied to distinguish InfVar from phonological patterns.

SHAPA successfully elicited a sample for phonetic and phonological analysis since its construction requirements nearly met the international construction criteria. However, it was a lengthy test with a low incidence of spontaneous naming. All quantitative and qualitative measurements revealed a general developmental tendency, with an acquisition rate almost identical to prior Arabic research.

Phonetic analysis indicated that SHA-speaking children achieved an average score of >80% for PCC and PCC-R by age 4;0, and they approached 90% by age 5;0. They acquired all shared UHA and MSA consonants (i.e. 75% accuracy) by age 3;11 and mastered them by age 4;5 (i.e. 90% accuracy), except for postalveolar /ʃ/ and /ʒ/. The stimulability task revealed that the phonetic inventory of the children in the oldest age group was not completed (for both UHA and MSA phones) because some phones were difficult to be imitated (e.g. the emphatic consonant and the trill /r/).

The number of Tokens, Types, and InfrVar decreased steadily with age. Twenty-seven patterns were identified in children's speech, with only four phonological patterns and one phonetic distortion existing across all age groups using both cut-off criteria. While all structural simplifications were overcome by 4;11, certain systemic simplifications and one phonetic distortion were still found when the upper cut-off criterion was used (i.e., fronting of /ʃ, ʒ/, devoicing, vowel substitution, and sibilants distortion /s, z, ʃ, ʒ/).

These findings highlighted that children's phonetic and phonological development was not completed by age 5;11. The theoretical and clinical implications of the results were discussed.

Acknowledgments

This thesis project would not have been possible without the support of many people. First and foremost, I am incredibly grateful to my supervisor Dr Silke Fricke and my advisor Prof. Annette Fox–Boyer. They read my numerous revisions and helped make sense of the confusion. I have learned a lot from their detailed feedback and comments, as well as our discussions during our meetings (face–to–face and online) were sources of inspiration. Their knowledge and ample experience have encouraged me during my whole PhD journey. They gave me endless support and were always there whenever I needed help.

Additionally, I thank my second supervisor Dr Traci Walker for her generous support, especially during the first year of my PhD journey. Without her help, I would not pass my confirmation review exam. I would like also to acknowledge the help of Dr Blanca Schafer, who supervised me at the beginning of my study.

Special thanks are due to Dr Yvan Rose, who did an excellent job teaching me how to use PHON software. His valuable support made my analysis step much more accessible and more straightforward. I also thank Dr Katie Finnegan for her proofreading services.

I would also express my sincere gratitude to King Abdul-Aziz University for giving me this opportunity to achieve my dream and sponsoring my education from September 2016 until I submit this thesis. Mainly thanks to Dr Nada Fakih, who helped me during my data collection trip and facilitated the paperwork needed for this journey. I also wish to thank Dr Nahla Dashash and Dr Yasser Alsa'bi, who assisted me during the data collection trips. A special thanks go to my SLT friends, who helped me with all reliability checks for the IPA phonetic transcriptions and phonological patterns.

Many thanks should also go to the Ministry of Education in Saudi Arabia, the director of kindergarten education, who gave me permission to visit kindergarten schools in Jeddah. I also very much appreciate all schools' principals and teachers who played a decisive role in my data collection. Special thanks to all interested parents and lovely children who participated in this research and helped to make this project possible.

My success would not have been possible without the support and nurturing of my parent, who have supported me in many ways, and always kept me in their prayer and thoughts. Thanks also to my little sister, Shaza. Her endless help during the last year of PhD after I had my new baby boy, Awab, was priceless. Finally, I would like to express my deepest gratitude to my lovely husband, Majed, for being a supportive person throughout the journey, and to my wonderful children, Jood and Saud. They have always supported me and never stopped encouraging and believing in me.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

AAT	Amayreh Articulation Test
CA	Classical Arabic
CDI	Communication Developmental Inventory
CCs	Consonant Clusters
CS	Connected Speech
HA	Hejazi Arabic
InfrVar	Infrequent Variants
IPA	International Phonetic Alphabet
JA-CDI	JISH Arabic– Communication Developmental Inventory
JAT	JISH Articulation Test
MAAT	Mansoura Arabic Articulation Test
MSA	Modern Standard Arabic
PCC	Percentage of consonants correct
PCCC	Percentage of consonant clusters correct
PCC-R	Percentage of consonants correct–Revised
PhonVar	Phonological Variants
PVC	Percentage of Vowels correct
SA	Saudi Arabia
SD	Standard Deviations
SHA	Saudi Hejazi Arabic
SHAPA	Saudi Hejazi Arabic Phonology Assessment
SLTs	Speech language Therapists
SSDs	Speech Sound Disorders
UHA	Urban Hejazi Arabic

Conventions

/ħu.sˤa:n/	target phonemic realisation of a word
[ħu.sa:n]	phonetic realisation of a word, actual realisation of the child
‘horse’	translation of Arabic word into English

Age display: 2;6 years= 2 years and six months

Declaration

I, Deema Turki, confirm that the Thesis is my own work. I am aware of the University's Guidance on the Use of Unfair Means (www.sheffield.ac.uk/ssid/unfair–means). This work has not been previously been presented for an award at this, or any other, university.

Introduction

Phonology is one of the critical dimensions involved in mastering a first language. Phonological development starts in the early ages of language acquisition and is almost complete by the age of 4 to 5 years (Abou-Elsaad, Afsah, & Rabea, 2019). Children's acquisition of phonology entails mastery of the production and perception of consonants, vowels, consonant clusters, prosodic elements, and the phonological rules of their language, leading to intelligible speech (McLeod & Crowe, 2018).

Acquiring intelligible speech requires children to master the phonological forms of words and phrases of their native language, as well as the articulatory and phonatory gestures required to generate these words and phrases in an adult-like manner. As a result, phonological development is seen to consist of two major components: the cognitive-linguistic component involved with learning the ambient phonological system and the development of speech-motor skills required for adult-like outputs (Stoel-Gammon & Sosa, 2007). A modern perspective of phonological acquisition incorporates many other components in addition to these two components, including social capacities and feedback from adult users of the ambient language that aid infants in exploiting their innate skills for acquiring phonological information and behavioural patterns (Davis & Bedore, 2013). However, researchers are still investigating the components that drive children's acquisition of phonology and if the patterns of acquisition are similar or different across languages. Phonological development theories have sought to give explanatory frameworks for phonological acquisition. Despite the fact that most modern phonological theories emphasise the vast degree of similarity among languages, they have attempted to account for language-specific developmental patterns. These theories agree on the occurrence of both phenomena; however, they argue on which aspects of speech development are of a universal nature, whether or not they are innate, and how children begin to learn the phonology of their languages, and what units are considered the building blocks of phonological acquisition.

While research indicates many similarities in phonological acquisition across languages (McLeod & Crowe, 2018), there are some highly significant language-specific aspects that could influence the phonological development. For example, some phonological patterns occur in many languages (e.g. velar fronting) are not necessarily overcome at the same age in each language (Clausen & Fox- Boyer, 2017). Additionally, other patterns can be language-specific ones (e.g. gemination reduction) that occur due to specific language feature (e.g. geminated consonants in Italian and Arabic) (Dodd, Holm, Hua, & Crosbie, 2003; Fox-Boyer, Lavaggi, & Fricke, 2021). In consonantal acquisition, for example, the rate of acquisition of a particular phoneme or syllable component can differ across languages (Hua & Dodd, 2006) (e.g. the Arabic consonants /f/, /t/, /l/ were found to be acquired earlier in Arabic than in English) (Amayreh & Dyson, 1998). This indicates that speech development might not be a purely innate universal process and it could be affected by the ambient language (Clausen & Fox-Boyer, 2017). Ingram (1991) proposed that the phonological system acquired may influence children's type of errors, and thus the individual systems of different languages deserve research attention. Therefore, cross–linguistic data can help in gaining understanding of which aspects of the speech acquisition process might be due to innate abilities and functions (i.e. universal tendencies) and which result from language–specific input and the influence of the ambient language phonological system (Hua & Zhu, 2002b).

In recent years, many studies have investigated the language-specific phonological development of children acquiring different languages. Such studies usually aim to address theoretical and applied information about language acquisition, particularly, phonological development. Theoretically, these data add valuable insights to the discussion about the nature of phonological development in typically developing children acquiring one or two languages. Clinically, they form a baseline for the identification of children with typical versus atypical development (Hua & Zhu, 2002a; Stackhouse & Wells, 1997). Another applied/clinical principle behind such studies is to provide speech-language therapists (SLTs) with a developmental norm on which they can decide whether a child's speech sound acquisition is developing typically or whether intervention is required (Bland-Stewart, 2003; Dodd, Holm, Crosbie, & Hua, 2005; Waring & Knight, 2013). These developmental norms for the typical acquisition of speech are considered essential applied knowledge for SLTs to assist their clinical decision–making, considering other areas including the impact of the environment and communicative intent, for example (Cohen & Anderson, 2011; Dodd, 2014; McLeod & Baker, 2014; Nojavan- Pirehyousefan, Zarifian, Ahmadi, & Pascoe, 2022).

Early identification and intervention of children with speech sound disorder (SSD) is essential because it is the most common difficulty facing children in preschool. McLeod and Baker (2014) reported that 40–50% of the SLTs' caseload is children with SSD, particularly in the USA, the UK, and Australia. In Saudi Arabia (SA), children with SSD represented almost 30% of school-age children (Al-Sabi, 2017), with nearly 67% of the caseload of SLTs in SA being SSD cases (Alanazi, 2017). It was found that preschool SSDs in some children are associated with problems of phonological awareness, which is an important skill for developing reading and spelling skills (Preston, Hull, & Edwards, 2013). Further, SSD may be associated with other forms of communication impairment such as language production difficulties and poor intelligibility, which may negatively affect social communication skills, school education, literacy skills, and employment prospects (McCormack, McLeod, McAllister, & Harrison, 2009). Therefore, accurate and comprehensive assessment of children for identification and remediation of SSD is essential to design the best intervention protocol and to avoid possible detrimental impacts on children's participation in society (Ahmadi et al., 2018; McLeod & Verdon, 2014).

Assessment of children with SSD involves several domains, such as an assessment of the phonological processes, speech intelligibility, speech stimulability, oro-motor function, and speech sound production in connected speech (CS) and at the single-word level (Nojavan-Pirehyousefan et al., 2021). Single- word naming tests with appropriate psychometric properties are applied routinely for the assessment of SSD (Ahmadi et al., 2018). However, such an assessment tool needs to be developed carefully with consideration of the culture and language system that is being assessed. The content of the speech sound assessment tool must be critically evaluated from both a cultural and a linguistic perspective before the tool is used to ensure that it is appropriate for the population it will be used with. Vocabulary items and pictures in the assessment tool must be familiar to the children who will be assessed. Adult target pronunciations presented in the tool need to be the same as those of the child's dialect (McLeod & Verdon, 2014). McLeod and Verdon (2014) identified a shortage of culturally appropriate assessment tests as a challenge for SLTs working with children from different cultures. Two major obstacles that motivate researchers to create a reliable and practical test for measuring the production of speech sounds are a lack of normative data on the acquisition of speech sounds and the lack of relevant assessment tools available to identify SSD (Kehoe, Niederberger, & Bouchut, 2021).

The phonological systems of two different dialects of a language usually vary in several significant respects (Velleman & Pearson, 2010). For example, two dialects of a language could differ in the frequency of occurrence of different consonants and word shapes, and this could result in differing timetables for the acquisition of the language's consonants depending on the dialect (Pearson, Velleman, Bryant, & Charko, 2009). McLeod et al. (2013) asserted that it is crucial to investigate the dialectal differences of a language for a valid diagnosis of SSD; however, such differences are understudied in linguistic research, particularly in speech-language therapy research.

Arabic is known for its different dialects, which show a wide range of variation. At the same time, they are similar enough to provide a basis for meaningful comparison of their phonological features (Hellmuth, 2013). Although many studies have investigated the phonological development and developmental norms of Arabic children who speak different dialects (e.g. Jordanian Arabic, Egyptian Arabic, and Kuwaiti Arabic), their results have shown some differences in the acquisition of Arabic speech sounds and in the identified norms among their samples, which could be due to dialectal differences. According to Watson (2002), Arabic dialects have 'a continuum spectrum of variation' that makes some dialects unintelligible for some Arabs unfamiliar with these dialects (p.8).

The few studies that have been conducted on the phonological development of Saudi Arabic-speaking children have used researcher-made tools in which psychometric properties are not evaluated, international criteria may not be followed or else are performed on relatively small samples (e.g. Abdoh, 2010; Al-Bader, 2009; Alawwad, 2009). For example, Al-Bader (2009) investigated the phonological patterns in the speech of 20 Saudi Najdi Arabic-speaking children using a single-word naming task that included 173 words. Despite the considerable number of words, her word list was designed only to test 16 phonological patterns without having the Saudi Najdi speech sound in every possible word position. It has been found that phoneme accuracy alters according to the number of syllables in words, the nature of sound sequences and the stress of the syllables in which the phonemes occur (James, 2001a). Therefore, it seems crucial to examine each speech sound in all possible syllable positions in a language to obtain a representative speech sample. Such linguistic criteria are presented in many instruments designed to assess the speech sound production of English-speaking children. However, a well-designed phonological assessment tool that is linguistically and culturally appropriate for Saudi Arabic-speaking children, Hejazi in particular, is still lacking.

In SA, SLTs face significant challenges when assessing speech sound skills in children due to the absence of normative data on the development of phonological abilities among Saudi children. The variety of dialects spoken in SA is another cause of difficulty, where there are five dialects spoken by adult Saudis (i.e. Najdi, Hejazi, Northern, Southern, and Gulf). This gap in knowledge causes challenges for SLTs assessing and evaluating typical versus atypical speech development, as well as diagnosing and analysing the phonological components of misarticulated speech in children. Because many different dialects are spoken in SA, this thesis considers the development of phonological features unique to Saudi Hejazi Arabic (hereafter 'SHA'), the dialect spoken in the western region of SA. The reasons for choosing this dialect are discussed below.

The phonological development of Saudi Hejazi Arabic-speaking children (hereafter 'SHA-speaking children') is worth studying for several reasons:

- First of all, SHA is one of the Arabic varieties that can be understood throughout the whole Arab world (Basalamah, 1990) because it shares many phonological features with Modern Standard Arabic (MSA), which makes its word structures similar to those of MSA words without the large amount of epenthesis found in other Arabic dialects. Identifying the developmental norms of phonological acquisition in children speaking this dialect could assist Arab SLTs by providing them with preliminary developmental norms based on Arabic-speaking children rather than using the English norms that are widely used.
- Second, almost 14 million people speak this dialect of Arabic, which is around half of the Saudi population. Therefore, research findings concerning speech sound development in children speaking this dialect may have implications for many Arabs.
- Third, there is a dearth of research into Arabic speech sound acquisition or disorders, and to date only a few Arabic dialects have been investigated in this area.
- Fourth, as mentioned previously, SLTs working with SHA-speaking children have minimal information about typical versus atypical speech sound acquisition and phonological patterns

produced by these children due to the lack of Arabic phonological development research in general and the phonological development of SHA in particular.

- Finally, SHA was chosen as it is the native dialect of the researcher, which supports the process of test development, assessment of children, and selection of settings where SHA is the main dialect spoken.

Therefore, the researcher was interested in establishing reference norms for the development of Arabic speech sounds in monolingual SHA-speaking children, which is the ultimate goal of this thesis. In addition, creating an assessment tool that is linguistically and culturally appropriate for assessing and evaluating speech sounds skills in SHA-speaking children is another goal for this study.

This thesis is divided into seven chapters. The first two chapters provide a literature review of the contemporary literature in this research area. Chapter 1 addresses an overview of the theoretical approaches of phonological development, as well as some aspects of assessing phonological development in terms of the tasks and measures that need to be applied and the principles of developing an assessment tool for phonological development based on international construction criteria. Chapter 2 is dedicated to aspects of the Arabic phonological system in general, and Saudi Hejazi phonology in particular, and their acquisition by reviewing previous studies investigating different Arabic dialects.

Chapter 3 outlines the study's research design and method, providing information about the participants and the recruitment process. A detailed explanation of the creation of the assessment tool used in this study (i.e. Saudi Hejazi Arabic Phonology Assessment 'SHAPA') and the scoring and analysis procedure followed in this study are described in this chapter.

The results and discussion of the study findings are presented in Chapters 4, 5, and 6. Chapter 4 provides results of the SHAPA evaluation and a discussion of its appropriateness as a reliable and practical assessment tool for clinical and research use. Chapter 5 presents findings and discussion on phonetic acquisition, including the percentage of consonants and vowels correct (PCC and PVC), the phonetic inventory, and consonantal co-occurrences (i.e. PCCC, and inventories). Chapter 6 describes the phonological variants (PhonVar) found in the speech of this study's sample, including identification of the developmental phonological patterns and infrequent variants (InfrVar), followed by a discussion regarding these findings.

The last chapter (Chapter 7) provides an overall discussion of the study's aims and findings, with an overview of the findings. Practical implications as well as a consideration of the study's strengths and limitations, including future steps, are also discussed in this chapter, followed by a summary and conclusion.

1 Literature Review I – Phonological development: Theories and Assessment

This chapter provides an overview of the contemporary literature on theories/models of phonological development, including the universality versus language-specificity of childhood phonological development. In addition, an overview of the assessment of expressive phonological development in terms of the tasks and measures that should be used to establish normative data for typical speech sound acquisition is also included. The first section reviews theories and approaches reported in the literature to explain phonological acquisition and development in monolingual children. Section 1.2 describes the universal aspects of typical phonological development, as well as highlighting the language-specific ones. The third section of the chapter (Section 1.3) provides a brief description of the tasks and measures used to collect speech samples for assessing speech sound development, as well as an evaluation of these tasks in terms of their purpose, scope, and advantages and disadvantages for determining children's expressive phonological skills. It also describes and critically evaluates the measures used to analyse these speech data. The last section (Section 1.4) presents detailed information about the essential psychometric principles required to develop an assessment tool based on international criteria.

1.1 Theoretical approaches to phonological acquisition

For decades, scholars have attempted to understand how the phonological system develops and is acquired in children. Their explanations and observations for phonology acquisition in children have contributed to the development of various theoretical phonological frameworks. While the pioneering approaches to understand children's speech acquisition was a formal, linguistically driven approach (cf. Jakobson, 1968), other fields of studies e.g. biology, cognitive science/psychology and education, have significantly contributed to the comprehension of the processes of speech acquisition over years, outlining the sophistication of this process during children's development (Davis & Bedore, 2013; Vihman, 2014). Consequently, a multidimensional view of speech acquisition has emerged which is reflected in recent literature, such as the emergentist models (Davis & Bedore, 2013, Vihman, 2014).

The main theoretical approaches to phonological acquisition in monolingual children are described and contrasted in the next section. It is critical to note that the purpose of this study is not to examine or confirm/disprove a specific theory. This review, however, provides a theoretical foundation for the current study data to be utilised as a reference point for the remainder of the literature review. This theoretical background should also justify the chosen tasks, measures, and criteria for developing the assessment instrument. Finally, to discuss the general outcomes of this study.

Several theoretical approaches have been described in the literature concerning the phonological acquisition. However, the dividing line between theories is not always clear, with many approaches borrowing one feature of a theory and incorporating it into another as the knowledge of phonological development has been evolving overtime (Stoel-Gammon & Sosa, 2007). Although phonological acquisition theories are quite diverse in detailing their underlying core assumption, many are conceptually linked, and they can be allocated to one of the following three approaches: **formalist**, **functionalist**, and **emergentist** approach (Davis & Bedore, 2013; Vihman, 2014). Table 1.1 lists the theories proposed under each approach, their main representatives, and their main aspects.

Table 1.1 *Theoretical approaches in phonological acquisition*

Theory's type	Theory's name	Authors and date	Main aspects
Formalist approach	**Structuralist theory**	Jakobson, 1941, 1968	1) Phonological acquisition is a set of mental representations of perceptual contrasts. 2) Phonological acquisition is based on universal and innate features or contrast hierarchies. 3) Discontinuity between babbling and first word periods. 4) Acquisition of phonological oppositions (unmarked contrasts are acquired before marked contrasts) has a universal character.
	Generative theory: Generative/ liner phonology Natural phonology.	Chomsky & Halle, 1968. Stamp, 1969. Smith, 1973.	1) Phonological acquisition involves "an innate endowment knowledge"; allows children to discover their language's structure with relatively little data from adult language. 2) Children are born with a set of innate and universal processes (i.e. phonological processes), and then learn to suppress the processes that do not occur in their languages. 3) Children are operating with adult-like *underlying representation*.
	Nonlinear phonology: Metrical phonology Lexical phonology	Bernhardt & Stemberger, 1998. Liberman, 1975. Liberman & Prince, 1977.	1) Focus on the hierarchical organisation of phonological units (i.e. words, syllables, segments and features) and the relationship among these features. 2) Nonlinear phonology allows for a description of underlying relationships that would allow one level (tier) of a unit to be governed by another (e.g. a segment dominated by a syllable). 3) Still, nonlinear phonology claims that the child has a set of innate universal phonological rules and processes described as confirmed templates that are set following the input language. 4) The child does not play an active part in phonological acquisition, as the process of perception and production utilises a 'passive filter'.

	Prosodic theory	Waterson, 1971 McCarthy & Prince, 1995	1) It is not assumed that the underlying representations are adult-like. 2) In the early phases of a child's development, speech perception is not adult-like, and develops gradually with speech production. 3) Suprasegmental characteristics such as syllable structure and stress patt eformed before segmental features as children attempted to replicate th t salient elements of the speech. 4) Children start talking with a template defined by the default configuration of all prosodic characteristics, and after enough exposure to the ambient language, these features are reset (reorganised), mapping perceived target onto an adult-like, more constrained template.
	Optimality theory[1]	Prince & Smolensky, 2004	1) Language is a system of conflicting universal constraints (i.e. markedness and faithfulness constraints), with the feature (subunit of phoneme) as the basic unit of phonological representation. 2) Speech acquisition proceeds by re-ranking a set of mentally available constraints (i.e. markedness and faithfulness constraints) in a language-specific hierarchy. 3) Observed 'surface' forms in child output and in languages arise from the resolution of conflicts between these ranked grammatical constraints.
Functionalist approaches	**Behaviourist theories**	Mowrer, 1952, 1960 Olmsted, 1966, 1971	1) The general ability to learn language is innate. 2) Contingent reinforcement holds an essential role in phonological acquisition. 3) Phonemes are acquired hierarchically according to the frequency of the occurrence of input and the ease of perception. 4) Language universal aspects should depend on a universal ability of perception, and language-specific phonological variations can be accounted for by the frequency of the input.
	Biological theories: Gestural phonology Self-organizing model Frame-Content Theory	Locke, 1980 Kent, 1992 Brownman & Glodstein, 1992 Davis & MacNeilage, 1995; MacNeilage & Davis,1990	1) Emphasise similarities between the pre-linguistic and linguistic (first word) period and similarities in the babbling repertoire across languages. 2) Emphasise the role of anatomical and motor skills in developing of a phonological system, as well as other types of non-linguistic learning and development. 3) In the biological approach, development is a process in which the child progressively applies available (innate) resources to imitate adult behaviours. 4) Child-parent interactions are an essential aspect of driving the PA process. 5) Learning from the environment is critical to creating precise ambient language regularities.

[1]Some models of Optimality theory are based on the Generative theory principles, and other are holding the emergence approach view (Stemberger & Bernhardt, 1999).

	Cognitive/ Usage-based phonology theory: Neighborhood density Whole-word phonology Linked-attractor model	Ferguson & Farwell, 1975. Menn, 1976 Kemmer & Barlow, 2000. Edwards et al., 2004. Vihman, 2007 Menn, Schmidt, & Nicholas, 2013	1) The child has an active role in the PA and chooses the words to say based on articulatory abilities and the phonology of his ambient language. 2) Babbling is seen as a practice stage for motor activities, which subserve speech production. 3) Exposure to ambient language will account for acquiring the language-specific phonemic system. 4) The child begins to talk at the word (or phrase) level available in his input language, and these early words are learned as phonetic forms. Later, these words are segmented and recognised as the basis of phonemes. 5) The construct of neighbourhood density emphasises linguistic processing and retrieval issues above the level of peripheral perceptual and production system operations, focusing on psycholinguistic processes rather than neural variables supporting cognition. 6) The child's speech is one input source to a developing phonological representation.
Emergentist theory	**Emergence approach**	Davis and Bedore, 2013	1) An emergence view asserts that phonological acquisition is a dynamic and active process. 2) It is a process founded within the operation of a complex physical, neural-cognitive, and social system. 3) The system operates based on shared domain-general resources that support diverse areas of function for humans.

Note: The sources of the "main aspects" column are: (Davis & Bedore, 2013a; Fox, 2000; Grech, 1998; Schwartz, 1992; Stoel-Gammon & Sosa, 2007; Velleman & Vihman, 2002; Vihman, 2014), [2] Alternative optimality theory approach; however, do not assume innate or universal status for constraints. Boersma (1998, as cited in Davis and Bedor, 2013), proposes that the differences in markedness constraints are based on frequency differences in the learner's input, not on innately available markedness or faithfulness constraints.

The following section will begin with a brief overview of phonological acquisition theories and basic phonological units. This will be followed by a discussion concerning potential influential factors on the order of phonological acquisition, focusing on the development of phonological universals and language specifics.

1.1.1 Formalist approaches

Formalist approaches are all rooted in adult phonology, which shows how adult language structure can be related to children's forms (Vihman, 2014). The formalist theories such as those of Jakobson (1968), Chomsky and Halle (1965), Stampe (1969) and Smith (1973) share the basic assumption that language acquisition, particularly phonological acquisition, is driven by a universal set of principles and parameters, rules or constraints. This innate set of abstract linguistic knowledge allows children to discover the structure of their ambient language with little data from adult's input, leading to a production of largely universal acquisition processes (Chomsky & Halle, 1968, as cited in Stoel-Gammon & Sosa, 2007). According to Jakobson (1968), the structural parameters of adult language are very similar across languages and dictate the relative order of phoneme acquisition in children across languages, that is a universal innate order of acquisition. He asserted that the acquisition of the phonemic contrast is based on implicational hierarchies regarding the nature of phonemic inventories in adult languages, unrelated to biological constraints, especially in the earliest stages of phonological learning (Stoel-Gammon & Sosa, 2007); that is, the presence of one phoneme in a language implies the presence of another phoneme. For example, the presence of the fricative /s/ implies the presence of the more basic plosive /t/, just as the acquisition of fricative by a child implies that plosive have been acquired. Therefore, the shared sounds across languages with common contrasts or features are assumed to be acquired before the infrequent, marked features (Davis & Bedore, 2013). According to Stampe (1969) and Smith (1973) children's phonological forms are produced as a result of strictly ordered, obligatory realization rules that are subject to phonetic rules to produce the child's output (Stoel-Gammon & Sosa, 2007). Smith described phonological development as a process of rule modification applied to stable, adult-like representations without evidence for an independent child-based phonological system (Stoel-Gammon & Sosa, 2007). In general, the formalist approaches focus on the phonetic-phonological structure of language without incorporating other aspects of language, child capacities, or cognitive and social function (Davis & Bedore, 2013).

Although formalist approaches almost all agree on the innate mental representations of language knowledge, some of these approaches (e.g. some alternative Optimality Theory perspectives (Prince & Smolensky, 2004), Nonlinear Phonology (Bernhardt & Stemberger, 1998), and Prosodic theory (McCarthy & Prince, 1995)) considered the role of the speech input and the frequency of phonological features in the ambient language to re-rank universal (markedness) constraints so that the output is adjusted according to the ambient language (Davis & Bedore, 2013; Stoel-Gammon & Sosa, 2007). Thus, markedness constraints are replaced by faithfulness constraints, which dictate the output to be similar to the adult form (Stoel-Gammon & Sosa, 2007). Consequently, the cross-linguistic variation is seen as a result of re-ranking these constraints (Davis & Bedore, 2013). On the other hand, some formalist approaches ignore the role of input (e.g. Natural phonology). In the case of Natural phonology, the phonological rules are innately known as 'phonological processes'. These processes are available in every child before any exposure to any language with the same set of natural processes (Vihman, 2014). Then, acquiring language-specific phonology involves language input, which allows learning the constraints a language imposes on these natural processes (Donegan & Stampe, 1979, as cited in Stoel-Gammon & Sosa, 2007). In short, formalist theories appear to be able to explain the linguistic surface patterns that define the mismatches between adult-like and children's speech, as well as the general patterns observed across languages.

Generally, numerous criticisms of formalist approaches have been cited. For instance, Menn, Schmidt, and Nicholas (2013) reported their concerns about using the term 'underlying representation' in the formalist approaches without discussing the concept of 'representation' itself. They questioned how this representation exists, as the assumption that perceptual abilities are fully developed at the onset of linguistic production is seen as problematic (see Menn et al., 2013, p.469 for detailed discussion).

Furthermore, Stoel-Gammon and Sosa (2007) reported that some of these formalist approaches (Structuralism theories and Generative theories in particular) do not have an adequate mechanism for accounting for the fact that phonological development is non-linear, variable (including inter- and intra-child variability), and highly individual. However, other formalist approaches, such as constraint-based approaches (e.g. Optimality Theory), have acknowledged continuity between the prelinguistic to linguistic period by positing constraints that operate during the babbling period and are then carried over to the meaningful speech emergence (Stoel-Gammon & Sosa, 2007). Individual differences and the role of ambient language input are also explained within the optimality theory up to a certain point. However, the origin of these initial constraints still is not clearly specified. Furthermore, formalist approaches mostly neglect the influence of the social and interactional skills of the child necessary to communicate effectively, as well as the child's physical capacities required for a successful acquisition of phonological systems (Davis & Bedore, 2013). Formalists generally believe that the acquisition process is primarily linear, guided by universal rules, and consistent across children, which is frequently not the case (Stoel-Gammon & Sosa, 2007).

1.1.1 Functionalist approaches

Functionalist approaches have in common the conviction that children play an active role in acquiring speech by interacting with their environment. They also consider the role of language input and use in developing and modifying the child's phonological system (Stoel-Gammon & Sosa, 2007). Such modifications allow children to shape their vocal production accordingly (Bybee, 2001; Vihman, 2014). Functionalist approaches provide some evidence to explain individual differences in phonological knowledge acquisition. Individual differences in the acquisition and the various phonological output forms produced within and between children as well as across languages were shown to be considered in functional approaches by considering the multi-dimensional biological resources for speech production (i.e. auditory perceptions, cognitive/comprehension, and articulatory production) (Bybee, 2001; Menn et al., 2013; Vihman, 2014). Therefore, phonological development is not just a matter of change in the system (rules) but may take place on a word-by-word basis reflecting the individual experience and preferences of the child (Stoel-Gammon & Sosa, 2007). In general, within functionalist approaches, acquiring phonological knowledge is viewed as an input-driven process that is required to instantiate neutrally and cognitive knowledge, taking into account the peripheral action system in instantiating phonetic behavioural patterns and growth (Davis & Bedore, 2013).

Despite these commonalities among the various functionalist approaches, the individual theories can be distinguished by their emphasis on the highlighted specialities. Biologically, theories emphasise the role of the physical articulatory/motor speech abilities for speech acquisition while acknowledging the role of non-linguistic learning and development (e.g. the cognitive and social-interactional components) (Stoel-Gammon & Sosa, 2007). Cognitive/usage-based theories focus on the child's role in the phonological acquisition, where he/she plays an active role (i.e. they select words based on their own articulatory abilities, test, and revise hypotheses regarding phonology based on linguistic experience and motor skills), as well as recognise the role of ambient language input and use (including the word frequency and neighbourhood density of individual word) (Menn et al., 2013; Stoel-Gammon & Sosa, 2007). Functionalist approaches generally emphasise self-organisation or language development through the interaction of the perceptual system and the production, speech-motor system, with no appeal to any specifically pre-existing linguistic knowledge (Vihman, 2014). They still differ, along with formalist approaches, in their explanations of the sources of individual differences and the influence of learning on phonological acquisition (Davis & Bedore, 2013a).

Several studies have investigated phonological acquisition among children speaking different languages to investigate whether children develop phonology along a universal path (i.e. following the formalist approach) and play an active role in the acquisition process (i.e. functionalist approach). For example, Locke (1983, as cited in Grech, 1998) reports a summary of investigations from 15 different language environments such as English, Afrikaans, Chinese, Dutch, Arabic and others. He found that babbling of children from varied language backgrounds shared more similarities than variances. Further, Locke found that these children produced more anterior consonants than back ones. However, he reported also language-specific patterns among these children. Many studies on children aged two and up indicated language-specific trends as well as diversity among the children (individual differences) (e.g. Anderson & Smith, 1978, So & Dodd, 1994).

Language exposure and functional load (i.e. the relative importance of each phoneme within a specific system; Hua & Dodd, 2006) were supported by most of the functionalist approaches (e.g. cognitive theory (Ferguson & Farwell, 1975)). However, Dyson and Amayreh (2000) could not justify the appearance of the back fricatives (e.g. /x/ and /ħ/) and the emphatic consonants in the speech of Jordanian Arabic-speaking children aged 2;0 according to the functional loud of these phones in the adult-speakers. They claimed that the frequent inclusion of these sounds in the child-directed speech by adult-Jordanian Arabic speakers and their many phonemic contrasts in that language might explain their early acquisition. Such evidence could prove that even functionalist approaches did not provide a full range of evidence for phonological acquisition considering cross-linguistic variations.

In short, the formalist and functionalist approaches differ fundamentally in which skills and mechanisms are involved in children's acquisition and which internal and external factors may influence the speech outcome, particularly as it relates to the driving forces of speech acquisition (i.e. innate vs learned/emerging), the development of linguistic perception, and the importance of ambient language input in explaining individual variability (Davis & Bedore, 2013).

1.1.1 The emergence approach

The emergentist theory implies a broader view of what must be acquired to master the practical use of phonological knowledge and behaviours. It assumes that child-intrinsic biological, social-interactional capacities and cognitive-neural foundations engage with communication's linguistic and cultural context (Davis & Bedore, 2013). Such a comprehensive view of phonological acquisition synthesises what different research traditions have focused on, including the consideration of underlying knowledge and rules (phonology), the physiological function of speech-related structures and functions (phonetics), or input processing, storage, and retrieval of output (psycholinguistics) (Davis & Bedore, 2013). Therefore, the emergentist approach can be considered a multifaceted approach that incorporates almost all disciplines of functionalist theories and some formalist approaches. It expands them by considering individual internal factors and external environmental factors as a strong interplay during the development of phonological knowledge. According to the emergence approach, speech acquisition research needs to be directed into studying other natural phenomena to test emergence as a potential solution to understand how phonology is acquired by considering the existing paradigms (Davis & Bedore, 2013).

Considering the necessary components for a theoretical model, the emergence approach meets the aspects that Stoel-Gammon and Sosa (2007) and Barlow and Gierut (1999) suggested. The emergentist approach emphasises the possible importance of the phonological system's functioning in characterising the child's acquisition. Considerations of the role of input, for example, are related to the frequency of input of ambient language phonological structures, rather than the function of parent-child interactions in facilitating and optimising the child's function in her environment through a growing phonological output repertoire and understanding of input that is coded using those phonological patterns. Like the formalist and functionalist approaches, the emergent approach describes the mismatch between children’s and adults’ productions, as well as explaining general phonological development observed across languages. It is trying to sufficiently explain individual variation (i.e. inter- and intra- variability) during development, which the majority of formalist approach theories and some of the functionalist theories (e.g. behavioural theory) do not consider. Formalist approaches also ignore the relationship between phonetics and phonology; in contrast, it is considered by the emergentist approach, which also provides predominantly satisfactory explanations for the developmental processes found in children’s speech output, children’s variations from the adult target and its variability across languages and children (Davis & Bedore, 2013).

Moreover, the emergence approach of the phonological theories would enable consideration of across language variations to understand how phonology is acquired. It also encompasses the three components that are recommended by Stoel-Gammon (1991) for a theory of phonological acquisition, i.e., auditory-perceptual, cognitive-linguistic, and neuro-motor-articulatory component. The emergence theory proposed that these three dimensions all are necessary to enable a fully functional phonological system (Davis

& Bedore, 2013). Davis and Bedore (2013) explain how the phonological system is acquired considering these three dimensions. They reported an example of how a child could say "ball", for example:

> '*She needs to retrieve and organize phonological components she has stored in her growing knowledge base. Her respiratory, phonatory, and articulatory systems have to coordinate so that she can talk about "ball" with others. Over time she learns to attach "ball" to all sorts of objects that fit within that semantic category and to use it in a variety of communication contexts'* (p.153).

Therefore, adopting such a comprehensive approach to analyse the phonology in SHA-speaking children would provide a general explanation for children's phonological acquisition considering the ambient language phonology as well as the social function of the language used (i.e. the dialect).

The main aim of phonological acquisition theories is to provide universally valid explanatory concepts for speech acquisition. All theories emphasise the similarity found across languages and highlight language-specific differences. However, theories do not agree on which aspects of speech development are universal, whether they are innate or not, and most importantly, how children begin to learn the phonology of their languages and what units are considered the building blocks of phonological acquisition. To answer this question regarding Arabic phonology, it is important to identify phonological universal and Arabic-language-specific features of the speech sounds acquisition by comparing the phonological development of Arabic to other languages in terms of rate and order of consonantal acquisition, syllable structure, and phonological patterns.

1.2 Aspects of universality and language-specific in the phonology of Arabic-speaking children

The phonological acquisition across languages involves both universal or near-universal and language-specific patterns as well as some individual patterns posed by each child. It is expected that children acquiring Arabic, independent of the dialect, will follow a universal order of phonetic acquisition and overcome their phonological patterns following the universal expectations. However, due to the complex structure of Arabic phonology, language-specific patterns of phonological development are also expected to occur among Arabic-speaking children. Because Arabic dialects differ in phonological features, some patterns are known to be dialect specific. These are considered typical in one dialect but not in other dialects.

Locke (1983) investigated phonological mastery for various languages, including English, German, Japanese, Russian, Italian, Arabic, Slovenian, Swedish, Norwegian, and Czech. He noticed a general trend in the order of acquisition among these children speaking different languages, which reflects a universal development in their phonological acquisition. For example, he found that anterior consonants were acquired earlier in all languages, and plosives seem to be appearing before fricatives while affricates and liquids are acquired later. Locke's findings matched Alqattan's (2015) review that showed that Arabic-speaking children tend to acquire plosives /b, t, d/, and nasals /m, n/ before fricatives /s, z, ʃ/. Alqattan (2015) reported that the order of acquisition of Arabic supported the notion of markedness (Optimality theory, Prince & Smolensky, 2008). Similarly, Dyson and Amayreh (2000) found that the alveolar /r/ tend to be substituted with the lateral [l], like children acquiring other languages that have a similar type of /r/ (e.g. Spanish, Italian, and Turkish).

Similarly, in terms of syllable structure acquisition, it was found that Arabic-speaking children follow almost similar patterns of syllable structures to children speaking English (Ayyad, 2011). Ayyad found that open syllables (e.g. CVV) were acquired before closed syllables (e.g. CVC). For instance, syllables such as CVV were found to be mastered by the age of 4;0 (i.e. by 90% criterion) in Kuwaiti Arabic-speaking children, while it was found that CVC(C) syllables were only acquired with 75% of accuracy by the same age. These findings were also true for children speaking English, according to Ayyad (2011) and Smit (2007, as cited in Ayyad, 2011). Disyllabic and multisyllabic words are common in adult Arabic speakers, as well as in many other languages (e.g. Turkish and Italian); therefore, children speaking Arabic acquired such word structures earlier in their life similar to what was reported for children speaking these languages.

Further, phonological patterns in children speaking Arabic showed some universal patterns that occur in Arabic and other languages such as English (Dodd et al, 2003), German (Fox, 2000) and Italian (Fox-Boyer et al., 2021). For example, fronting, stopping, devoicing, weak-syllable deletion and final consonant deletion were among these universal patterns reported in the speech of children. However, most of these universal patterns were found to be overcome in children's speech with different frequencies of occurrences at different ages across languages. For example, stopping occurs in the speech of Kuwaiti Arabic-speaking children and English-speaking children with a relatively similar frequency at age 2;0-2;5 (Alqattan, 2015; McIntosh & Dodd, 2008).

Many language-specific phonological developmental paths were identified in the Arabic literature, which are due to the Arabic-specific features. For example, some aspects of speech sound acquisition in Arabic-speaking children were related to the ambient-language input, such as the early acquisition of back fricatives (/x, ɣ, ħ/) before the front fricatives (/s, z, ʃ/) (Amayreh & Dyson, 1998). It was also found that Arabic-speaking children acquire voiceless plosives before their voiced counterparts (Alqattan, 2015). Furhtermore, Arabic-speaking children tend to correctly produce back fricatives earlier than front fricatives (Dyson & Amayreh, 2000). These examples provide some evidence that Jakobson's proposals about the universal order of the acquisition, such as the presence of the velar implies the presence of the labial, did not acknowledge cross-linguistic variations. Therefore, a language-specific order of phonetic acquisition is apparent in Arabic-speaking children, and such evidence adds to the limitations of formalist approaches.

Dyson and Amayreh (2000) found that the inclusion of the glottal stop /ʔ/ and the emphatic sounds (/tˤ, dˤ, ðˤ, sˤ/) in the Arabic phonemic inventory could lead to the appearance of some language specific patterns. For example, a glottal stop replacement pattern was reported for children speaking Egyptian and Jordanian Arabic, and de-emphasisation was reported for children speaking Jordanian and Kuwaiti Arabic (Ammar & Morsi, 2006; Dyson & Amayreh, 2000; Ayyad, 2011). Although glottal stop replacement /ʔ/ is one of the Arabic language-specific patterns occurring in the speech of children speaking Egyptian and Jordanian Arabic, it would not be expected to be overcome in their speech if the glottal stop /ʔ/ is replacing the voiced plosive /g/ (e.g. /gamar/-[ʔamar], 'moon'). Such a pattern is common in the adult's speech of Egyptian and Jordanian Arabic, unlike other Arabic dialects that do not allow for such a pattern. Again, these findings support the suggestion for using emergentist theory in analysing Arabic phonology, as the effect of the ambient language, the role of adults' input, and the social context, which were strongly highlighted within this theory, were supported in the Arabic phonological acquisition studies.

Also regarding syllable structure acquisition, Alqattan (2015) found that many children acquiring Kuwaiti-Arabic master the production of words with longer structures (e.g. CVCVC) before children acquiring English. Owaida (2015) also found that the percentage of correct production of multisyllabic words was higher for children developing Syrian Arabic compared to children acquiring English.

As reported in Section 1.1, 'universalist' theories (i.e. formalist), such as those of Jakobson (1968), Locke (1983) and Stampe (1969), propose that children develop phonology along a universal or near-universal path. This theory's proposal was not fully supported in Arabic phonology studies. Other theories that are cognitively based (e.g. functionalist), such as those offered by Ferguson (1978) and supported by Macken and Ferguson (1983), Stemberger (1992) and Waterson (1971), were found to be partially supported in Arabic studies. Still, the effect of the ambient language on phonological acquisition that was shown in the Arabic studies, was not acknowledged within the functionalist theories, which suggest that children play an active role in the acquisition process by formulating and testing hypotheses regarding the sound system being acquired. Instead, the emergentist theory neutralizes the gap between innate capacities (i.e. the universal pattern of development) and the role of the environment (i.e. results in language-specific patterns). This neutralization of boundaries between 'nature and nurture' is the result of s heterarchical and dynamic system enabled by social feedback (e.g. the adult's input of the ambient language) (Devis & Bedore, 2013). Several studies conducted on Arabic-speaking children have found many similarities in segmental phonology between these children and across languages (e.g. Dyson & Amayreh, 2000). This also applies to studies of phonological development in children, where universal path of development was found in the speech of Arabic-speaking children. At the same time, many language-specific phonological development were documented in these studies, emphasising the suggestion of the emergence approach where both paths of development are necessary for the phonological acquisition (Davis & Bedore, 2013).

Therefore, the current study's findings could add to the evidence of language-specific patterns in phonological acquisition. In addition, it helps in describing the trajectory of speech acquisition in typically developing children in terms of the universal, language-specific, or dialectal-specific path of acquisition under the scope of the emergence approach. The strong acknowledgement of the ambient language phonology, interactional instinct and cultural input within the emergentist theory of phonological acquisition provides the opportunity to explain bilingual children's acquisition (Iglesias & Rojas, 2012, as cited in Albrecht, 2017). Consequently, because Arabic phonological development is impacted by the diglossic nature of the spoken language (Hussein, 1980), this gives the children speaking Arabic the 'bilingual' nature of exposure as they are exposed to two (or more) Arabic forms (e.g. MSA and UHA, see Chapter 2, Section 2.1.2). Thus, considering a specific Arabic dialect in this study (i.e. SHA) allowed for examining the effect of the ambient language and the environmental/social context on the speech development of children speaking this dialect and comparing their speech to children speaking other Arabic dialects.

1.3 Tasks and measures

The tasks used by SLTs to collect representative speech data need to include a range of methods; such tasks include single-word naming tasks, elicitation of connected speech (CS) samples, and stimulability tasks to identify a child's functional skills across a broad array of areas (McLeod & Baker, 2014). These measures can be used to describe the child's phonetic inventory and consonantal acquisition, the number and type of errors made, and their speech error patterns compared to normative data for children of their age (Holm, Sanchez, Crosbie, Morgan, & Dodd, 2021). However, there is ongoing debate in the literature concerning the optimal sampling mode for a comprehensive assessment of SSD in children that would lead to efficient and effective treatment (Masterson, Bernhardt, & Hofheinz, 2005). Below a critical evaluation is provided of each task's application and advantages and disadvantages for establishing normative speech data. Following this, a description of different analyses will be presented, including measures for *independent analyses* such as phonetic and phonemic inventories, together with *relational analyses* such as percentage of consonants and vowels correct (PCC and PVC), and measures of phonological variants (PhonVar).

1.3.1 Tasks for speech sample elicitation

Many studies on phonological assessment have relied on three methods of speech elicitation: single word naming tasks, elicitation of CS (Hua & Dodd, 2006), and stimulability testing (de Castro &

Wertzner, 2012). However, Morrison and Shriberg (1992) mentioned that the type of context used to sample each form of elicitation can vary across studies; for example, three types of contexts are used to elicit CS: spontaneous, retelling, and imitation. Each of these methods has advantages and disadvantages as discussed below; however, the appropriateness of a selected speech sampling task depends on the purpose of the assessment (Eisenberg & Hitchcock, 2010) because the speech sample can bias the analysis in a way that fails to elucidate basic phonological feature (Bernhardt & Holdgrafer, 2001a).

1.3.1.1 Single-word naming task

A single-word naming task elicits representative speech data that contains all possible speech sounds of a given language, making it feasible for purposes of comparisons and generalisation (Eisenberg & Hitchcock, 2010; Wolk & Meisler, 1998). SLTs most frequently use this sampling method when assessing SSDs by directing the child to name a certain number of items in pictures spontaneously or through imitation in a single-word naming task (Yeh & Liu, 2021). This method has two main advantages that make it the most popular method among SLTs and researchers for eliciting speech data from children. First, it is an easy method to administer because of its structure, where a predetermined word list assists the test-giver in predicting the child's production. Moreover, the task length is usually short and time-efficient, necessary requirements due to SLTs' time constraints and to maintain the child's attention during the test (Hua, 2006; Tyler & Tolbert, 2002; Wolk & Meisler, 1998). It is also considered an easy task to phonetically transcribe the child's production because of the predetermined Using such a list across multiple clients and area of investigation enhances intra- and inter client reliability, allowing identification of impairment and standard evaluation of outcome over time (Eisenberg & Hitchcock, 2010; Masterson et al., 2005).

The second advantage of the single-word naming task is that it yields consistent, representative speech sample data (Masterson et al., 2005; Wolk & Meisler, 1998; Yeh & Liu, 2021). Such data provides a balanced but not equal frequency of occurrences of phonemes in each possible position as they need to be sampled in relative proportion to their frequency of use in the language under study (Flipsen Jr & Ogiela, 2015). The predetermined word list controls the speech sample so that it is consistent across all test takers, especially if the choice of test items is made based on the selection of words within children's lexicon, which leads to consistent naming of the test items across all participants (Stoel-Gammon & Williams, 2013). It can be constructed to include all word structures and segments (phonemes) in various positions and phonetic contexts of the target language, yielding greater content validity concerning this language (Masterson et al., 2005; Wolk & Meisler, 1998). Supporters of the single word naming task contend that, compared to CS samples, this method of elicitation (a) produces controlled speech data that are feasible for comparison and generalisation, (b) allows for the generation of norm-referenced scores to help make eligibility decisions regarding treatment goals, and (c) guarantees a sample of all of the relevant phonemes in a language (Flipsen Jr & Ogiela, 2015; Hua, 2006).

The major weakness of the single-word naming task is that it may overestimate the child's actual abilities because its design could result in elicitation of a limited type and range of words (Eisenberg & Hitchcock, 2010; Wolk & Meisler, 1998). Sometimes single-word naming tests do not sample all the speech sounds of the language under study in all possible positions, especially sounds more prone to error, which leads to the collection of a speech sample with limited phonetic contexts. Skahan, Watson, and Lof (2007) argued that published single-word naming tests often do not yield sufficient opportunity for the productions of phonemes in a variety of phonetic contexts, which could lead to incomplete data about the children’s speech. The words usually selected for such a test have simple phonotactic shapes (e.g. monosyllabic words) (Eisenberg & Hitchcock, 2010; James, 2001a), and such word shapes fail to reflect children’s performance in real-life communication (Morrison & Shriberg, 1992; Wolk & Meisler, 1998). Lastly, the predetermined word list in the single-word naming task may not be familiar to children, which may negatively impact their performance and consequently limit the speech sample collected (Yeh & Liu, 2021).

To avoid incomplete speech sample data, it has been recommended to include words with different shapes and lengths, such as multisyllabic words, as such words could provide opportunities to examine children’s productions of different phonological variables (e.g. non-final weak syllables, within-word consonantal sequences at syllable boundaries, and two-footed words) (James, Ferguson, & Butcher, 2016). Furthermore, transcribing the whole production of a child rather than just the target sounds may maximise the information obtained from the test and provide more data about segment production variability (James, 2001b; Klein, 1984). It has also been recommended to sample each consonantal and vocalic phoneme repeatedly in each word position, where appropriate, in various phonetic contexts to better understand features of normal speech development. Examining a single targeted production of speech sounds may not be sufficient to reflect the variability of production that characterises disordered speech (Eisenberg & Hitchcock, 2010; James, 2001a; Klein, 1984). Wolk and Meisler (1998) stated that a well-designed single-word naming task could tap into the child's phonological system more deeply and give the clinicians/researchers the maximum control over the speech sample, resulting in a more comprehensive set of data.

1.3.1.2 Connected speech

Connected speech (CS) samples are used to elicit speech data from a child that reflect his/her performance in real-life natural communication. There are differences in application for this form of elicitation. Morrison and Shriberg (1992) explained that there were differences in method in many studies that reported using CS samples, included differences in the purpose for the child talking (e.g. naming, repeating, or informing), the mode of talking (e.g. labelling, telling/retelling, and describing), the level of spontaneity (e.g. spontaneous, evoked, and imitative), and the situational support for talking (e.g. object, picture, or events). Such differences in task design may lead to different results when analysing the data collected. However, most studies have used spontaneous conversational speech (e.g. Wolk & Meisler, 1998).

The most-reported strength of CS samples is the availability of various phonetic contexts in the child's spontaneous speech sample, which are thought to be important in phonological assessment (Wolk & Meisler, 1998). CS provides information about the typical production of a child in terms of lexicon, phonology, syntax, and pragmatic interaction and allows for integrated analysis of the phonological skills of a child (Morrison & Shriberg, 1992; Stoel-Gammon & Williams, 2013). Prosodic factors (such as rate, rhythm, intonation, pitch, and loudness) are also reported to be evaluated only in CS samples (Bernhardt & Holdgrafer, 2001b). Masterson et al. (2005) explained that CS samples, conversational one in particular, are potentially representative of a child's everyday speech, which increases their ecological validity and provides a valid basis for planning effective and efficient treatment. Fabiano Smith (2019) recommended that CS samples need to be supplemented by a single-word naming task to ensure that sufficient opportunities are provided to produce each sound of a language in order to develop the phonetic inventory. Furthermore, CS samples could provide SLTs with more details about the specific phonological patterns in children's speech that have been reported to appear only in CS samples (Masterson et al., 2005; Yeh & Liu, 2021).

Despite the advantages of CS samples, this sample type has several disadvantages that may result in SLTs relying on data from single-word tests to analyse speech production and determine treatment goals. First, the time required for phonetic transcription of CS samples is usually more than that available in most clinical settings. Masterson et al. (2005) estimated that conversational samples take three times longer to transcribe than single-word naming tasks. Second, transcription of conversational samples may not be possible for highly unintelligible children since the target words may not be identified or predicted, which reduces the content validity of this mode of elicitation. In addition, collecting the recommended number of intelligible words in a speech sample, which should be between 50 and 100 words (Morrison & Shriberg, 1992), may take longer than the time needed to administer and score a single-word naming task, especially if a child is reluctant to talk. Stoel-Gammon and Williams (2013) added two other difficulties with such a task: the lack of comparability over time and across samples because the data are not based on a single measure like in a single-word naming task. According to Morris, Wilcox, and Schooling (1995), CS samples can effectively obtain the necessary speech sound information, but the instability of a child's conversation over time can hinder a speech assessment that only uses this type of measure.

In a conversational sample, the child has control over his/her CS and may avoid some difficult words/sounds (Morrison & Shriberg, 1992), which leads to a restricted range of word shapes and sounds being produced (i.e. restricted phonetic contexts) (Eisenberg & Hitchcock, 2010). This is because children tend to avoid words that contain elements that are absent from their inventory, instead producing words with familiar sounds and shapes (Yeh & Liu, 2021). However, SLTs can, up to a certain point, control the elicitation of a conversational speech sample by centring the conversation around a common set of objects, similar to Wolk and Meisler's (1998) method, thus increasing the reliability of their comparisons. In the same vein, Yeh and Liu (2021) found that elicitation of CS using a story retelling picture description task may produce a sample that reduces the linguistic processing demand required for producing CS compared to the conversational tasks used in the literature. For example, in their study they found that most of the speech sample in the SSD group elicited by CS consisted of single words. This indicates that using a guided CS sample, such as a picture description task, could have similar reliability to the result of single-word naming tasks.

While the majority of studies have reported strengths and weaknesses for both modes of speech elicitation (single-word naming and CS samples), according to Wolk and Meisler (1998) and Masterson et al. (2005), both modes of elicitation result in the identification of the same potential treatment targets. Their findings indicated that task differences did not appear to relate to phonological profile differences. In contrast, the recent study by Yeh and Liu (2021) found that single-word samples were more informative in detecting significant differences between typically developing children and children with SSD on all the measures they used: intelligibility, speech accuracy, phonemic inventory, and the phonological patterns. An extensive single-word naming task may provide an adequate sample of phonological behaviour and at the same time avoid problems associated with the collection and transcription of CS (Eisenberg & Hitchcock, 2010; Wolk & Meisler, 1998). Edwards and Beckman (2008) clarified that because of the drawbacks involved in CS samples, the primary data for most studies on consonantal acquisition are elicited using single-word naming tasks rather than CS. McLeod (2007), for example, reports on speech sound acquisition in 36 different languages or dialects. Based on 37 studies, 16 of these languages provides data about the age of acquisition of consonants. Seventy-five per cent of these studies examined speech samples based on single-word production only, an additional 14% examined a combination of single–words and CS, and only 11% examined CS alone. Smit (1987) clarified that the two kinds of normative data elicited by single-word naming or CS samples provide different kinds of information needed by clinicians, but that they complement each other.

As is clear from this review, CS has been characterised as a time-consuming task. It does not allow the prediction of the targeted items and yields speech data that are not comparable across children. On the other hand, a well–designed single-word naming task that captures different phonological behaviours among participants would have a high degree of inter–participant reliability and adequate content validity concerning the target language (Masterson et al., 2005). According to Tyler and Tolbert (2002), it is recommended that SLTs collect speech sound data using both modes of elicitation to increase assessment efficiency. However, Morrison and Shriberg (1992) reported that the large–scale studies of typically developing children as well as children with SSD most often used single-word naming tasks in the form of picture naming to collect their speech samples. Furthermore, Yeh and Liu (2021) asserted that a single-word naming sample proved to be a more informative than a CS sample in differentiating children with SSD from typically developing children.

Given the above, the current study adopted a single-word naming test to collect the speech sample from typically developing SHA-speaking children. The aim was to obtain normative data on phonological acquisition for the Hejazi population, where the collected speech sample needed to be comparable and generalisable. At the same time, this type of task was chosen to control the collected speech sample so that it included all phonological characteristics of the language and dialect under study. The guidelines and criteria adopted for developing this single-word naming task are presented and discussed in Section 1.4.1.

1.3.1.3 Stimulability task: Phone imitation

Stimulability is one of the fundamental tasks required for a comprehensive assessment of phonological development in children, providing supplementary data about the nature of speech problems (McLeod & Baker, 2014; Tyler & Tolbert, 2002). Different linguistic environments can be included in stimulability tasks, such sounds being elicited in isolation or in syllables, words, sentences, or CS (Glaspey & Stoel-Gammon, 2007). Phone imitation is part of a stimulability examination designed to examine the child’s ability to immediately imitate, in isolation, a sound (phone) absents from his/her phonetic inventory following an examiner’s model (Miccio, Elbert, & Forrest, 1999; Powell & Miccio, 1996). Glaspey and Stoel-Gammon (2007) described different cues usually used by clinicians when applying stimulability tasks. These cues might include placement instructions, a verbal and/or visual model, or a tactile model. They explained that verbal and visual cues are often used, where children are instructed to watch the clinician’s face when verbal cues are presented.

The purpose of stimulability testing is to predict the prognosis of SSD, diagnose SSD, or select treatment targets and plans (Glaspey & Stoel-Gammon, 2007). It is a possible indicator of the child’s physiological readiness to learn a new sound, even sounds outside those of their ambient language (Dodd et al., 2003; Miccio, Elbert, & Forrest, 1999). Assessing the stimulability inventory was the most-reported independent analysis used by the 39 informal measures reviewed by Limbrick et al. (2013). It was used to assess children's ability to imitate speech sounds without comparing their production to that of adults' targets. This method can also identify possible difficulties in the articulatory production of missing sounds from the phonetic inventory (de Castro & Wertzner, 2012).

SLTs usually conduct stimulability tasks during the speech sound assessment to detect whether children can produce sounds that were found to be absent from their phonetic inventory after an imitation task. However, the inclusion of such a task as another measurement for establishing the phonetic inventory of typically developed children has rarely been reported (Powell & Miccio, 1996; Storkel, 2019). Some researchers, who conducted stimulability tests when assessing children with and without SSD, found that the phone imitation skill is linked to typical speech acquisition and provided evidence of articulatory competence and intact oro-motor speech processing (de Castro & Wertzner, 2012). de Castro and Wertzner (2012) clarified that this is because the ability to produce a sound is dependent on the development of oral-motor control, which involves a synergy between the lips and jaw movements during the vocal tract constriction. Such a task may help to predict the occurrence of speech sounds in a child's phonetic-phonological system and indicate the typical development of speech sounds among school-aged children (Miccio et al., 1999; Powell & Miccio, 1996).

When assessing children with SSD, stimulability test scores could differentiate between those sounds that can be imitated in isolation and therefore are likely to be acquired without intervention, and those sounds that cannot be accurately imitated and thus need to be targeted in treatment (Miccio et al., 1999). The sounds that are stimulable and can be imitated correctly are more likely to be generalised and acquired even without intervention, as imitation is a prerequisite to generalisation (Powell & Miccio, 1996). In their study, de Castro and Wertzner (2012) concluded that the difficulty in producing sounds in the stimulability task reflects difficulty with the phonological representation of these sounds. They found that children who had more absent sounds (i.e. not stimulable) were children who had severe SSD. Thus, applying a stimulability task to diagnose SSDs is helpful since it provides information about the linguistic organisation and oro-motor speech skills that need to be addressed during intervention.

Eisenberg and Hitchcock (2010) claimed that some tests that are designed to differentiate between children with and without SSD might not be appropriate for forming a phonetic inventory because some test content (i.e. types and range) may be unsuitable for determining which sounds a child is and is not capable of producing. They explained that words with difficult phonetic contexts might prevent a child from producing a consonant or vowel that the child is able to produce in simpler words. Thus, it could be argued that including a stimulability task in addition to single-word naming test would support SLTs in completing a phonetic inventory analysis. Storkel (2019) recommended that multiple measures need to be applied when establishing speech sounds normative developmental data because speech sound acquisition is too complex to be reduced to a single data point to represent typical development. A stimulability task is one of these recommended measures, as such information could provide more details about the speech sound acquisition and assist in establishing the child's phonetic inventory (Dodd et al., 2003; Hua, 2006). Applying such a task on only sounds found to be absent from a child's phonetic inventory in a single-word naming task may not reveal a complete picture of the child's phonology, but it could provide information if these sounds are emerging (which would be the case if they could be imitated) or weather these sounds are absent from the child's speech sounds system (Storkel, 2019). Sometimes a child will produce a sound in a single-word naming task unconsciously (i.e. under the influence of contextual factors) (Klein, 1984), while he/she may not correctly imitate it during the stimulability task. Powell and Miccio (1996) explained that poor stimulability does not necessarily indicate that the motoric movements (i.e. the articulator skills) are not under the child's control. It could be related to poor sensory input skills, poor linguistic conceptualisation, or other interfering behaviours.

This review has mostly presented the advantages of the stimulability task; however, there is not much reported in the literature regarding the disadvantages of this task. In addition, it is not common to apply stimulability testing to compile a phonetic inventory in studies conducted to establish phonological normative data. However, Eisenberg and Hitchcock (2010) suggested that sounds produced via auditory or visual stimulation should be excluded from a child's phonetic inventory. Moreover, Morris et al. (1995) explained that imitated responses could overestimate the speech behaviour of a child as such responses mirror adults' productions. It can be argued that since the phonetic inventory considers children's phonetic ability to articulate the sound (Hua, 2006) independently (Stackhouse & Wells, 1997); then including imitated responses (such as the responses of a stimulability task) will not affect the quality of the analysis results. In short, the limitations of such a task in normative studies are still unknown.

When applying the stimulability task to establish normative data, researchers are usually interested in assessing all sounds of the targeted language. In contrast, when they clinically evaluate a child's speech sounds, this task usually targets the missing sounds from his/her phonetic inventory. Given the discriminative ability of this task to distinguish between typical and atypical speech sound development (Miccio et al., 1999), it seems to be a valuable supplement when assessing speech sound acquisition in monolingual children. It provides a complete picture of the phonetic inventory when integrated with other modes of eliciting speech samples (e.g. single-word naming task).

1.3.2 Types of analysis of speech sample

Many studies on phonological development and assessment have reported using different models of analysis and analytical frameworks. Independent and relational analysis are the two modes of analysis available for SLTs to analyse the collected speech data in order to diagnose and treat SSDs (Stoel Gammon & Dunn, 1985). The following sections evaluate both types of analysis and describe their tasks.

1.3.2.1 Independent analysis measures

Independent analyses of a child's speech aim to describe the child's phonology without considering whether his/her production is correct relative to the adult target form (Stokes, Klee, Carson, & Carson, 2005). Such an analysis provides the researcher with information about a child's ability in terms of speech production (i.e. consonants, vowels, and word shapes) without comparison to any previously existing model (Stoel-Gammon & Dunn, 1985). Phonetic and phonemic inventories are examples of independent analysis reported in the literature to assess and describe the speech of typically developing children and children with SSD (e.g. Dyson, 1988; Gierut, Simmerman, & Neumann, 1994). Although a phonemic inventory is usually reported as a type of relational analysis (e.g. Dodd et al., 2003; Fox, 2006), Gierut et al. (1994, 1996) described phonemic inventories using an independent analysis of a child's sound system. They reported that obtaining a phonemic inventory requires a large-scale analysis of each child's production. Thus, they proposed analysing children's usage of minimal pairs to identify each phone's contrastive and consistent function needed to understand the phonemic system. Their definition of minimal pairs was based on the child's production rather than adult target forms, which means using independent analysis for compiling the phonemic inventory. Gierut and colleagues give the following example for their definition of the minimal pairs: "a child's production of [did] and [vid] is considered as a minimal pair, even where the adult targets' 'them' and 'read' are not" (p.297). Therefore, such a complex analysis, which requires a description of each child's sound system, may make it challenging to obtain a phonemic inventory for a normative sample that usually includes a large number of typically developing children.

1.3.2.1.1 Phonetic inventory versus phonemic inventory

The *phonetic inventory* is one of the most fundamental independent measures frequently used by SLTs to describe children's phonological system (Dyson, 1988). Skahan, Watson, and Lof (2007) reported that most SLTs in the United States use phonetic inventories as part of their phonological assessment. A phonetic inventory is a catalogue that lists all the sounds a child can physically articulate regardless of whether the sound is used correctly (McAllister-Byun & Rose, 2016; Powell & Miccio, 1996), which is determined by articulatory and motor skills (Priester, Post, & Goorhuis-Brouwer, 2011). Analysing this phonetic repertoire yields the sounds and features that are available to the child to form words (Stokes et al., 2005). Thus, it includes speech sounds used correctly and incorrectly by the child. For example, suppose a child fails to produce the target /s/ and realises it as [t] all the time but uses [s] in place of the target /ʃ/. In those cases, this child can phonetically produce /s/, but his/her phonemic (phonological) knowledge of this sound /s/ is not yet complete.

Determining a typical phonetic inventory profile in preschool children would describe the speech sound system by identifying the phones the child can make and the set of sounds that are never produced by the child (Powell & Miccio, 1996; Stokes et al., 2005). Therefore, establishing normative data for the phonetic inventory is essential to provide standardised results data accumulated from various typically developing children speaking the same language (Dyson, 1988). This means that the phonetic inventory could be analysed either following relational analysis by comparing child's production to the adult's model, or by independent analysis describing the child's sound productions in detail by their place, manner, voicing, syllable shapes, and any constraint of the syllables and words.

Several studies have reported data about the phonetic inventory using independent analysis to examine the speech of English-speaking children (e.g. Dinnsen, Chin, Elbert, & Powell, 1990; Dyson, 1988; Stoel-Gammon, 1985). These studies have shown that certain sound classes predominate the phonetic inventory of early meaningful speech. For example, Stoel-Gammon (1985) found that the specific sound classes that predominate a typically developing two-year-old child’s phonetic inventory are stops, nasals, and glides, and these sounds are included in the phonemic inventories of older children (e.g. those aged three years). By age 36 months, Stoel-Gammon and Sosa (2007) found that the typical phonetic inventory of English-speaking children expanded to include consonants from nearly all place and manner classes. For phonetic inventory reported for Arabic-speaking children, Amayreh and Dyson (2000) found that the phonetic inventory of two-year-old Jordanian Arabic-speaking children shared many characteristics with those reported for English. However, some unexpected consonants appeared in the inventory of these children, such as the lateral /l/ and the back fricatives /x/ and /ħ/, which is evidence that rejects the universal path proposed by Jakobson (1968). They explained such differences as reflecting the influence of the ambient language.

Some researchers have included an accuracy criterion for attributing sounds to a child’s inventory, which adds a relational component to the inventory analysis (Eisenberg & Hitchcock, 2010). Thus, the phonetic inventory could be compiled from speech data elicited by production of a sound in isolation using a stimulability task, in words using a single-word naming test, or in a CS sample (Dodd, Holm, Hua, & Crosbie, 2003; Skahan et al., 2007; Stokes et al., 2005). For a sound to be credited as existing in a child’s phonetic inventory and considered as acquired, many normative studies of single-word production have used different criteria. Some counted the production of the target phone either once (e.g. Dodd et al., 2003; Gangji, Pascoe, & Smouse, 2015; Maphalala et al., 2014), twice, which is the most frequent criterion used in the literature, (e.g. Clausen & Fox-Boyer, 2017; Dinnsen et al., 1990; Dyson, 1988; Eisenberg & Hitchcock, 2010; Smit, Hand, Freilinger, Bernthal, & Bird, 1990; Stoel-Gammon, 1987), or three times in three different positions (Templin, 1957 as cited in Eisenberg & Hitchcock, 2010). Eisenberg and Hitchcock (2010) argued that children who have acquired a sound but limit its occurrence across word positions would not be credited as having that sound in their phonetic inventory if their sound production is tested in only two- or three-word positions. Therefore, for each sound to be counted in the phonetic inventory, they recommended that the speech sample must contain sufficient occurrences of consonants and vowels with at least two opportunities in all permissible word/syllable positions for the production of each sound. However, there is no research evidence on such a cut-off criterion (Eisenberg & Hitchcock, 2010).

Regarding phonetic acquisition at the group level, many studies have reported different criteria for the percentage of children in an age group who produce a sound correctly. Some studies reported that they using three criteria: age of acquisition (75% of children in an age group), age of mastery (90%), or age of customary production (50%) (Priester et al., 2011); others reported two criteria: 75% for acquisition and 90% for mastery (e.g. Clausen & Fox-Boyer, 2017; Fox, 2006); while others reported only one: 90% of children in an age group (e.g. Dodd et al., 2003). The choice of 90% has a clinical basis because estimates of developmental speech disorders are generally fall between 3% and 10% of the normal English-speaking preschool population (Dodd et al., 2003). Similarly, Smit et al. (1990) supplied some validation for the 90% criterion. They noted that the children's performance tended to plateau after the 90% criterion was reached. This means that if a child does not acquire a sound by the age where 90% of responses are judged as acceptable, then it seems unlikely that the child will acquire this sound without treatment (Storkel, 2019). On the other hand, the two criteria of 75% and 90% have been chosen by some researchers for the purpose of cross-linguistic comparison (Hua, 2006). Using two levels of criteria is in agreement with Sander's (1972) suggestion of using age ranges (i.e. 90% for the upper age limit and 50% for customary production) to identify the acquisition age for a phone. For example, if a sound is reported to be acquired at 3;0–5;0 using 50% and 90% criteria. This means that at least 50% of three-year-old are able to produce this sound, but it is not until children are five years old that 90% or more of that group can produce this sound. Hua (2006) clarified that these criteria are all arbitrary and chosen to achieve the need to set a minimum percentage for the number of children in a particular age group who produce a sound correctly.

Storkel (2019) asserted that the diagnostic accuracy of these different acquisition criteria (e.g. 50% versus 75% versus 90%) had not decided. Thus, the effect of selecting different cut-offs for finding a proper age of acquisition is also unknown. Storkel (2019) noted that any selected cut-off is arbitrary, similar to selecting a cut-off score on a standardised test without bearing in mind the ability of that cut-off score to discriminate between normally developing children and children with an impairment. While these different acquisition criteria are also intended to be used to discriminate between typically versus atypically developing children, Plante and Vance (1994) argued that these arbitrarily derived cut-off scores on a standardised test could lead to the misidentification of children with an impairment as typically developing children, or vice versa, because of a lack of diagnostic accuracy.

On the other hand, *phonemic inventory* refers to speech sound use, which is determined by linguistic experience and the organisation of the speech sound system (Priester, Post, & Goorhuis-Brouwer, 2011). It lists the speech sounds that a child can use correctly in their speech and provides an essential measure of productive phonological knowledge (Child Speech Disorder Research Network, 2017). A phonemic analysis aims to show not just the sounds produced by a child (i.e. phonetic inventory), but the sounds used contrastively, and which enable the child to make distinctions between and among various words (McAllister-Byun & Rose, 2016; Stokes et al., 2005). This analysis also holds information about sound contrasts, or the minor differences in speech sounds, which enable the child to make distinctions between them (Stokes et al., 2005). According to Stokes et al. (2005), figuring out the phonemic contrastive profile of typically developing two-year-old children using relational analysis would provide a metric description of the contrastive feature systems of children with phonological disorders based on the performance of these typically developing children. Therefore, identifying the phonemic inventory of typically developing children is essential for the knowledge of the developmental hierarchy of contrastive phonemic features. This leads to a better classification system of phonological deviant variants to show how delayed or disordered a child is achieving meaningful contrasts (Stokes et al., 2005).

Concerning developing descriptions of phonological systems, Gierut et al. (1994) stated that most previous research emphasised descriptions of the phonetic inventory of children while excluding the phonemic system due to methodological difficulties associated with obtaining appropriate data about phonemic status. One of the methodological difficulties is that sounds have to occur as minimal or near-minimal pairs to be considered as phonemes, which is the main speech unit used to establish phonemic contrasts according to Ferguson and Farwell (1975). Therefore, obtaining a phonemic inventory using a conversational speech sample may not be optimal since children tend to avoid using minimal pairs in their CS (Gierut et al., 1994). Furthermore, obtaining a phonemic inventory requires analysis of each child's production and examination of whether he/she is using the phoneme function contrastively and consistently. Gierut et al. (1994) and Stokes et al. (2005) concluded that the best study design for identifying the phonemic system is a longitudinal study that better understands the phonemic acquisition sequence and underscores the range of individual differences.

Ingram (1986; as cited in Grech, 1998) proposed a possible concept that could be used to determine the acquisition of contrasts, which is 'the definition of the phonological substitution'. Grech (1998) reported

Ingram's substitution definition, which means that a sound is considered a phoneme if used frequently, occurring in minimal or near minimal pairs, and assigns a specific contrast on analysing a child's substitution patterns. Therefore, when a child uses a phone to signify contrasts (i.e. differences in meaning), this phone could be counted as a phoneme and then included in the phonemic inventory of that child (Grech, 1998).

In terms of compiling a phonemic inventory, Stokes et al. (2005), who adopted the criteria for phonological assessment used by Ingram and Ingram (2001), noted that a child has to correctly produce a consonant at least twice in a specific word position in two different words, with at least 50% accuracy for that segment, in order for that consonant to be included in the child's inventory. Using this criterion also requires counting the number of opportunities to produce each phoneme in the speech sample. Such a calculation requires identifying the percentage phoneme used by each child. For example, if initial /k/ occurs only twice in a speech sample (i.e. minimal occurrences of a sound), and one production is correct and the other incorrect, then the segment /k/ would not be included in the phonemic inventory for the initial word position as the correct production occurred only once. Similarly, if initial /k/ is produced correctly in two out of five opportunities in a speech sample, then it would meet the required criterion. However, the accuracy rate is still less than 50% and thus /k/ would not be counted in this child's phonemic inventory (Stokes et al., 2005). To judge the contrastive feature of these phonemes, the child has to use two contrastive sounds in two different words. For example, the presence of initial /p/ and initial /b/ would constitute a voicing contrast (Goad & Ingram, 1987; as cited in Stokes et al., 2005). Fox (2006) clarified that a phoneme is included in the inventory if it is produced by each child at least two out of three times correctly in the correct word positions, similar to the criterion reported by Smit, Hand, Freilinger, Bernthal, and Bird (1990). At the group level, two criteria were used by Fox: when 75% (for acquisition) and 90% (for mastery) of children within one age group were able to use a phoneme correctly in its correct position. These criteria have been reported by many studies where the acquisition of a certain sound was determined when 75% or more of the children in an age group used this sound correctly (Priester et al., 2011).

It is clear from this review that a description of both the phonetic and phonemic structure of speech sounds is important to provide a comprehensive description of phonological acquisition in children speaking a specific language. However, because large-scale analyses are needed to obtain a proper phonemic contrast inventory, where children need to produce minimal pairs to identify each phone's contrastive and consistent function needs to be analysed to understand the phonemic system; this study adopted only a phonetic inventory in its analysis. Such information is still missing for many Arabic-speaking children, especially SHA-speaking children. Phonemic acquisition is described in this study using a relational framework that analyses the phonological patterns.

1.3.2.2 Relational analysis measures

In a relational analysis, the child's production is compared with the adult form (Stokes et al., 2005). It provides the researcher with information on how the child's speech compares with the adult or target phonology (Baker, 2004). It can also focus on incorrect productions by comparing the adult and child pronunciations and describing the differences in terms of segments, features, or phonological processes (Stoel-Gammon, 1985). Within relational analysis, the percentage of consonants, vowels, and phones correct, and examination of phonological patterns are two types of measurement that based on this type of analysis.

1.3.2.2.1 Percentage of consonants, vowels, and phones correct

One of the most common relational analyses measures used to evaluate children's phonological abilities is the *percentage of consonants correct* (PCC) introduced by Shriberg and Kwiatkowski (1982). PCC is a metric that expresses the percentage of intended consonantal sounds articulated correctly in a conversational sample with a minimum of 50 utterances (Shriberg & Kwiatkowski, 1982). It is used to estimate the severity of a child's SSD (severity of involvement) by dividing the number of correct consonants by the total number of consonants in a CS sample. It is noteworthy that this metric was validated using a conversational sample for the speech data. However, it has been reported that for clinical use and efficiency, many researchers have applied PCC to represent the percentage of correct consonants in a well-designed single-word naming test (Fabiano-Smith, 2019; Shriberg, Austin, Lewis, McSweeny, & Wilson, 1997a). However, Masterson, Bernhardt, and Hofheinz (2005) found that the average PCC score associated with a conversational task was significantly higher than the average PCC score for a single-word naming task. Shriberg et al. (1997a) clarified that the proportional distribution of intended consonants in single-word naming samples differs from the distribution of consonants in a conversational speech sample; therefore, PCCs based on different sampling contexts are not directly comparable. However, Masterson, Bernhardt, and Hofheinz (2005) claimed that this result might be related to the structures of the single-word samples compared to spontaneous conversation.

The PCC metric originally developed by Shriberg and Kwiatkowski (1982) has some limitations. For example, one of the concerns with the PCC metric is that the conversational speech sample collected from reluctant talkers may not be linguistically rich. The speech rate of these children may be too slow or unintelligible for the examiner. Shriberg et al. (1997a) suggested using appropriate pictures to elicit speech samples from children who are not good candidates for conversational speech samples. They also addressed a concern regarding phoneme distortions being weighted similarly to omission and substitutions errors in the PCC score. Shriberg (1993) described consonantal distortion as being biologically and cognitive-linguistically more mature articulation errors because they reflect a transient or persistent difficulty in representing allophonic detail and/or with sensory–motor aspects of articulatory precision. In his previous study, Shriberg found that many of the young children he tested expressed a sufficient proportion of common distortions errors, which yielded an overlap between the standard deviation of the PCC score of typically developing children and the score of children with speech delay. Such an overlap places typically developing children at risk of being wrongly identified as having a SSD, and grants children with delayed phonological development the same priority for intervention as those with disordered development (Dodd, 2014).

Shriberg et al. (1997a) made changes to the PCC metric to address distortion errors. They proposed two metrics: percentage of consonants correct–adjusted (PCC-A) and percentage of consonants correct-revised (PCC-R). The adjusted score of the PCC (PCC-A) ignores all consonant allophones (i.e. common clinical distortion errors) that are considered phonemically correct (Shriberg et al., 1997). Examples of clinically common distortions are dentalisation (e.g. /s/, /z→ [s̪, z̪]), and lateralization (e.g. /ʃ/, /ʒ/→[ʃˡ, ʒˡ]). PCC-A has been reported to be used in studies focused on phonology rather than phonetic variants (Clausen & Fox-Boyer, 2017). In the revised PCC (PCC-R), common and uncommon clinical consonant distortions are considered correct. For example, the palatalization of / ʃ/ →[ʃʲ], an example of uncommon clinical distortion, is counted as a correct production using the PCC-R metric. Therefore, PCC-R is considered a measure of phonemic accuracy in which allophonic errors (i.e. consonant distortions) are scored as correct. In contrast, all deletions and substitutions are scored as errors (Campbell, Dollaghan, Janosky, & Adelson, 2007). In other words, the PCC-R metric counted all clinical distortions that used to be transcribed by SLTs using diacritics (e.g. dentalized [̪], lateralized [ˡ]) and uncommon clinical distortions (e.g. palatalized [ʲ], lengthened [ː]) as correct, and these are differentiated from the substitution and deletion errors (e.g. substituting /r/ with [l] or [w], or delete it) which were counted as errors (Shriberg et al., 1997a). Therefore, the PCC-R metric has been described as the best measure of articulatory competence in three– to eight–year–old children (Shriberg, Austin, Lewis, McSweeny, & Wilson, 1997b). The discrepancies in PCC-A and PCC-R scores represent the proportion of all distortion errors that are unusual clinical distortions (Shriberg, et al., 1997a).

Another concern addressed by Shriberg et al. (1997a) was about the representativeness of the PCC regarding articulatory competence given that it only includes consonants. To address the representativeness concern, Shriberg (1993) added the percentage of vowels/diphthongs correct (PVC) metric to his speech profile series to provide additional information about the articulation of vowels and diphthongs of American English. Like the PCC-R, the PVC-R is calculated with all vowel distortions (e.g. [ɝ̜], where [̜] indicates a derhotacised r–coloured vowel) being considered correct (Shriberg et al., 1997a).

None of the above–described metrics describe the total index of all phonemes (i.e. combined consonants and vowels together). To meet this need, Dollaghan, Biber, and Campbell (1993) proposed a percentage of phones correct (PPC) measures to reflect the percentage of consonants and vowels/diphthongs articulated correctly in responses to a nonsense-word reptation task used in their research on assessing phonological memory. However, the PPC metric has also been used to calculate the percentage of phones correct in responses to single-word naming tasks (e.g. Dodd et al., 2003). Like the PCC-R calculation, the PPC-R also scores both correctly articulated speech sounds and distorted-speech-sounds as correct. In general, the PPC metric is recommended to be calculated for children presenting with difficulties across both the consonantal and vocalic systems (Child Speech Disorder Research Network, 2017).

To summarise, selecting an appropriate measure from the above–mentioned metrics depends on the assessment need and research question. Shriberg et al. (1997a) suggested some conditions for the proper application for each metric. For example, the PCC is appropriate when all targeted children are between three and six years old and have speech delays. However, when comparing children of diverse ages and diverse speech skills, the PCC-R is the most appropriate metric. It is also more sensitive to differences in phonological involvement than the PPC-R, mainly because both typically developing children and children with speech delay use a high proportion of correct vowels (Shriberg et al., 1997a). Therefore, in the current study the PCC, PCC-R, and PVC were only included in the analysis of the collected speech data.

1.3.2.2.2 Phonological variants: Infrequent variants and phonological patterns

Phonological variants (PhonVar), a term adopted from Albrecht (2017), means that children's speech productions are deviant from the adult target. According to Albrecht (2017), PhonVar is a measure similar to PCC, whereas PCC measures correctness, PhonVar measures the deviations. However, the number of PhonVar is not the reverse count of the number of PCC since a single consonant can be affected by more than one phonological pattern. These PhonVar can be differentiated into infrequent variants (InfrVar) and phonological patterns, which are described below.

InfrVar refer to the phonological variants that occur in children's speech but are not frequent enough to be counted as a phonological pattern (Fox-Boyer, Lavaggi, & Fricke, 2021). They have been reported by Dodd et al. (2003) as errors in a child's speech that may occur by chance or occur due to developmental fluctuations. Although McReynolds and Elbert (1981) addressed the importance of distinguishing between such infrequent errors, which they described as "surface error patterns" and actual phonological patterns, InfrVar are usually not considered in the literature on analysis of phonological patterns. Fox (2016) reported that such variants may shed light on the variability and consistency in children's phonological productions because the occurrences of InfrVar in children's speech were found to relate to language competence and the PCC score. Albrecht (2017) found that the bilingual German-Turkish speaking children who produced more InfrVar in Turkish were those reported to have lower Turkish language proficiency and a low PCC-A score. This could reflect the variability in children's production, indicating the instability of their phonological system, which may also be considered for categorising SSDs. Therefore, this measure of InfrVar could add some quantitative information to the qualitative phonological variation analysis (Albrecht, 2017).

The other type of phonological variation is phonological patterns, sometimes referred to as phonological processes in other literature. Stampe (1969) proposed the term 'phonological processes' for the first time when he introduced his seminalwork describing rule-based changes in child phonology (see Chapter 1, Section 1.1.1). However, the recent literature in child phonology does not use the term 'processes' as described by Stampe's theory, but rather to analyse systematic changes in the child's speech concerning adults' realisations of target words (Kirk & Vigeland, 2015); they have become the most common type of relational analysis used by SLTs to describe children's speech (Dodd et al., 2003). In contrast to InfrVar, *phonological patterns* are defined as the error patterns that apply to natural classes of phonemes linked by a common feature or syllable position in a child's speech by substituting one sound class with another or omitting sounds or syllables (Brosseau-Lapré & Rvachew, 2018). Holm et al. (2021) reported that phonological patterns reflect children's understanding of the phonological system of the learned languages. Dodd et al. (2003) preferred the term 'phonological error patterns' to avoid the theoretical assumptions associated with 'processes'. However, Albrecht (2017) claimed that the term 'error' should be avoided too when describing developmental speech. The phonological pattern term is adopted in this research using two cut-off criteria (see below for more details).

The variants in terminology and definitions of phonological patterns are among the most significant challenges faced SLTs analysing children's speech. For example, final single consonants deletion (FCD) should be clearly defined and discriminated from the final consonant cluster deletion (Kirk & Vigeland, 2015). Similarly, a fronting pattern could include two types of patterns: velar fronting and palatals fronting, with velar fronting being resolved earlier than palatals fronting in typically developing English–speaking children. Further, the classification of the consonant cluster reduction pattern is unclear in the literature and is sometimes confused with cluster deletion or cluster substitution. However, Kirk and Vigeland (2015) clarified that these different classifications of consonant cluster patterns need to be differentiated from each other and not be collapsed with the pattern of cluster reduction because each class has a different age of occurrence. For example, cluster reduction, which involves reducing the target cluster (either two or three to one or two consonants), is known to be the last pattern to disappear in the speech of typically developing children, unlike cluster deletion, which is considered as a pattern that rarely occurs in the speech of children. Failing to precisely define each pattern and using a clear description for a pattern in terms of the sound class involved causes huge confusion, especially when establishing normative data, because it may affect the age of suppression and occurrences (Holm et al., 2021). Thus, the definitions of phonological patterns used in a norm study or examined in an assessment tool need to be clearly described to avoid incorrect normative data or misdiagnosis of the phonological disorders.

Another concern with phonological patterns analysis is the method of identifying error patterns. Some studies required a single occurrence of an error to identify a pattern in a child's speech (e.g. Gangji, Pascoe, & Smouse, 2015; Hodson & Paden, 1981). According to McReynolds and Elbert (1981), identifying the pattern without specifying the number of occurrences of this pattern may not distinguish between articulation errors versus error patterns truly resulting from the operation of phonological processes. This is because analysing phonological patterns using single occurrence criteria is similar to the 'surface error pattern analyses' used for analysing articulation errors. For example, labelling a single occurrence of error as a pattern, such as labelling the substitution of stop for fricatives as 'stopping', is similar to relabelling an articulation error not as an identification of the operation of patterns. Therefore, McReynolds and Elbert (1981) suggested the use of the cut-off criteria of four occurrences for an error to be considered as a 'pattern'. However, they reported that this quantitative criterion is arbitrarily chosen and is not stringent, but that it was reasonable as it did not give an unrealistic picture of the children's error pattern. This criterion has been used by many researchers (e.g. Clausen & Fox-Boyer, 2017). Another cut-off criterion frequently reported in the literature for identifying phonological patterns is a higher criterion of at least five occurrences of a pattern (i.e. ≥ 5) (e.g. Albrecht, 2017; Dodd et al., 2003;

Fox-Boyer, 2016); however, no justification has been provided regarding the reasons for selecting this higher criterion.

Difficulty arias, however, when an attempt is made to compare the results from these different studies that based their phonological patterns identification on an unempirical derived cut-off criterion. To date, there has been no research showing what constitutes a sufficient number of occurrences for determining that a child's speech displays a particular phonological pattern (Kirk & Vigeland, 2014). Although Kirk and Vigeland (2015) acknowledged the lack of empirical data on the optimal number of opportunities for a pattern in single-word naming tests to assess the occurrence of phonological patterns, they based their tests evaluation on checking whether a test has at least four opportunities for the occurrence of a pattern, following McReynolds and Elbert's (1981) recommendation. At the same time, they recommended setting a higher cut-off value, especially for patterns involving a large number of different consonants, such as the stopping of fricatives (Kirk & Vigeland, 2015). Therefore, two cut-off criteria are adopted in this current project, one similar to McReynolds and Elbert's suggestions and another one with a higher value following Kirk and Vigeland's recommendation. These two cut-off criteria are used to evaluate the effect of applying two criteria for identifying the phonological patterns of the target population.

1.3.3 Summary of tasks and measures

This section has reviewed the literature about the different tasks and measures used by SLTs to evaluate, diagnose, and set treatment plans for children with SSDs, or to study the typical development of the children's phonology. It could be summarised in the following points:

- SLTs routinely collect speech samples elicited by SWN, CS, and/or stimulability tasks. Each method has advantages and disadvantages regarding its representativeness of the child's speech and its practical application.
  - The SWN task yields a speech sample that sufficiently reflects children's phonological abilities if it is designed based on international constructions criteria considering enough representation of all phones of the target language.
  - A CS sample is usually known for its advantages of having a sample representing the child's daily speech; however, it is difficult to compare different children's samples, and it is not recommended for normative studies.
  - Stimulability tasks are reported to assess the child's ability to imitate a sound absent from his/her phonetic inventory. However, it found that speech sound stimulability is linked to the typical acquisition of phonology since it provides evidence about the development of oro-motor skills (de Castro & Wertzner, 2012; Miccio et al., 1999).

- In this current study, only the SWN task and stimulability task were chosen to collect the speech data from SHA-speaking children for the reasons discussed in Section 1.1.1.1 and Section 1.1.1.3.
- To analyse collected speech samples, it is recommended to use a variety of measures and analytical procedures to identify the child's phonological development.
    - A phonetic inventory is the most independent measure used to list all the sounds a child can physically articulate; however, a relational component could be added to the phonetic inventory by setting criteria for a sound to be included in the inventory.
    - PCC, PVC, and PPC, have been reported as one of the most frequent analytical procedures used to analyse children's phonology.
    - Phonological patterns were also reported to be one of the most analytical procedures used by SLTs (Skahan et al., 2007).
- The quantitative criteria that used to be used by SLTs to discriminate children with SSDs from typically developing children are almost arbitrarily chosen and lacking any empirical evidence.

1.3.4 The theoretical implication of task s and measures

To understand the theoretical background of the tasks and measures chosen in this study, this section discusses the link between these chosen tasks (i.e. SWN and stimulability tasks) and measures (phonetic inventory, PCCs' measures, and phonological patterns) along with the phonological theory that followed in the analysis of this study, that is emergence approach as reported in Section 1.1.3.

Firstly, the traditional SWN tests were developed based on the theoretical concept that claims the unit of phonological description is the individual segments or features (i.e. generative phonology, Chomsky & Halle, 1968). However, according to the emergence approach, words used in child's speech and sampled via SWN task are composed of syllabic sequences with a specific internal structure which is more or less constant in a language. A speaker of this particular language must recognize such a sequence to activate the corresponding meaning (Davis & Bedore, 2013). It is also known that the SWN task usually yielded distinctive features analysis (Schwartz, 1992), however; such an analysis was rejected by Grunwell (1987) and Fee (1995), who argued that the unit upon which the description of structural patterns is based on the syllable and not the segment. Most of the traditional SWN tests that were developed are based on the segmental units and do not identify the contrastive system needed for the phonemic inventory elicitation. Also, they do not provide enough opportunity for each phone to be represented in different phonotactic contexts, which makes them inappropriate for compiling the phonetic inventory either.

Despite the criticism of the SWN task, constructing such a test based on current international construction criteria (see Section 1.2.1) could lead to the analysis of the collected speech sample based on an emergence theoretical framework, including phonetic inventory and phonological patterns analysis. The phonetic aspect of phonological acquisition has been considered in the emergence approach, emphasising the integral importance of peripheral perception and production capacities (Davis & Bedore, 2013). Similarly, according to the emergentist theory, the features of phonological knowledge and behaviour seen in children learning a language imply that the intimate physical, social and cognitive surroundings of the child are relevant and critical to the type of patterns that appear within the complex system of language (Davis & Bedore, 2013).

Therefore, it is necessary when analysing phonetic and phonological aspects of an ambient language acquisition to consider the nature of input (e.g. the dialect of the language children are exposed to and the role of the ambient language) and that children may not have adult-like underlying representations as their perception skills are still developing (Brosseau-Lapre & Rvachew, 2018; Davis & Bedore, 2013; Waterson, 1971). Thus, they may have some patterns that are more complex than the discrete set of natural phonological patterns (cf. Natural phonology, Stampe, 1969). These complex patterns may reveal basic properties of the adult phonological system (Davis & Bedore, 2013). For example, the gemination of consonant sequence pattern in the speech of Arabic-speaking children reflects the generalization of the gemination feature in Arabic phonology (Khattab & Altamimi, 2013), unlike natural phonological patterns (e.g. stopping and weak syllable deletion), which could be related to physical maturation constraints. The SWN task designed based on the ambient language phonological aspects could be a valuable tool to analyse such patterns considering the emergence approach.

Secondly, the emergentist theory could explain the rationale behind using the stimulability task to compile a phonetic inventory along with the SWN task and examine children's oral motor skills. Within this theory, phonetic acquisition is prioritised by phonological systems because this type of acquisition reflects the development and functioning of peripheral speech structures and observable speech behaviours in children (Davis & Bedore, 2013). Phonetic perspectives of the phonological system could explain the potential of a direct relationship between the structure and function of the oral-motor organs within children and what they know and do with that knowledge (Davis & Bedore, 2013). Therefore, analysing the data of the stimulability task (i.e. phonetic imitation task) could explain the relationship between the children’s phonetic perceptions (i.e. the examiner input) and their productions (i.e. children’s response). This task reflects children’s peripheral speech structure capacities, according to emergence approach, which an important aspect in understanding the phonological acquisition process.

1.4 Test construction principles and psychometric properties

When assessing phonological skills in children, international requirements regarding psychometric properties and how these tests are constructed should be considered (Flipsen & Ogiela, 2015). The most important test construction principle is to consider a significant number of linguistic criteria when developing valid and reliable normative data for examining phonological acquisition in children speaking a specific language. These linguistic criteria are one of the most important characteristics of content validity, a psychometric property that needs to be met in a test. Kirk and Vigeland (2014) asserted that test developers need to pay special attention to the representativeness of the normative sample, the validity of the construct on which the test is based, and to have evidence of the test's diagnostic accuracy. These major psychometric parameters are needed for a test to be an accurate diagnostic tool for SSD (Fabiano-Smith, 2019). The following subsection reviews different aspects of constructional international guidelines and criteria, the needed to develop an assessment tool to evaluate the phonological skills in SHA-speaking children, as well as the requirement regarding psychometric properties.

1.4.1 Constructional guidelines and linguistic criteria for developing an assessment tool

One of the most common speech sampling tasks for SLTs world–wide is the single-word naming task (see Section 1.3.1.1). There are many construction guidelines to be considered when developing such a tool, including on test–items selections and linguistic criteria for selecting the test content (McLeod, 2012). However, these criteria may vary depending on the type of analysis applied to the collected speech data, such as compiling phonetic/phonemic inventories, PCC analysis, and analysing phonological variants. The essential construction criteria based on conceptualisation steps, which are needed to create a single-word naming test used to establish phonological information (i.e. phonetic inventory and identifying phonological patterns) (McLeod & Verdon, 2014), are reviewed in detail below.

To ensure that a single-word naming test elicits a representative speech sample, the selected words must be phonetically controlled and include representative opportunities for production of all consonants and vowels of the language under study. Thus, to compile a phonetic inventory, Eisenberg and Hitchcock (2010) suggested that the test words should provide a non–harmonic singleton context. This is because assessing final consonants in words with the same first and last consonant might lead to a higher rate of final consonant production; for example, in English, a word such as "cake" would not be appropriate for a consonantal inventory. McLeod (2012) clarified that different phonetic contexts create different co-articulatory situations (e.g. in English, if the only word that is elicited for assessing /j/ sound is 'yellow', then it is likely that /j/ will be misarticulated due to co–articulatory effects with /l/). In addition, a word-initial consonant needs to be followed by two different vowels, and a word-final consonant needs to be preceded by two different vowels. As noted by Eisenberg and Hitchcock (2010), this is needed to establish a consonant as a separate phoneme and to rule out holistic productions with a single vowel. Therefore, and as reported by James (2001a), consonants need to be repeatedly sampled in all word positions (initial, medial, and final) in a variety of phonotactic shapes.

Words length is another aspect that needs to be considered when developing a test, as the latest recommendation is to include multisyllabic words in a test (James et al., 2016). James (2001a) claimed that such word types should be sampled for many reasons. First, early children's lexicons usually contain simplified renditions of multisyllabic words. Second, including multisyllabic words in a test enables observation of the juxtapositioning of phonological constituents such as stress and syllables which cannot be observed in the production of monosyllabic words. Third, the production of multisyllabic words draws on skills of word syllabification as does spelling and reading, whereas monosyllabic word production does not require syllabification skills. James (2001a) argued that words with complex phonotactic shapes, as well as varied stress and number of syllables, would yield superior ability than less complex words to discriminate children with age-appropriate skills from those who have SSD.

On the other hand, Eisenberg and Hitchcock (2010) mentioned that multisyllabic words do not offer a good opportunity for inventorying word-initial consonant production even if this initial syllable is stressed. This is because it was found that consonants in such a position are produced with less accuracy than consonants in disyllabic or monosyllabic words. They suggested, therefore, to use bisyllabic words to form an inventory of production, choosing a target consonant in a stressed syllable. Further, Eisenberg and Hitchcock argued that unstressed syllables are not suitable for testing consonantal production, since such consonants are more prone to error when they occur in unstressed syllables, thus reducing opportunities for consonants to occur in certain positions (e.g. word-initial consonants in words with an initial weak/unstressed syllable).

Moreover, consonants in a cluster cannot be used to assess single consonants because, according to Eisenberg and Hitchcock (2010), consonant with clusters are not appropriate context for assessing single consonants. Some children tend to omit the marked consonant in a cluster, while others may produce a consonant only within a highly co-articulated context. Therefore, the consonant clusters context could lead to underestimate or overestimate production of consonant; similarly, the morpheme structure could affect the production of the consonant.

To elicit a speech sample that meet the previously stated criteria in a comprehensive and accurate way, there is agreement among researchers that it has to allow for a large number of occurrences of phonological features for subsequent analyses. However, different authors have suggested different ranges of sample size. For example, Smit et al. (1990) included 80 photographs in their test to target 108 phonemes. Stoel-Gammon and Dunn (1985) reported that a representative speech sample needs to include approximately 75–100 words to ensure the inclusion of all structures and segments (phonemes) of the target language. Eisenberg and Hitchcock (2010) noted that the sample size of the single-word naming task is mostly smaller than that recommended for CS samples, including fewer than 90 to 100 words. James (2001a) reported that the number of test items needs to be about 200 as this number is considered to yield a representative speech sample. McReynolds and Elbert (1981) found that 100 words in a sample was considered adequate for obtaining representation of all consonants in a number of contexts. However, most of the available single-word tests have far fewer than what is suggested in the literature as they aim for practical time efficiency. On top of this, as mentioned in Section 1.3.1.1, the test length needs to not be too long as this can reduce the content validity due to tiredness and fatigue, where the child could lose his/her attention which could affect their performance on later items (McCauley & Swisher, 1984). Plus, the limited time available for SLTs may prevent them from using such a long test (Hua, 2006; Tyler & Tolbert, 2002; Wolk & Meisler, 1998). In short, it is important to balance the comprehensiveness of the obtained speech sample with the reality of the time constraints of clinical practice (Eisenberg & Hitchcock, 2010).

Of more importance than the number of test items is the content and structure of these items so that they contain sufficient opportunities for each sound to occur. In terms of a representativeness number of occurrences of a sound, as reported in Section 1.3.2.1.1, many researchers required a phone to occur at least two times in two different words because the effect of surrounding phonemes can influence a child's ability to produce the target sound (Eisenberg & Hitchcock, 2010). Rvachew and Andrews (2002) created a word list that elicited each English consonant in four different positions and considering the word stress in order to examine the influence of the syllable position on the sound production of preschool–age children. They found that the phonetic context of a sound may affect its production and the type of error produced, and the syllable position of that sound may influence its production. James (2001a) thus recommended for all consonants need to be sampled in a variety of phonotactic shapes in words with differing number of syllables. Therefore, it is ideal to include at least three opportunities for each phone in each possible position with different vowels while considering the word shape and word stress. At the same time, it is important to strike a balance between the comprehensiveness of a test and the reality of the time constraints of clinical practice and the suitability of a task to be used with young children (e.g. 2;6 year–olds) (Rvachew & Andrews, 2002; Stoel-Gammon & Williams, 2013).

Regarding including words with consonants in medial position in the speech sample, James (2001a) reported that this would ensure a variety of phonotactic shapes in the speech sample. She explained that medial consonants behave differently compared to those in the initial and final position; for example, an intervocalic /t/ that initiates a weak syllable following a stressed syllable is often realised as a flap [ɾ] (e.g. 'lettuce' realised as [leɾə]). Grunwell (1985) argued that the medial position is, in fact, two positions: syllable–final and syllable initial within a word, which is named as a consonants sequence (i.e. heterosyllabic CC). James (2001a) criticised Grunwell's definition as it did not consider when the medial consonant occurs between vowels (e.g. the medial /p/ in the word 'happy'). Thus, she recommended that the medial position needs to be sampled with both with abutting (heterosyllabic) consonants and between vowels. Interestingly, Rvachew and Andrews (2002) found that ambisyllabic/heterosyllabic consonants are not like the other word-medial consonants (that occur between vowels). They asserted that a failure to distinguish heterosyllabic consonants from the other medial consonants may lead to an inaccurate description of a child's phonological system. Kirk and Vigeland (2015) evaluated the word-medial position as a non–cohesive phonological structure; therefore, this position was not considered as providing a reasonable opportunity for the elicitation of any pattern. However, Grech's (2006) findings, which resulted from investigating Maltese phonological development, supported James's point of view. It is important to acknowledge that Maltese is a Semitic language, and it is originated from Arabic. Grech (2006) considered the medial position in her analyses and found that Maltese children tended to either geminate consonant sequences or delete initial consonants in a syllable, including ones in sequence. Further, Amayreh and Dyson (1998) concluded that the medial position is important and should always be included in any Arabic phonology test, as they found that consonantal production in this position was more accurate than that of initial and final consonants. Therefore the medial position needs to be included in a speech sample intended to analyse phonological patterns considering both intervocalic and heterosyllabic CCs.

Another consideration that needs to be taken into account is the accessibility of the spontaneous naming of the test items. This type of response is preferred over imitating because imitated production may be more adult–like than spontaneous production, which could lead to overestimating a child's phonological ability (Morrison & Shriberg, 1992). There is evidence that spontaneous naming is closer to the habitual speech mode than other elicitation techniques such as imitation (James, 2001a). Spontaneous naming could be facilitated by clear, child–friendly, coloured photos, whereas line drawings have been found less effective for children eliciting test targets. Phonological test developers need to select familiar and recognisable words for the target age range under study and, most importantly, include words that are part of the productive vocabulary of the target age of typically developing children. Because children's phonological ability and vocabulary size are linked in some fashion (Flipsen & Ogiela, 2015; Smith, Mcgregor, & Demille, 2006; Stoel-Gammon & Williams, 2013), it would be inappropriate to use a test that included difficult vocabulary that is not part of children's receptive lexicon. Further, if the target was not named spontaneously by the tested children, it would be difficult to elicit the target word using other types of cues, including forced choice or delayed imitation, if this target is absent from their receptive lexicon (Stoel-Gammon & Williams, 2013). Therefore, it is recommended that a systematic item analysis for the test is reported by examining item difficulty and item discrimination because such an analysis could promote evidence of the validity and reliability of a test (Flipsen & Ogiela, 2015; James, 2001a; McCauley & Strand, 2008; McCauley & Swisher, 1984). A detailed description of this analysis is provided in Section 1.4.2.3.

It is important for a speech sample intended to be used to analyse phonological patterns to include sufficient opportunities for production of the common phonological patterns in a language. Kirk and Vigeland (2015) recommended that a test should include at least four opportunities of each error pattern. Although this cut-off criterion is not empirically proven, it is still widely used in the literature (see Section 1.3.2.2.2). Additional consideration should be given to "positional asymmetries" and the type and position of the target phonemes responsible for some patterns (Kirk & Vigeland, 2015). For example, the word ‘rain’ in English is considered to provide an opportunity for the gliding of /r/, but ‘crab’ is not because /r/ in ‘crab’ occurs as consonant cluster. The fronting of velars is an example of the positional asymmetry because word-initial velars are fronted more commonly than word-final velars. Such phonetic contexts may affect the number of opportunities for this pattern to occur in a child’s speech. Therefore, sufficient opportunities have to be provided in both word-initial and word-final positions for the production of these asymmetry patterns (Kirk & Vigeland, 2014).

Ideally, representative test items should include all the phonemes of the respective language in all phonotactically possible positions, including clusters and gemination, to obtain a complete picture of children’s production skills of individual phonemes (Bernhardt & Holdgrafer, 2001a). Thus, the ability to truly analyse children’s speech is directly related to the type of speech sample the clinician/researcher elicits. Therefore, the linguistic criteria reviewed above need to be achieved in a test to support its content validity. A further review about the psychometric properties of a phonological assessment tool is presented in the following subsection.

1.4.2 Psychometric properties

Having reviewed the most important principles for a representative speech sample and the linguistic criteria that need to be followed, the major psychometrics properties that need to be considered when assessing speech sound skills in children will now be reviewed.

1.4.2.1 Normative sample

A normative sample is the sample of children included in a normative study to establish normative data (Fabiano-Smith, 2019). It is the basis for comparisons of obtained data and is used as a reference in clinical decision–making. This sample should represent a range of variability existing in the general population under study (Kirk & Vigeland, 2014). Dodd et al. (2003) point out that normative sample need to be based on two significant characteristics in order to be a norm that differentiate children with speech and language delay from typically developing children. These characteristics are related to the sample size and composition of this sample which are discussed in the following sections.

1.4.2.1.1 Sample size

It is essential to include a sufficiently enormous number of children in each subgroup when collecting normative data for the standardised norm-reference test. Dodd et al. (2003) clarify that a large sample size is required to achieve a compromise between individual variants in development and representativeness of the norm sample. Having a large number of children in an age group increases the probability that the sample represents the variability and distribution of the test scores of the wider population (Kirk & Vigeland, 2014). Flipsen and Ogiela (2015) emphasised that the number of children in each subgroup is a more important criterion than the number of children in the entire sample. The number of children in each subgroup need to big enough to increase the likelihood that the sample represents speech variability and distribution of scores of the wider population (Kirk & Vigeland, 2014). Similar to many other studies, McCauley and Swisher (1984) recommended that the ideal number of children in each subgroup needs to be at least 100 children. Fewer than one hundred children in a subgroup allows an excessive amount of variability within the sample, creating an unreliable comparison between the child being tested and children included in the sample (Fabiano-Smith, 2019). Although many authors have emphasised the importance of this criterion (i.e. at least one hundred children in a subgroup), Kirk and Vigeland (2014) found that it was seldom met among the tests they reviewed, as this number is too big and need greater time and financial resources which are not always available. Thus, this recommendation it may not feasible or practical. They noted that some of the reviewed tests included a normative sample with one hundred children from each one-year age interval. However, the reported scores and percentile ranks were for six-month or even three-month intervals. As a result, the number of children in a subgroup could be as few as 25 children (Kirk & Vigeland, 2014), which still did not meet the McCauley and Swisher’s criterion.

In general, most of the reported recommendations in the literature regarding sample size have focused on the number of children in an age group, without specifying the number of participants in the whole sample, or the age range that is needed to be included in the sample. However, McLeod and Crowe (2018) recommended that the normative sample needs to include the minimum and the maximum age of the participants to capture children’s acquisition of consonants with eliminating the basal and ceiling effects as much as possible. They also suggested to use a six-age-band to report the age of acquisition of speech sound. However, not all researchers followed these recommendations where some divided their age groups into either to one-year-age band (e.g. Ammar & Morsi, 2006; Pearson et al., 2009), or three-month intervals (e.g. Alqattan, 2015) without clarifying their rationale for choosing such an age-band. A one-year interval between each age group is not preferable because it may not be specific enough for a closed, focused observation of the children’s speech sound development that captures phonological acquisition over a short period in children under the age of three years (Elrefaie, Hegazi, El-Mahallawi, & Khodeir, 2021; Watson & Scukanec, 1997). Flipsen and Ogiela (2015) noted that the narrow age band is intended to capture the rapid rate of change that occurs in many domains during the developmental period; however, they did not specify how narrow the band should be. In short, most of the recent studies followed McLeod and Crowe (2018) recommendation (e.g. Ceron, Gubiani, de Oliveira, & Keske-Soares, 2017; Fox-Boyer et al., 2021; Nojavan-Pirehyousefan et al., 2022). This criterion is followed in this current study as well.

1.4.2.1.2 Sample composition

One important criterion for creating a proper normative sample is to clearly define the participants in the sample to ensure that they are representative of the variability in the general population under study. Many variables need to be considered when sampling children for a normative study, for example, age, socioeconomic variables (including family background), gender, language background, and sibling order (Hua, 2006).

One decision researcher in normative studies has to make is to define children's phonological ability level in the normative sample (i.e. whether to include children with SSD in the sample). Since the publication of McCauley and Swisher (1984), the issues of how to profile a study sample and whether to include atypical developing children in the normative sample have been addressed in the literature. Although the inclusion of children with atypical developing speech sounds is still subject to debate, Pena, Spaulding, and Plante (2006) argued (among other things) that the normative sample must be composed of only typically developing children if the goal of the standardised test is to identify children with disordered speech skills, not severity estimation. They explained that the inclusion of the impaired children in a normative sample of a test that aimed only to find the existence of the language impairment, not to diagnose its severity, may shift the distribution of the children's performance in ways that undermine a clinician's ability to identify the impairment. According to Pena, Spaulding, and Plante (2006), including children with SSD in the normative sample of a test that is intended to identify the existence of SSD in children may negatively affect the accuracy of the test in identifying children with SSD, especially if the number of children in each age group of the normative sample is limited (e.g. less than 40 children, which is the case in my study). Similarly, Fabiano-Smith explained that including children with atypical speech sounds in the sample lowers the mean scores, causing under-identification of children with disorders (Flipsen & Ogiela, 2015). Thus, including children with SSD may shift the distribution of the norm in ways that undermine the clinician's ability to identify SSD in the tested children (Pena, Spaulding, & Plante, 2006).

On the other hand, including a 'full-range sample' (i.e. individuals with disordered speech) is needed if the test is to judge the severity of a particular disorder (Fabiano-Smith, 2019). For example, Kirk and Vigeland (2014) asserted that the normative sample should include children with SSD because educational testing is often used to determine whether or not a child has a disability. Details of the sample type and sufficient information about its composition must be reported in the test manual. It is also found that selecting only typically developing children could limit the clinical application as excluding children with SSD who fall at the end of bell-shaped normal distribution could inflate the normal average that results from such a study (Law et al.,2000, as cited in Hua, 2006). Pena, Spaulding, and Plante (2006) concluded their argument by suggesting that mixed group norm is advantageous because adding impaired cases to the distribution broadens the variability within the normative group.

For this study, the researcher decided to depend on parental judgment on whether the child has been diagnosed with SSD to be excluded. No formal assessment determined their typical speech performance. This approach could be viewed as a broad exclusion criterion (Pena et al., 2006). Depending solely on parental reports cloud be inevitably resulted in including some children with SSD to some extent, but without shifting the outcome towards a false positive (e.g. some patterns are atypical, but because of the effect of including children with SSD in the study sample, these patterns identified as developmental patterns).

It is also important to include children from different age ranges to represent the developmental trajectory over increasing age. To determine the lower age band to include in the normative sample, it is important to have children with consistent speech production. Although it is difficult to pinpoint when children can speak, the average lexicon size for a two–year–old typically developing child is 250– 300 words (Stoel-Gammon & Williams, 2013). Once they have acquired this amount of vocabulary, the nature of their phonological errors' changes, indicating a change in learning strategy (Dodd et al., 2005). Their speech become more consistent, and their pronunciation errors can be described in terms of phonological patterns. For example, they started to realise all fricatives (e.g. /s/) as stops (e.g. /t/). When reorganisation from whole word to segmental phonology occurs (McIntosh & Dodd, 2008), a child begins to develop a 'productive speech sound system' that allows consistent production of a recognisable target (Ingram, 1976). This usually occurs after the age of two when children can integrate information to derive phonological rules (McIntosh & Dodd, 2008). Therefore, a study investigating the phonological system of children needs to carefully select the age of the participants, including the age that correctly reflects the children's phonological abilities. According to Dodd and McIntosh (2009), it is not before two years that children's production can be described in terms of systematic rules that are shared by most children. In addition, children's attention span by this age (i.e. 2;0 years) is usually sufficient for conducting a single-word naming task and obtaining a speech sample that can be phonologically analysed (van Haaften, Diepeveen, van den Engel–Hoek, de Swart, & Maassen, 2020). In contrast, Fox (2007) argued that children younger than 2;6 years may not be responsive to such a task because of their restricted vocabulary. Therefore, she recommended that the normative sample obtained by a single-word naming test should include children from the age of 2;6, which is followed in the current study.

Regarding the upper age band to include in the normative sample, there is a wide variety of different ages reported in the literature. For example, all the tests reviewed by Kirk and Vigeland (2014) included different upper age bands in their normative samples, ranging between 7;11 and 21;11. However, they failed to comment on the effect of the normative sample age range on the accuracy of the psychometric properties they reviewed. Dodd et al. (2003) assigned the age 6;11 as the upper age band among their participants. They found that 90% of English–speaking children over six years of age had error–free speech. Similarly, Owaida (2015) found that Syrian Arabic speaking children no longer produced developmental errors by the age of 5.5, which was the oldest age in her sample. Elrefaie, Hegazi, Elmahallawi, and Khodeir (2021) reported that Egyptian Arabic-speaking children acquired all the dialectal speech sounds by the age of 6;0. In agreement with Amayreh's (2003) findings, Elrefaie et al. found that the standard Arabic sounds /ð, ðˤ, θ, q/ were acquired by children in their sample after the age of 6;0. Since the current project is interested in dialectal Hejazi Arabic, and according to the results of previous Arabic studies, it was assumed that the older children included in the normative sample needed to be 5;11 years.

Another characteristic of sample composition is the demographic variation of the sample. The demographics of the normative sample must be comparable to those of the entire population who may take the test (Kirk & Vigeland, 2014) as some variants exist between speakers of the same or similar ethnic dialects (Flipsen & Ogiela, 2015). McCauley and Swisher (1984) argued that if the child’s dialect differs from the standard dialect or the dialect on which norm-referenced tests are almost always based, his or her performance on such a test may be adversely affected. Similarly, Flipsen and Ogiela (2015) asserted that it is essential to establish local norms that take into consideration potential differences in the development of speech skills and phonotactic constraints (e.g. syllable structure stress patterns and co–occurrence restrictions among segments). Such differences have been proven to exist between the Mainstream American English (MAE) and African American English (AAE), where speakers of these two dialects have different developmental trajectories in the rate and order of acquisition of speech sounds (e.g. accuracy of production of an initial consonant in an iambic word and a cluster in the initial and final position) (Pearson et al., 2009). In terms of Arabic, it is important to consider the large number of varieties of this language when establishing normative data for Arabic-speaking children, especially for speech sound development. Similar to MAE and AAE, it has been found that there are differences in the trajectories of phonological development between children speaking different Arabic dialects. For example, final consonant deletion was shown to rarely occur in the speech of Egyptian Arabic-speaking children aged between 3;0 and 4;0 (Ammar & Morsi, 2006). Therefore, using normative data established on children speaking a particular dialect may not reflect the actual picture of the performance of children speaking a different dialect (McCauley & Swisher, 1984). For this reason, setting speech sound normative data for SHA-speaking children is much needed.

Finally, normative data need to be updated every 10 to 15 years at maximum to accurately represent the current population due to the continuous change in the population demographic over time (Holm et al., 2021; Kirk & Vigeland, 2014). In their recent work, Holm et al. (2021) found qualitative differences in the speech development of a 2015 cohort of children compared to previous developmental norms. These differences could be due to the effect of some factors thought to affect the language learning environment of children over time, including exposure to another language, early preschool education, and increasing in the time of the screen exposure. Holm et al. (2021) concluded that the normative data of standardised assessments for SSD need to be periodically re-evaluated to ensure the validity of normative data.

Based on the literature reviewed above, the current study aimed to investigate the phonological development of SHA-speaking children using a normative sample that took into considering the reviewed criteria whenever possible. For example, it included 235 children who spoke a similar dialect. The age range as well as the size of the subgroups were all factors considered when creating the sample for this study (see Chapter 3, Section 3.2, for more details).

1.4.2.2 Reliability

Test reliability refers to the consistency of test results when the testing procedure is repeated over time (test-retest reliability) and across different examiners (inter-rater reliability) with consistent test scores across the test items (internal consistency) (Kirk & Vigeland, 2014; McCauley & Swisher, 1984). Reliability is one of the most important characteristics that provides necessary, but not sufficient, evidence for the test validity. There is a debate in the statistical literature about what are the appropriate indices that should be used to report reliability (Shweta, Bajpai, & Chaturvedi, 2015). Belur, Tompson, Thornton, and Simon (2021) reported that it has been recommended that reliability should be measured or estimated using multiple calculations to ensure that the estimated reliability is not due to chance. The decision on choosing the most appropriate measure is based on the purpose of the analysis and the complexity and the type of data required for the reliability measure (Belur et al., 2021; Shweta et al., 2015). In this section, the most important forms of reliability that should ideally be measured will be reviewed.

Test-retest reliability (test consistency) is the ability of a test to yield the same score when given more than once over a short period (Fabiano-Smith, 2019). This form of reliability aims to distinguish true score variance from random and transient measurement error. Such transient error can result from time of the day, attention, fatigue levels, and respondents’ mood; all can contribute to transient error and, consequently, affects the test result (Flipsen Jr & Ogiela, 2015; Polit, 2014). Ideally, to determine test-retest reliability, the test designers must test the same group of individuals twice within a relatively short time, and then calculate the reliability correlation coefficients and the point-to-point percentage of agreement between the test scores. If the two test scores result in a reliability coefficient of at least .90 as the minimum standard, which is statistically significant at or beyond the 0.05 level, then evidence of the test's stability over time is met (Kirk & Vigeland, 2014; McCauley & Swisher, 1984). When checking the reliability of a test designed for young children, the interval between the tests being administered need to be carefully considered (i.e. two weeks at the maximum). Test learning and developmental growth factors should also be minimised and not affect the test scores (Kirk & Vigeland, 2014). Although this form of reliability is essential evidence for a reliable test, Flipsen and Ogiela (2015) found that fewer than half of the tests they reviewed reported this form of reliability with correlation coefficients of at least .90. They argued that the tests failed to achieve this criterion because it is too high, and most of the tests reported correlations between .70 and .80, which are considered to show acceptable reliability. Therefore, lowering the criteria could raise the number of tests meeting this reliability aspect.

Even though this type of reliability measure is recommended, it was not applied in this thesis for the following reasons:

1. The data collection was carried out in Saudi Arabia, and the trip had a strict time limit.
2. It was not possible to retest 10% of the whole sample within the recommended time limit, which is a minimum of two weeks after the first testing, according to Carmines and Zeller's (2011) recommendations.
3. As reported, the main goal was to prioritise collecting a large enough primary sample.

Inter–rater reliability (IRR) is another type of reliability that assesses the extent to which test scores are consistent across different examiners and how much the examiner's style and/or rapport with the child influences scores (Flipsen Jr & Ogiela, 2015; McCauley & Swisher, 1984). This aspect of reliability is either assessed by having different examiners administer the test with the same participants within a short interval or by letting other examiners analyse the performances of the same children (Kirk & Vigeland, 2014). Although there are no specific recommendations in the literature regarding the number of raters or participants that need to be included in the sample for IRR testing, it is common to measure this form of reliability by retesting/reanalysing a randomly selected 10% portion of the whole sample by a qualified rater (Albrecht, 2017). Belur et al. (2021) argued that multiple raters are not always suitable for reliability measures, especially if the data are too complicated and are not nominal, as their judgement could involve subjective interpretation. Therefore, the current study assigned only one rater to check the reliability of its data.

One of the metrics used to calculate IRR is percentage agreement, which is calculated by point–to–point agreement (Eagan, Brohinsky, Wang, & Shaffer, 2020). This refers to how often raters agree on the exact rating and can vary between 0% and 100% (Shweta et al., 2015) The concept of percentage agreement among raters is easily calculated, the simplest to understand compared to other measures, and directly interpretable (McHugh, 2012; Shweta et al., 2015).

There are no rigid rules regarding the level of agreement needed to judge whether the IRR is sufficient (Shweta et al., 2015). However, McCauley and Swisher (1984) reported that the minimum standard of IRR is a high correlation coefficient of 0.90 or higher (significant at $p <.05$) which is required for a test to be considered reliable for making clinical decisions. In terms of percentage agreement, it has been reported that 75% to 90% agreement demonstrates an acceptable level of agreement (Shweta et al., 2015). However, for research purposes, reliability criterion of 0.70 is considered sufficient, while 0.90 is the criterion required for individual clinical diagnosis (Nunnally,1999, as cited in Schaefer et al., 2009). Since the main goal of the test created for the current study is for it to be used for a clinical purpose, a 90% agreement cut-off criterion was set for the reliability check.

Similar to inter–rater reliability, McLeod (2012) explained that a rater reliability can also be applied when the same test results obtained by the same examiner and scored at a different time to ensure *intra-rater reliability*. Because ensuring reliability requires more than one measure (Belur et al., 2021), both of these forms of reliability (inter-and intra-rater) were applied in the current study.

Internal consistency is another type of reliability metric that refers to how a group of test items designed to measure the same overall construct produce similar scores (Kirk & Vigeland, 2014; McLeod, 2012). One proposed measure for this kind of reliability is splitting the targeted test into two halves and then correlating the participants' scores on the two halves, which is called split-half reliability. Carmines and Zeller (2011) clarified that the results of such a reliability measure might differ depending on how the test is split, as there is no one way to divide a test. However, these methods are not commonly used to evaluate the reliability of tests that measure speech and language ability. Albrecht (2017) explained that this type of reliability is not suitable for a single-word naming test. This is because words included in the assessment are purposely chosen to very in terms of their phonetic structure and how this structure allows for a variety of phonological patterns so that maximum scores per item cannot be predicted.

Another essential measure related to reliability reported by Flipsen and Ogiela (2015) is *standard error measurement* (SEM). This measure estimates the error margin around any particular score, and it represents the standard deviation that would be obtained if an average person took the test many times (Flipsen & Ogiela, 2015). Therefore, SEM provides valuable guidance for clinicians who test phonological ability in children and assess their speech abilities one single time and document and monitor their developmental progress. It assists SLTs to interpret tests results by providing a possible range of scores within which the true scores lie (Flipsen Jr & Ogiela, 2015). This measurement refers to the standard deviation of scores a child would produce if he/she was asked to repeat the test several times (Albrecht, 2017). The standard deviation (SD) of the whole age group's score was calculated for this current study. McCauley and Swisher (1984) suggested that the means and SD values for each subgroup need to be provided because the mean is the average score of members in the normative subgroup, while the SD gives the test users an estimate of the subgroup members' score variants.

To ensure a good level of reliability for a given test, it is always recommended that the test manual provides clear instructions to test users in a consistent and standardised way. It is also important that sufficient information on the qualification and training needed to professionally administer the test is provided (McCauley & Swisher, 1984), as well as the type of analysis required by the test. Lastly, it is important to report reliability evidence for each subgroup in the normative sample because the reliability coefficients scores may vary across the groups (Kirk & Vigeland, 2014; McCauley & Swisher, 1984), for example, the younger age groups may have lower reliability. However, Kirk and Vigeland (2014) reported that the reliability correlation coefficient for the total scored group could be strong and positive even when there is a lack of agreement point between individual items. Therefore, the reliability measures used in this study are reported for all age groups together, without giving an individual score for each age group.

1.4.2.3 Validity

A test is considered psychometrically valid if it accurately measures what it is designed to measure (Fabiano-Smith, 2019; McCauley & Swisher, 1984). Different aspects of validity are usually considered important for any test that measures behaviour or which is used to make inferences about underlying abilities. Test developers need to design their assessment to meet certain aspects, including construct validity, content validity (involving content relevance and content coverage), criterion validity, and diagnostic or identification accuracy (Eisenberg & Hitchcock, 2010; Fabiano-Smith, 2019; McCauley & Swisher, 1984; McLeod, 2012). Messick (1980) argued that 'construct validity' is the basic meaning of validity. He stated that construct validity is the unifying concept of validity that integrates content validity and criterion validity into a common framework for testing the hypotheses about theoretically relevant relationships' (Messick, 1980, p. 1015). Some evidence about the validity aspects of phonological assessments is provided below.

The first criterion that needs to be met is *test content validity* evidence. McLeod (2012) defined content validity as the degree to which the test items represent the domain of concern. For content validation, a test developer needs to provide any of the following pieces of evidence to meet the content validity criterion: a) content relevance, b) content coverage, and c) item analysis (McCauley & Strand, 2008). Content relevance involves specifying a test's domain (Eisenberg & Hitchcock, 2010), and measures whether the test content is relevant to the measured behaviour (McLeod, 2012). For an assessment that is intended to assess the phonological development of monolingual children, it is important to consider the structure of the language's phonology (e.g. the number of elements allowed in a consonant cluster and its permitted word position, either initial or final) to ensure that the test content is relevant to what is being assessed (Fabiano-Smith, 2019). Content coverage relates to the degree of representativeness with test samples in that domain (Eisenberg & Hitchcock, 2010). To meet the criterion of content coverage, it is important to provide enough opportunities to test each speech sound production. For example, it is necessary to target all vowels in a target language with enough frequency to accurately represent a child's abilities. Similarly, sufficient opportunities for the production of all common phonological error patterns in the target language should be provided (Kirk & Vigeland, 2015).

In general, to present adequate evidence of content validity, a phonological test has to consider a significant number of linguistic criteria (Kirk & Vigeland, 2014). Some of those criteria are reported in Section 1.4.1, and may vary depending on the type of analyses applied in the studies, including phonetic versus phonemic inventories, percentage of consonants correct, and phonological error patterns, as these set the boundaries of their test domains (Fabiano-Smith, 2019).

Another issue related to test content validity is item analysis (see Section 1.4.1), which is rarely included in the psychometric information in published tests (James, 2001a). This type of analysis can be done either through statistical analysis during test development and/or may involve some systematic field testing of items to evaluate children's real-world responses (Flipsen & Ogiela, 2015). The measures and scoring for such an analysis have not been finalised in the literature, since there are different cutoff criteria used to judge test-item difficulty, either 50% or 80% (e.g. Ceron, Gubiani, de Oliveira, & Keske-Soares, 2020; James, 2001a; Nojavan-Pirehyousefan et al., 2021; Stoel-Gammon & Williams, 2013). Similarly, different measures are administered for analysing test items, including counting the instances of spontaneous production, examining whether item pictures are recognised by the children during their testing, or calculating the naming agreement with the intended target (e.g. Ceron et al., 2020; James, 2001a; Nojavan-Pirehyousefan et al., 2021). James (2001a) conducted item analysis on a cohort of words chosen to represent all Australian English phonemes. She applied two aspects of item analysis: item difficulty and item discrimination. Item difficulty is defined as the proportion of persons responding appropriately to an item. Ideally, for a test designed to sample children's speech, it should have a low measure of item difficulty. A low measure of item difficulty indicates that children are familiar with the words included in the test, and they can quickly and spontaneously name the item. Spontaneous naming of the test items means citing an item with as little cueing as possible, which is ensured when the items have low item difficulty. James (2001a) applied systematic field testing on 59 typically developing Australian English–speaking children, where she coded their word naming for accuracy of naming to measure item difficulty and accuracy of articulation to measure item discrimination. Accurate naming was credited for spontaneous naming, while accurate articulation was coded when children articulated a word accurately. James (2001a) applied a cut-off criterion of 50% or more of children naming an item spontaneously for it to be deemed to have an adequate item difficulty. Her results showed that 70% of pictures were spontaneously named correctly as the intended name by 50% of children from both cohorts. The 30% of words not named correctly were unfamiliar words, difficult, or difficult to illustrate, such as verbs and articles. In the final version of her test, she converted these verbs to nouns because she found that nouns are the most accessible words for the children to name.

Similarly, Ceron et al. (2020) conducted a study to analyse the recognizability of items in their newly developed phonological assessment tool. Their sample had 48 children aged between 3;0 and 8:11, with eight children in six age groups, who were older than the youngest age group of James's study. The researchers required a score of 80% of their test items to be named and recognised spontaneously. Ceron and colleagues (2020) classified the percentage item recognition rate based on three cut-off criteria:

a) 80% or more: If an item was recognised and named by 80% of the children, then it was retained in their instrument.
b) 60–79%: If an item was named by 60–79% of the children, changes were made to the image or prompt to ensure the target word was produced.
c) 60% or fewer: When an item was named by 60% of children or less, it was removed from their word list.

However, they did not justify their choice of these reported cut-off criteria. The most–reported cut-off criterion for an item to be considered a familiar word and included in a test was to be named by at least 50% or more of the children (James, 2001a; Stoel-Gammon & Williams, 2013).

Considering the youngest age of children who participated in the reported studies, it is evident that Ceron et al. (2020) is the only study that did not include children aged between 2;0 and 2;11; their youngest age was 3;0. Thus, using a strict cut-off criterion to identify test-item familiarity in a study including children aged between 2;0–2;11 may not be suitable, because these young children may not have all test items in their productive lexicon (Stoel-Gammon & Williams, 2013) and their inability to name items spontaneously is because of their limited lexicon inventory. Moreover, Ceron and colleagues counted all words named correctly, either spontaneously or by prompts such as sentence completion cues, unlike other studies that only counted the spontaneous naming of items. They only excluded words named by direct imitation. This calculation may also have been taken into consideration when justifying their high cut-off criterion (i.e. 80%).

Another validity aspect that needs to be evident is the *criterion–related validity,* which involves the collection of empirical evidence showing that the test scores are related to some other measure of the behaviour being assessed (McCauley & Swisher, 1984). In other words, criterion validity is the extent to which one can predict an individual's score on another measurement from his/her performance on a given instrument (Raykov & Marcoulides, 2010). Concurrent validity is one kind of criterion-related validity that is usually reported in test manuals. It is determined by assessing how closely an individual's test score is related to his or her score on another already validated test measuring the same or similar abilities (McLeod, 2012). The correlation between these assessments needs to be high to demonstrate that the assessments would generate similar results, and for making accurate diagnostic and intervention decisions (McLeod & Verdon, 2014). According to Albrecht (2017), such a high correlation could be problematic since the necessity of this new test would be questionable if it did not show any advantage over the one it was compared to. Moreover, this type of validity cannot be achieved for assessments that are the first of their kind when there is a lack of other linguistically or psychometrically suitable assessment tools available for comparison, which is the case in the current study.

Predictive validity is another kind of criterion–related validity, according to McCauley and Swisher (1984). It is examined by assessing how closely an individual's test score can be used to predict future performance on a criterion measure, as well as if it displays a developmental trend (McCauley & Swisher, 1984). To meet this criterion of predictive validity, a test manual needs to report any outcome for which competency with speech production might have an impact (e.g. speech skills, reading skills, or literacy skills) (Flipsen Jr & Ogiela, 2015). However, Kirk and Vigeland (2014) criticised this type of validity as a weak measure because so many skills and individual qualities are developmental.

The last type of evidence to consider about test's validity is *diagnostic or identification accuracy*, which refers to evidence showing that the test is valid for making diagnostic decisions (Kirk & Vigeland, 2014). The evidence needed to support this diagnostic decision concerns test sensitivity and specificity (Flipsen & Ogiela, 2015; Kirk & Vigeland, 2014; Peña, Spaulding, & Plante, 2006). Sensitivity refers to the degree to which a test correctly identifies children who actually have SSD as having a disorder. Specificity refers to the degree to which a test accurately excludes typically developing children from the disorder diagnosis (Flipsen & Ogiela, 2015; Kirk & Vigeland, 2014; Peña et al., 2006). The item discrimination measure relates to sensitivity and specificity, as it measures how test items differentiate children with age-appropriate skills from those who have a speech impairment. James (2001a) defined item discrimination as the degree to which items differentiate between subjects on the behaviour being measured. She concluded that a test with a cohort of words that repeatedly samples all phonemes (consonants and vowels) in all words positions with varied phonotactic shape, number of syllables, and stress will have a high item discrimination measure.

1.4.3 Summary of test construction principles and psychometric properties

The most important aspects of test construction principles for assessing expressive phonology in children were reviewed, and it includes the following:

- Considering the following international construction criteria in developing the SWN test could ensure its validation and reliability evidence:
  - Test items selected for such a tool should meet constructional guidelines in including phonetically controlled words (e.g. word structure, shape, and number of opportunities for each phone).
  - To sample the phonemes with representative frequency within a test with at least one to five occurrences in different words depending on the frequency of phonemes in the language/dialect under study (Flipsen Jr & Ogiela, 2015).
  - To provide at least four to five opportunities for producing each phonological pattern common in typical and atypical phonological development (Kirk & Vigeland, 2015).
  - Detailed information about test reliability needs to be provided, especially regarding test–retest reliability and inter/intra-rater reliability.
  - The content validity of a specific test is a fundamental criterion that must be achieved to ensure the adequacy of the single-word naming test.
- Psychometric principles need to be taken into account when developing a test, including normative samples, reliability, and validity:
  - The normative sample needs to include a sufficient number of children in each age group to represent the variability in the targeted population.
  - The inclusion of the atypical developing children (e.g. children with SSD) in a normative sample is a debatable issue. The normative sample should children with different speech sounds skills to ensure a normal distribution of the mean test score.
  - Test reliability needs to be calculated using appropriate measures that suit the purpose of the analysis and the type of data collected by that test.
  - Validity evidence needs to be met for a test, particularly the content and diagnostic validity, to be a clinically useful tool for SLTs.

1.4.4 The theoretical implication of test construction principles and psychometric properties:

The criteria for developing a test to measure children's phonological acquisition is usually based on a theoretical background which is needed to analyse the collected data. This section discusses the theoretical rationales of the criteria selected to develop SHAPA (the phonological test used in this study).

Firstly, the word selection of SHAPA, was generally based on the assumption that the central phonological unit is the syllable and that there is a need to expand the phonological analysis of phonemes by considering the syllable position of the phone (Rvachew & Andrews, 2002), and not focusing only on the individual phoneme (i.e. the paradigmatic nature of phonemes; McCormack, 1994, as cited in

James, 2001). This selection includes choosing words with multisyllabic structures and phonetically controlled contexts, as well as presenting enough opportunities for each phone to occur (see Section 1.2.1). It is suggested in modern phonology to consider these criteria (i.e. whether a range of different syllable shapes was achieved and manipulated in the test's words) to test the phoneme in a wide range of phonological constituents (James, 2001). Hence, examining the syntagmatic nature of phonemes in a specific language is recommended by considering the number of syllables in words, the nature of the sound sequence, the stress of the syllables in which phonemes occur, and the different sequence of phonemes (James, 2001). The resulting speech sample, therefore, could capture many features of speech development in children. These phonotactic aspects of adult-speech develop within a child's speech when s/he transfers from word-based to phoneme-based lexical representations (Brosseau-Lapre & Rvachew, 2018). This is when the child's production changes from undifferentiated syllables to differentiated syllables then into individual articulatory gestures, according to the self-organisation model, a model of the emergentist theory (Davis & Bedore, 2013). For further details on the self-organisation model refer back to Davis and Bedore (2013), p. 49-52.

Another criterion used in developing a phonological test is the inclusion of familiar words into the test corpus. According to the emergentist approach, phonological development is interacting with many other areas of language development, including lexical development (e.g. Sosa & Stoel-Gammon, 2012). Vihman (2010) reported that infants, as early as 11 months, are supposed to be more sensitive to familiar words regarding onset consonants. Whereas unfamiliar words that are not used frequently in the infant's environment present a more significant challenge to the memory and reflect less accurate representation that is often 'holistic' or 'gestalt-like' rather than 'fully specified' or detailed. Thus, the production of these familiar words is assumed to be more stable because of the well-practised motor routine, which may involve enhanced abstraction of phonological information (Sosa & Stoel-Gammon, 2012). As a result, familiar words should also stimulate spontaneous responses, which are always preferred over imitative responses in the SWN task because they are expected to reflect the child's underlying representation of his phonology.

Third, the composition of the normative sample is an important requirement in theoretically based test development. It has been established that including children under two years old in studies of phonological development in children is not beneficial for various reasons. According to Vihman and Greenlee (1987), it is improbable that children's emergent phonological systems (i.e. at the age of one year) are founded on complex representations and linkages between adult aims and child output at their earliest stage of word formation (i.e. at the age of one year). Grech (1998) also found from her evaluation of the research that phonological analysis of children's speech before the age of two is not advised. She discovered that by the age of 2;0, the phonetic and phonemic organisation might be approaching. This is because children may not be expected to have basic phonemic representations until the developmental age of 2;0. This has immediate implications for this study, as analysing the speech data of children aged two would typically involve phonetic repertoire (or inventory) measures (i.e. independent analysis). Furthermore, their performance could not be compared to the performance of adults' speech (i.e. relational analysis) or even to older children using the same measures (Brosseau-Lapre & Rvachew, 2018; Vihman & Greenlee, 1987). This could lead to the conclusion that there is no use in analysing children's phonology in terms of phonemic and phonetic inventories before the age of two because the foundation of the phonological organisation begins when children enter their word production period (Vihman & Croft, 2007).

For the current study, it was essential to review all linguistic criteria of the language under investigation and understand its phonology. The following chapter reviews Arabic phonology, with a detailed description of Saudi Hejazi Arabic. This is followed by examining the available literature about phonological development in Arabic-speaking children, taking other Arabic dialects into consideration.

2 Literature Review II – Arabic phonological system and development

This chapter introduces Arabic phonology in general, together with a detailed description of the Arabic dialect under study: Saudi Hejazi Arabic (SHA). The first section introduces the phonological system of the Urban Hejazi Arabic (UHA) and Modern Standard Arabic (MSA), while the second section reviews the current state of knowledge on the typical phonological development of Arabic, including studies on other Arabic dialects. The final section of this chapter presents an overview of existing assessment tools used to examine productive phonological skills in Arabic monolingual children. Following this, the limitations of the available Arabic articulation tests, will be explored and the need for a new assessment tool to explore the phonological skills of SHA-speaking children will be highlighted. The chapter will conclude with the study aims and research questions.

2.1 Phonological system of Arabic

Understanding the phonological acquisition of Arabic-speaking children requires knowledge of the phonological structure of the language. It is widely known that differences in phonological development stem from the phonological characteristics of the ambient language; therefore, it is important to identify these characteristics, which reflect the nature of the specific phonological system in order to fully understand the phonological acquisition of children exposed to this language (Hua & Dodd, 2006).

The purpose of this section is to provide a comprehensive description of the phonology of Arabic, the language of interest in this study, specifically that of UHA. A brief history of the Arabic language and its different forms is provided first, including Classical Arabic (hereafter, 'CA'), MSA, and the dialectal variants. Following this, sociolinguistic background information is given to show how the Arabic language is used within Arabic societies, with a thorough description of the use of UHA versus MSA in Saudi Arabia (SA) (see Section 2.1.1 below), the country in which this study was conducted. It also presents a brief comparison between UHA and the other varieties of Arabic in terms of their phonological features, including consonantal and vowel inventories, consonants clusters, phonotactic features, and stress patterns. These verities include MSA, Egyptian Arabic, Jordanian Arabic, Kuwaiti Arabic, and Najdi Arabic. Since most of the phonological features of Arabic reported in the literature are taken from MSA, this section provides the phonological description of UHA based on a comparison with MSA, highlighting the similarities and differences between UHA, MSA, and other Arabic dialects. The other Arabic dialects are included in this comparison because most of the available Arabic phonological acquisition literature is based on studies carried out on children exposed to these dialects (see Section 2.2).

2.1.1 The historical background of Arabic

Arabic is a South Semitic language that fall within the Afro-Asiatic language family and is closely related to Aramaic and Hebrew[2]. It is the official language of 25 Arab countries situated in much of North Africa and the Middle East, including the Arabian Peninsula (Watson, 2002). Around 422 million people speak Arabic as their native language, with a significant variation in their dialects (IstiZada, 2020).

Historically, Arabic has undergone different stages of development and progress over the centuries. Researchers agree that the birthplace of the Arabic language is the central and northern region of the Arabian Peninsula (Lipinski, 1997; Versteegh, 2006; Watson, 2002). Birnstiel (2019) referred to the Arabic spoken in the Arabian Peninsula before the sixth century as 'Old Arabic', while Watson (2002) agreed with Lipinski (1997) that Old Arabic had three different forms: Hejazi, Najdi, and a dialect spoken by other Arabian tribes in the northern Arabian Peninsula. With the beginning of Islam, the Holy Quran was revealed in Arabic, particularly in the form of 'Old Hejazi' (Al-huri, 2015; Watson, 2002). Old Hejazi is described in the literature as the purest dialect of Old Arabic, one which had not been contaminated by other Sematic or non-Sematic languages, unlike other forms of Old Arabic (Al- huri, 2015; Watson, 2002). It is important to note that the Old Hejazi dialect is distinct from the modern Hejazi that is presently spoken in the western region of SA.

In the seventh century when Islam began to spread outside the Arabian Peninsula, CA began to emerge. This form of Arabic was primarily based on the language of the Quran, with some interference from the pre-Islamic poetic koine (Al-huri, 2015; Watson, 2002). CA holds the most prestige compared to other Arabic varieties among all Muslims worldwide due to its religious and historical status of being the language of both the Quran and the literary heritage of Arabs (Al-huri, 2015). It is mainly perceived today as the written and oral language used when reciting the Quran and is the language of all Islamic speech and activities (e.g. prayers).

In the eighth century, Arabic grammarians began the standardisation and codification process of Arabic to preserve its purity after the linguistic mixing of native Arabic and non–Arabic speakers (Al-huri, 2015). The term 'Modern Standard Arabic' (MSA) then emerged, describing a form of Arabic that was originally the modern descendant of CA. Many linguists view MSA as a modified edition of CA that maintains the latter's phonology, morphology, and syntax (e.g. Holes, 2004; Versteegh, 2006). However, MSA uses only a subset of the syntactic structures of CA. The lexis and stylistics feature of

[2] Arabic and Hebrew are two unique examples of the living languages from this language family (Afro–Asiatic family) (Al-huri, 2015).

MSA are also somewhat different from those of CA; MSA has come to have a reduced lexicon after dropping some of the archaic words and phrases and replacing them with new technical and scholarly lexis (Al-huri, 2015; Bin–Muqbil, 2006; Versteegh, 2006; Watson, 2002). Overall, most researchers consider CA and MSA as two registers of one form of Arabic because of their very close linguistic and orthographic similarities (Al-huri, 2015; Birnstiel, 2019). However, Birnstiel (2019) argued that CA, rather than MSA, is considered the mother language of all spoken varieties of Arabic in the Arab world.

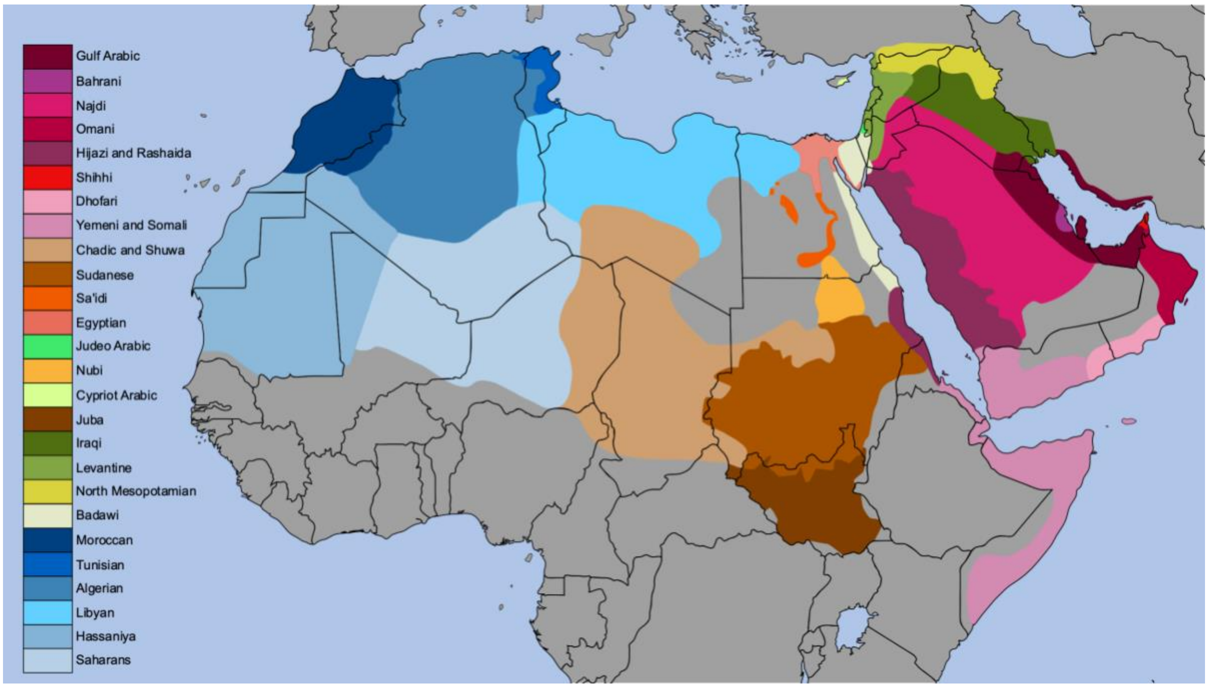
In term of the emergence of the other Arabic dialects, Versteegh (2006) claimed that Arabic varieties began to emerge from MSA because of the spread of Islam during the second half of the seventeenth century, when Arabic came into contact with many different languages (e.g. Barbar, Persian, Greek, and South Arabic). In contrast, Al Nassir (1985) and Lipinski (1997) argued that the varieties of Arabic began to emerge in the sixth century, at a time when different forms of Old Arabic already existed. However, Birnstiel (2019) concluded that it is difficult to obtain an accurate detailed historical picture of the emergence of the Arabic varieties because several of the morphological and syntactic features of certain varieties are more related to CA than to MSA. Moreover, Holes (2007) claimed that MSA is viewed as a pan-Arabic 'lingua franca' used whenever dialectal differences lead to lack of intelligibility.

MSA differs considerably from the Arabic dialects in terms of phonology, morphology, syntax, and lexicon, which are slightly modified in each dialect and differ from one dialect to another (Omar, 1975; Watson, 2002). However, Al Nassir (1985) stated that it is not easy to determine the degree of linguistic difference between Arabic dialects. This linguistic phenomenon is termed in the literature as 'diglossia', first introduced by Ferguson (1959), and describes a well-known situation whereby MSA is intertwined with the colloquial or dialectal Arabic within a community (Basalamah, 1990). According to Saiegh-Haddad (2004), Arabic is a clear example of diglossia as it has two distinct sets of functions performed by two different linguistic codes that have strict socio-functional levels. In other words, these two different forms of the language (i.e. MSA and the dialectal form) serve two different functions within a community. MSA is the language of education and official occasions and is described as a 'high' or 'elite' level of the language. In contrast, the dialectal forms of Arabic are used for daily conversation in public areas, such as in casual gatherings, when shopping and in family conversations, and are seen as 'low' levels of the language.

As in all other Arabic countries, the western region of SA experiences diglossia involving MSA and UHA, the dialect of this region, where UHA is the spoken language and MSA is the language of education and the formal writing system. However, some informal Arabic writings can use the UHA dialectal form, such as on social media and in some newspapers (Holes, 2007). The following section will provide more information about the sociolinguistic situation regarding the use of MSA versus dialectal Arabic in the Arabic world, with specific information concerning the use of UHA.

2.1.2 Sociolinguistic aspect of Arabic

Arabic varieties are classified according to the region in which they are spoken as follows: Maghrebi, Sudanese, Egyptian, Mesopotamian, Levantine, Arabian Peninsula, and Peripheries (see Figure 2.1). Each regional variety has been influenced by some languages depending on the original languages of each region. In Figure 2.1, each region is illustrated with a particular colour group (e.g. Arabian Peninsula dialects are coloured in different shades of red).



(https://istizada.com/complete–list–of–Arabic-speaking–countries–2014/)

Figure 2.1 *Different Arabic varieties in the Arab world*

Ingham (1971) also classified the Arabic dialects spoken in SA into four main groups: the northeast Arabian dialects, the southwest Arabian dialects, the northwest Arabian dialects, and west Arabian dialects (see Figure 2.2).

Figure 2.2 *Main dialects spoken in Saudi Arabia (from MGHAMDI.COM, 2014*)

In Arab world, no one is brought up speaking MSA as a mother tongue (Al-huri, 2015; Birnstiel, 2019; Watson, 2002). Instead, children develop knowledge and fluency in the dialect of their region because it is the primary spoken language in their homes. At the same time, children start learning MSA formally at schools or home, as part of their education. MSA is confined to formal written and spoken occasions, and children need to be taught this form of Arabic to be able to read and write in Arabic. In some cases, early exposure to MSA can begin before schooling commences because most TV programmes and other media materials for children (e.g. cartoons, CDs, YouTube channels) are in MSA.

In SA, colloquial Arabic is used within family’s conversation and, therefore, dialectal forms of Arabic are the primary family input for children. When children commence education, their exposure to MSA increases, even in kindergarten. For example, in the western region of SA, all kindergarten teachers teach primarily in MSA (e.g. at circle time or when storytelling). However, the children’s conversations during all other activities (e.g. free-playtime, mealtime) are in dialectal Arabic, which is supposed to be UHA. However, children can also be exposed to many different Saudi dialects (e.g. Najdi, Southern) as Jeddah, the main city of this region, is the second biggest city in SA and many Saudi’s families move to it of either work or education. Jeddah is considered as a multi-cultural area with around 2.13 million expats from 1.9 million nationals who came to the region either for work, to live, for Hajj, or came as legal or illegal immigrant (Arabian business, 2016; De Bel-Air, 2018). Therefore, the children in SA exposed to wide dialectal varieties and foreign accented speech which are spoken in their environment (e.g. schools, community, friends). They get ‘bilingually’ influenced, or as Levy and Hanulíková (2019) argued, they were bidialectal or multidialectal since they were exposed to MSA, UHA as well as other Arabic verities and foreign accented speech at an early age in their life. In such a context, children simultaneously develop two linguistic systems: MSA and the everyday dialectal form (e.g. UHA), and these two systems may vary in the extent of their phonological structure (Saiegh-Haddad & Haj, 2018). Still, children exposed to these two linguistic systems were considered as monolingual according to Levy and Hanulíková (2019) who considered children exposed to one of the German dialects (i.e. Swabian) and the Standard German as monolingual because they were “heterogeneous” group with passive and active knowledge of German varieties. Levy and Hanulíková added that these varieties were roofed over by one standard form which means that these children are not using two distinct languages (Levy & Hanulíková, 2019, p.6). Similar explanation could be applied on SHA-speaking children as they exposed to UHA and MSA as mentioned earlier.

In the following section, some historical information about the dialect of interest for this current study (UHA) and how the linguistic diversity in this area affected this dialect are provided.

2.1.2.1 Urban Hejazi Arabic

Urban Hejazi Arabic (UHA) is one of the Hejazi Arabic (HA) dialect spoken in the Arabian Peninsula by the people in the western region of SA, which includes the cities of Jeddah, Makkah, Taif, Madinah and Yanbu (see Figure 2.2). Al Nassir (1985) claimed that HA dates back to early Islam, where it was the dialect of the old ‘Quraysh’ tribe in Makkah. Since then, classical HA has developed markedly, and modern HA is quite distinct from the classical form of Arabic (Watson, 2002). There is current agreement among linguists that modern HA is widely used in the Arabian Peninsula throughout SA for specific purposes, such as for governmental and commercial purposes, and that it is widely understood across the Arabian Peninsula and the whole Arab world (Basalamah, 1990; Omar, 1975). HA is spoken in SA by almost 14.5 million people which is almost 41% of Saudi Arabic’s total population (Languages and dialects in Saudi Arabia, 2020).

HA dialects can be subdivided into two types along the Bedouin-Urban axis, similar to many other Arabic dialects. Watson (2002) differentiated Bedouin dialects from Urban dialects in that Bedouin dialects tend to be more ‘conservative and homogenous’, and resist innovation in terms of pronunciation, while Urban dialects are more adaptive and show clear intra-dialectal differences based on age, gender, social status, and religion. Bedouin Hejazi Arabic (BHA) is spoken more in the southwestern cities of the western region, such as Taif, and is also known as Tihama (Ingham, 1971), whereas UHA is mainly spoken in the three main cities of the western region: Jeddah, Makkah, and Al Madinah. However, Alhazmi (2018) argued that these two different Hejazi dialects (UHA and BHA) are spoken side by side in the Hejazi region of SA. She found that the differences between these two forms are more social than regional in nature, similar to all other Saudi dialects which could be spoken in the western region as well. However, the focus of this study is UHA because the data collection was conducted mainly in Jeddah, where UHA is the original spoken dialect and most families who speak this dialect have settled in the city of Jeddah.

In the literature, the structure of UHA is described as a combination of the phonology and morphology of several Egypto-Levantine dialects, which have had a large influence on UHA, while its syntax and lexis are similar to those of the rest of the dialects in the Arabian Peninsula such as Najdi and Eastern dialects (Ingham, 1971; Omar, 1975).

As reported earlier, because Makkah City, the home of Islam, is located in Hejaz (western) region of SA, and it hosts pilgrimage traffic every year, the populations of the whole Hejazi region has been influenced by the dialects of other Muslims from different countries (Arab and non-Arab) who have come to perform pilgrimage activities, and some of whom have settled in the area (Omar, 1975). For all the foregoing reasons, it could be argued that UHA is the Arabic dialect with which all Arabic speakers are most familiar with.

Although there are many linguistic differences between Arabic dialects in all language domains (i.e. phonology, morphology, syntax, and lexicon), these varieties are structurally related to MSA (Saiegh-Haddad, 2004). According to Saiegh-Haddad and Haj (2018), in the lexical domain, Arabic lexis can be categorised as follows: a) *identical words*, which that have a similar phonological form in MSA and the vernacular form (e.g. in UHA and MSA /batˤtˤiːx/ 'watermelon' and in UHA and MSA /kursiː/ 'chair'), b) *cognate words*, which are phonologically related and gave show some overlapping phonological forms when used in each of UHA and MSA. The phonological differences in cognate words are either in consonantal or vocalic phonemes or syllabic structure, or in both (e.g. UHA /dahab/ versus MSA /ðahab/ 'gold'; UHA /ɣara/ versus MSA / ɣaraʔ/ 'glue' are examples of consonantal phoneme change, and UHA /toːb/ versus MSA /θawb/ are examples of consonantal and vocalic phoneme change, and UHA /dˤahir/ verses MSA /ðˤahr/ are examples of syllabic structure change). c) *unique words*, which have a unique form in MSA that differs entirely from that used in the vernacular variety (e.g. UHA /ʃubbaːk/ versus MSA /nafiða/ 'window' and UHA /xu.ʃum/ versus MSA /ʔanf/ 'nose'). These differences in unique words can be attributed to many variables, such as historical use of language, social status, and many other factors which are beyond the scope of this study. Similarly, in the phonology domain Arabic phonemes could be categorised into three classes which are reported in the following section.

The next section presents a detailed description of the phonology of UHA, including a comparison of its phonological characteristics with those of MSA and some other Arabic dialects (i.e. Egyptian, Jordanian, Syrian, Kuwaiti, and Najdi Arabic). These Arabic dialects are included in the comparison because, as reported earlier, the phonological acquisitions of children speaking these dialects has been investigated in well- cited studies, to the best of the researcher's knowledge, while this is not the case for UHA (see Section 2.2). The comparison will present the full inventory of the consonants and vowels in UHA, MSA and these dialects as well as some of their distinctive phonological aspects in terms of syllable patterns and word stress. The main reason for a such a comparison is to understand the phonological system of UHA in comparison to MSA and the other Arabic dialects. Such an understanding is a prerequisite for examining the phonological development of children speaking UHA.

2.1.3 Phonology of Arabic

Like other Semitic languages, Arabic (CA and MSA) is marked by its rich of consonantal system and limited vocalic system (Watson, 2002). As noted earlier, MSA retains the basic phonological system of CA. It has 28 consonantal phonemes (two of which are semivowels) in nine places of articulation, and very small vowels inventory which includes only three short vowels (/i/, /a/, and /u/) and their long versions (/iː/, /aː/, /uː/), as well as two diphthongs (/ai/, /au/). Naser (1993) documented that consonant in Semitic languages, including Arabic, carry the main semantic content of the utterance, and vowels are used in combination with them to modify this content. Similarly, UHA has more consonants than vowels, with 25 consonantal phonemes and eight monophthongal vowels (Basalamah, 1990). UHA is also similar to MSA in having the consonant co-occurrences structure in their words, where consonants occur either as a word-final consonant cluster (CC), heterosyllabic CC (i.e. word-medial consonant sequence), or as gemination (i.e. word medial consonant doubling). A full description of UHA phonology is provided below, and a comparison with MSA and the other dialects is given for each feature.

2.1.3.1 Consonantal system of Arabic

As reported earlier, the 27 UHA consonantal phonemes may belong to one of the following categories according to Saiegh-Haddad and Haj (2018):

a) Only–UHA phonemes, which are used in the Hejazi dialect but are not within the phonemic inventory of MSA; this category includes two phones /ɡ/ and /zˤ/.
b) Only MSA phonemes, which are not used in UHA; this category includes interdental fricatives: voiced /ð, ðˤ/ and voiceless /θ/, and the uvular /q/.
c) Shared MSA and UHA phonemes; this category includes 25 consonants.

MSA includes 28 consonants, four of which are replaced by other consonants in UHA (see Table 2.1).

Table 2.1 *Consonantal Variants between UHA and MSA*

UHA	t/ s	d/ z	zˤ/ dˤ	ɡ
MSA	θ	ð	ðˤ	q

It is important to note that the consonantal phonemic inventory of UHA consists of 27 phonemes, but its consonantal phonetic inventory includes 28 phones. The MSA consonant phonemic inventory has 28 phonemes, while its consonantal phonetic inventory includes 33 phones (Alghamdi, 2015; see Table 2.2). According to McCarthy (1994), all forms of Arabic have a full set of guttural consonants (/ʔ, h, ħ, ʕ, χ, ʁ/, and /q/) as well as a set of 'emphatics' or 'pharyngealized' coronal consonants (/tˤ, dˤ, sˤ, ðˤ/) which are unique characteristics of Arabic consonants. McCarthy (1994) defined this set of guttural phonemes as 'the consonants that are produced with a primary constriction in the posterior region of the vocal tract' (McCarthy, 1994, p. 12). This is defined as a broad region that encompasses the area from the larynx to the oropharynx. Similarly, McCarthy described the 'emphatic' consonants as those produced with a secondary articulation, in which the root of the tongue is retracted towards the back wall of the pharynx. Table 2.2 summaries the UHA and MSA consonant phones. The phones that exist only in UHA are given blue font, while those that exist only in MSA are given in red font. The shared consonants are presented in black font, illustrating the phonetic system of these two forms.

Table 2.2 *MSA and UHA consonant inventories (Al Oufi, 2016; Alghamdi, 1998; Holes, 2004; Watson, 2002)*

Manner of Articulation	Place of Articulation																			
	Bilabial/ Labial		Labio–dental		Dental		Alveolar		Post alveolar		Palatal		Velar		Uvular		Pharyngeal		Glottal	
Voicing	–	+	–	+	–	+	–	+	–	+	–	+	–	+	–	+	–	+	–	+
Plosive		b					t	d			ɟ		k	g	q				ʔ	
Emphatic							tˤ	dˤ												
Nasal		m						n												
Trill								r												
Tap or Flap								ɾ												
Fricative			f		θ	ð	s	z	ʃ	ʒ			x	ɣ	χ	ʁ	ħ	ʕ	h	
Emphatic						ðˤ	sˤ	zˤ												
Affricative										d͡ʒ										
Lateral approximant								l												
Glide approximant	w										j									

Key: (–) voiceless, (+) voiced, MSA only (in red), UHA only (in blue), shared consonants (in black)

Comparing the consonantal inventory of UHA and MSA with respect to the manner of articulation clearly shows that they share the same divisions of manner of articulation. In terms of place of articulation, UHA and MSA share the same classification, with two differences: the dental set is missing in UHA, while the uvular set of MSA is replaced in UHA by a velar set (Basalamah, 1990: see Table 2.2). Two phonemes that exist in both UHA and MSA have allophonic variation:

/r/ → the tap or flap alveolar [ɾ] and the trill [r]

/ʒ/ → the affricative postalveolar [dʒ] and the fricative postalveolar [ʒ]

In terms of the five other Arabic dialects, they all maintain the use of bilabial /b, m/, labiodental /f/, alveolar /t/, /d/, /n/, /s/, /z/, /l/ and their emphatic (pharyngealised) counterparts /tˤ, sˤ/, pharyngeals /ħ, ʕ/, glottal /h, ʔ/ and the glides /w, j/. They also have /x/ and /ɣ/ in their consonantal inventory either with a velar or uvular classification (see Table 2.3). There is some variation among the different dialects regarding the remaining MSA consonants. Table 2.3 illustrates the major phonemic variants of these consonants across the dialects.

Table 2.3 *MSA phonemic variation across the Arabic dialects*

MSA	Egyptian	Jordanian	Syrian	Kuwaiti	Najdi Saudi	UHA
q	ʔ	g/ʔ/k/ɢ	ʔ/ g	g/ ɣ	g	g
dˤ	dˤ	d/ðˤ	dˤ	ðˤ	ðˤ	dˤ
ʔ	deleted word medially	deleted word medially	deleted word medially	deleted word medially	deleted word medially	deleted word medially
r/ɾ	r	r/ɾ	r	r/ɾ	ɾ	r/ɾ
θ	t	t	t/s	θ	θ	t/s
ð	d	d	d/z	ð	ð	d/z
ðˤ	zˤ	dˤ	dˤ/zˤ	ðˤ	ðˤ	dˤ
ʃ	ç	ʃ	ʃ	ʧ/ ʃ	ʃ	ʃ
dʒ/ʒ	ɟ /g	dʒ/ʒ	dʒ/ʒ	dʒ	dʒ	ʒ
χ	x	χ	x	χ	χ	x
ʁ	ɣ	ʁ	ɣ	ʁ	ʁ	ɣ

As can be seen from the table, the phonetic system of some dialects is not identical to their phonemic inventory. For example, the Jordanian dialect has a larger phonetic system (i.e. 31 phones) than its phonemics system (i.e. 25 phonemes). In short, it could be concluded that UHA has 26 consonantal phonemes and 28 consonantal phones, while MSA has 28 consonantal phonemes and 32 consonantal phones.

Appendix A provides a detailed descriptions of the consonantal classification according to place of articulation in the different Arabic forms: UHA, MSA, and the other five Arabic dialects (hereafter ‘other dialects’).

2.1.3.2 Consonantal co–occurrences

In UHA, there are three types of consonantal co–occurrences: word-final consonant cluster (CC), heterosyllabic CCs, and gemination, all of which are restricted to two elements. Word-initial CCs are prohibited in UHA. In terms of consonantal co–occurrence, MSA is similar to UHA in two respects only, and also prohibits word-initial CCs. Although these types of consonantal co-occurrences are also found in other dialects, there is some variation in the CC position, which can occur in word-final or word-initial position. In general, all Arabic varieties do not allow consonants to appear in a cluster within words. Although word–internal consonants sequences do occur in multisyllabic Arabic words such as /mustaʃfa/, they are syllabified (e.g. /mus.taʃ.fa/), yielding a heterosyllabic CC (Saiegh-Haddad, Shahbari-Kassem, & Schiff, 2020). In the following sections, detailed descriptions are provided of these three types of consonantal co-occurrences in UHA, MSA, and the other dialects.

2.1.3.2.1 Word-final consonant cluster (CC)

In UHA, CCs are only permitted in word-final position, and are restricted to two elements (Jarrah, 1993). However, the available data on word-final CCs in UHA are not satisfactory and lacking a clear description of the permitted consonantal elements (cf. Al–Mohanna, 1998; AL Qahtani, 2014; Jarrah, 1993). What is known about word-final CCs in UHA is the specific phonological environments in which short vowels (anaptyctic vowels) are inserted between the two elements of a cluster in MSA words (see Table 2.4 for examples; Al–Mohanna, 1998; AL Qahtani, 2014; Alfaifi, 2019; Jarrah, 1993). Such a process is explained in the Arabic literature as an ‘epenthesis process’; however, it is not always a real epenthesis process which requires the insertion of a schwa /ə/. What is meant by epenthesis in the Arabic literature is the insertion of a vowel that destroys the cluster and changes the word/syllable structure completely of sonority scale principles (SSP,) which are based on ascending sonority to a peak (vowel) and descending towards syllable boundaries (codas) (Clements, 1990).

Concerning the SSP[3], Jarrah (1993) stated that vowel insertion in UHA is triggered by a word-final CC with rising sonority values in MSA nouns that have the canonical shape CVCC (see Table 2.4). AL Qahtani (2014) noted that some dialects, such as UHA and Najdi Arabic, permit the occurrence of vowel insertion to repair SSP violation in a syllable. Table 2.4 provides some examples of this SSP violation repairer strategy in UHA.

[3]Sonority scale: Vowels > Laterals >Trills > Nasals > Fricatives > Plosives (for more information on the sonority scale, see Parker (2012, p.321).
← more sonorous less sonorous →

*Mora is the minimal prosodic unit (μ) that organise sounds into units and is based on length. For more details about the prosodic units and mora, see Watson (2002, p. 128–133).

Table 2.4 *Examples of words with a final CC in MSA but not in UHA, where they are non–CC words*

MSA	Manner of CC	Sonority type	UHA	Glossary
/miʃtˁ/	F–S	raising	**/muʃutˁ/**	Comb
/nimr/	N–T	raising	**/ni.mir/**	Tiger
/raʔs/	S–F	raising	**/raːs/**	Head
/riʒl/	F–L	raising	**/riʒil/**	Leg
/ðˁahr/	F–T	raising	**/dˁahir/**	Back
/ʔuðn/	F–N	raising	**/ʔudun/**	Ear
/maqsˁ/	S–F	raising	**/ma.gasˁ/**	Scissors

Key: L= Lateral, T= Trill, N= Nasal, F= Fricative, S= Plosive

In a study conducted by Alfaifi (2019), adult Hejazi speakers tended to maintain MSA's word-final CCs without epenthesis when their phonetic segments have falling or plateaued sonority[4] (see Table 2.5).

Table 2.5 *Examples of MSA word-final CCs that are preserved in UHA according to sonority type*

MSA	Manner of CC	Sonority type	UHA	Glossary
kalb	L–S	falling	kalb	Dog
gird	T–S	falling	gird	Monkey
ʃams	N–F	falling	ʃams	Sun
milħ	L–F	falling	milħ	Salt
waʒh	F–F	plateau	waʒh	Face

Key: L= Lateral, T= Trill, N= Nasal, F= Fricative, S= Plosive

In terms of the type of vowel insertion in word-final C_1C_2, Jarrah (1993) described three rules for inserted vowels in UHA, where the inserted vowel is determined by the lexical vowel and the second consonant (C_2) in the word-final cluster. First, if the lexical stem vowel is either /i/ or /u/, then vowel harmony occurs, as in /ʔuðn/ ('ear'), which pronounced as [ʔuðun], where the inserted vowel is like the lexical stem vowel (see Table 2.5). Second, if the lexical stem vowel is /a/ then the epenthetic vowel that inserted after C_1 depends on C_2. If C_2 is /l/, then the epenthetic vowel is [i], as in /ħabl/ which is articulated as [ħab**i**l]. Third, if C_1 is /r/, the epenthetic vowel is [u], as in /tamr/ which is articulated as [tam**u**r]. Another condition that triggers epenthesis of word-final CC is the suffixation of a word. For example, in a word like /kalb/ ('dog'), after adding a suffix to the word, as in /kal.ba.ha/ ('her dog'), vowel insertion occurs after the second consonant in the cluster, resulting in the addition of another

[4] Raised sonority means that the coda's peripheral segment has a higher sonority value than the segment closer to the nucleus. Fall sonority is when the CC's peripheral consonant has a lower sonority value than the first consonant in a cluster. Plateaued sonority is when both consonants of a cluster have a similar sonority level (Alfaifi, 2019).

syllable to the word (Jarrah, 1993).

MSA is similar to UHA, where word-final CC is restricted to two elements. Al-Ani (1970) reported in his book all word-final CC possibilities, including 312 variants that are allowed in MSA. He documented that some consonants cannot occur in a cluster, including:

- /bf/, /bm, /fb/, /fm/ and /fh/.
- /mf/
- /nr/, /nl/, /rl/, /ln/, /lr/; and
- all the velar, uvular and laryngeal consonants /k, q, x, ɣ, ħ, ʕ, h/ which do not occur with each other.

For other Arabic dialects, as reported earlier, the position of the CC within the word varies among different dialects. For example, in Egyptian Arabic CCs are prohibited in word-initial or syllable–initial position and are only allowed word-finally with two elements and with no restrictions on what those two consonantal phonemes can be. This word-final cluster can be of any sonority sequence combination (falling, level, rising) (Ragheb & Davis, 2014). In Jordanian Arabic, CCs are allowed in word-initial and word-final position and like in MSA are restricted to two elements. In Kuwaiti Arabic, CCs are permitted in word-initial and word-final position and are limited to two elements only. All possible Kuwaiti consonantal combinations in a cluster were reported by Al Qenaie (2011, p. 71–83). Najdi Arabic is another dialect that permits an initial CC and a word-final cluster. AL Qahtani (2014) reported that most of the rules applied to CCs in UHA are similar to those for Najdi Arabic, especially the rules concerning vowel insertion in word-final CCs. In the Jordanian and Egyptian dialects, word-final CCs are restricted to two elements with some prohibited combinations, similar to MSA and UHA. Syrian Arabic is the only dialect that prohibits word-final and word-initial CCs.

2.1.3.2.2 Heterosyllabic CCs

Heterosyllabic CCs occurs in all Arabic dialects and are a common word shape in MSA. Some Arabic literature describes heterosyllabic CCs as consonantal sequencing (e.g. Dyson & Amayreh, 2000) or as medial true CCs (e.g. Alqattan, 2015; El-badarin & Bani-Yasin, 1993). In contrast, Al Qenaie (2011) argued that no Arabic dialect permits medial CCs (i.e. the true CCs); they are only found word initially and finally. All medial CCs described in the Arabic literature are heterosyllabic or abutting clusters that include consonants spreading over two neighbouring syllables.

Regarding UHA, similar to the case with word-final CCs, there is no available literature describing the permitted consonantal elements for heterosyllabic CCs (cf. Al-Mohanna, 1998; Jarrah, 1993); however, it is a common word shape in UHA (see Table 2.6). In contrast, Al-Ani (1970) reported all heterosyllabic CC possibilities for MSA, which include 657 variants. He noticed that heterosyllabic CCs are found more frequently than final clusters.

Table 2.6 *Some examples of heterosyllabic CCs in UHA and MSA*

Heterosyllabic CC	**Glossary**
ma**sʒ**id	mosque
ma**tˁb**ax	kitchen
ma**nd**iːl	wipe
bu**rt**uqaːl	orange
ʔu**sˁb**aːʕ	finger
ʔi**θn**eːn/ ʔi**tn**eːn	two
fu**st**aːn	dress

In terms of the other dialects, heterosyllabic CCs occur in all Arabic dialects. For example, in Jordanian Arabic, heterosyllabic CCs are reported to be the most frequent CC, while word-final CCs are the least frequent (El-badarin & Bani-Yasin, 1993). Amayreh and Dyson (1998) acknowledged this type of word shape (i.e. CVC.CV/C) in their study, where their word list included six medial consonants that are part of heterosyllabic CCs, which they termed 'consonant sequences'. However, Amayreh and Dyson (1998) clearly stated that such sequences are not clusters in Arabic, since one consonant terminates the syllable while the second initiates the following syllables. El-badarin and Bani-Yasin (1993) disagreed with Amayreh and Dyson (1998) instead considering the consonant sequence as a 'medial' heterosyllabic CCs. However, true medial CCs (e.g. the intervocalic coda cluster /bint.na/ 'our daughter') could occur in Jordanian Arabic in some words after the addition of certain affixes (e.g. /–na/ 'our'), and this is true for other Levantine dialects as well (such as Lebanese Arabic) (Ragheb & Davis, 2011). In Kuwaiti Arabic, Alqattan (2015) reported that words with both a heterosyllabic CC (i.e. CV**C**.**C**V) and the true medial CC (i.e. CV.CCV) occur among her participants, which indicates that Kuwaiti Arabic allows for true medial CCs in addition to heterosyllabic CCs. Although there is no direct description of heterosyllabic CCs in Najdi Arabic, AL Qahtani (2014) reported that syllables with a coda are allowed in Najdi Arabic, and syllable onset is obligatory. Such a description indicates that heterosyllabic CCs do occur in Najdi Arabic.

2.1.3.2.3 Gemination

Gemination is when two identical sounds co–occur at a word boundary (Saidat & Khlifat, 2019). It is one of the main characteristics of Arabic syllable structure and has been defined differently by various scholars. Al-Ani (1970), for example, described a geminated consonant as ~~t~~ involving prolonging continuants and the longer closure of plosives. On the other hand, Crystal (2008) defined gemination as a sequence of identical adjacent segments of a sound in a single morpheme. He clarified that a geminated consonant cannot be simply regarded as a ‘long’ consonants; the transcription used for it is, e.g. [–bb–]. In this thesis, Crystal’s definition is followed. However, there is a lack of literature on UHA in terms of the phonetic environment and position of occurrence of the geminated consonants. Recently though, some researchers have become interested in describing geminated consonants in UHA (e.g. Bokhari, 2020).

In general, in UHA, all sound classes can be geminates, except /r/ and /ʔ/ (Jarrah, 1993), and the geminated consonants can occur in intervocalic position (e.g. /tuf.faħa/ ‘apple’), but not in word-final position (Bokhari, 2020). Ingham (1971) reported that in UHA word-final geminates are reduced to single consonants in paused position (e.g. /ħagg/ ‘right’→ /ħag/ in isolation). In MSA, similar to UHA, all sound classes can be geminates (Khattab & Al-Tamimi, 2012) but in addition to the word-medial position, geminated consonants can occur in word-final position in monosyllabic words only (e.g. /rabb/ ‘Lord’) (Davis & Ragheb, 2014). Regarding other dialects, there is no agreement on the position of geminated consonants. Some authors have noted that word-final consonants could be geminated in other dialects (e.g. Lebanese, Syrian, Jordanian, and Egyptian Arabic). They claimed that a word-final geminated consonant could appear as a cluster (CVCC) (e.g. Davis & Ragheb, 2014), but others have argued that it is only a prolonged individual consonant in a coda position (e.g. Basalamah, 1990; Khattab & Al-Tamim, 2012; Saidat & Khlifat, 2019). Abu-Abbas, Zuraiq, and Abdel-Ghafer (2011) reported that for Jordanian Arabic, in addition to word-medial and word-final positions, geminated consonants, can occur in word-initial (e.g. /ddaːr/ ‘house’); word-initial geminated consonants have not been reported in any other Arabic dialect. In short, the medial geminated pattern is more frequent and salient than other positions in all Arabic dialects because it is used in the derivation of nouns of a profession from the triliteral verb templates (i.e. CVCːV (C) the causative form, e.g. /xabbaːz/) (Khattab & Al-Tamimi, 2012).

Furthermore, there is an important morpho–phonemic role that trigger gemination in Arabic words and needs to be highlighted, which is the insertion of the prefix /ʔal/, the definite article, before Arabic nouns. The lateral consonant /l/ in this definite article is assimilated before dental, alveolar, and postalveolar consonants, which means it is assimilated (geminated) to the initial consonant of the word. These consonants are /t, tˤ, d, dˤ, n, r, s, sˤ, z, ð, ðˤ, θ, ʃ, ʒ, l/. For example, /ʔal/ + /ʃam.ʕa / →/ʔaʃ.ʃam.ʕa/. On the other hand, the /l/ consonant in this article is pronounced when it is inserted into words initiated with any other consonants (bilabial, labiodental, palatal, velar, pharyngeal and glottal); for example, /ʔal/ + /ħu.sˤaːn/→/ʔal.ħu.sˤaːn/ without geminated the first consonant.

2.1.3.3 Vowel phonemes

As reported earlier, the vocalic system in Arabic is relatively impoverished compared to the consonantal system. The vowel inventory includes only three short vowels and their three counterpart long vowels. A full description of the vowels based on their classification in MSA and UHA follows.

Short vowels: The three short vowels in Arabic are the high–close front /i/, the high–rounded back /u/ and the low–open central /a/ (Al-Ani, 1970). These are present in UHA, MSA, and the other dialects. However, Al-Ani (1970) and Ingham (1971) described allophonic variants of these short vowels for both UHA and MSA. Table 1, Appendix B illustrates all the allophonic vowel variants and their environments in MSA and UHA.

Long vowels: MSA has three long vowels: /iː/, /uː/, and /aː/. UHA has similar long vowels, plus /eː/ and /oː/, which are the reflexes of the two MSA diphthongs /ai, au/. It is noteworthy that the vowel length in all forms of Arabic is a phonemic feature which differentiates meaning. For example, /ful/ 'Arabian jasmine' and /fuːl/ 'broad beans' are two different words that have the same consonants and vowels but with different vowel length. The difference between short and long vowels is not in the quality but in the doubled duration of the long vowel production (Alghamdi, 2015; Ryding, 2010). This view is supported by Almurashi, Al-Tamimi, and Khattab (2020), who confirmed that UHA monophthongal vowels are dynamic and that vowel duration is the most useful feature to differentiate between short and long vowels in Arabic. Al-Muhanna (2009) reported that in UHA if the final syllable in a word is an open syllable with a long vowel (CVː), this promotes the vowel–shortening process CVː→ CV. He claimed that such processes represent CV as the syllable template (see Table 2, Appendix B for more details about syllable structure).

According to Al-Ani (1970), the allophones of the high–front /i/ and /iː/, and the high–back /u/ and /uː/, vowels differ from their long counterparts in quantity alone, while the low–central short vowels /aː/ differs in both quantity and quality (see Table 2 Appendix B). Similarly, Ingham (1971) described the vocalic system of UHA as having a wide allophonic range and that the exact vocalic feature articulated is governed by the adjacent consonants. Table 2, Appendix C presents these long vowels and their allophonic variation in MSA and UHA according to the adjacent consonants. Although these allophonic variants of short and long vowels have been reported by Al-Ani (1970) and Ingham (1971), no other study has reported these variants, instead other studies mostly focusing on broad vocalic transcription while ignoring allophonic variation (e.g. Al-Muhanna, 2009; Alghamdi, 2015; Basalamah, 1990; Jarrah, 1993).

According to Ingham (1971), UHA has no diphthongs, and that the diphthongs in MSA are replaced with two long vowels /eː/ and /oː/, which are not included in the MSA vowels inventory. However, some words in UHA have diphthongs such as:

- /au/ in /tˤ**au**la/ "table", /ʒ**au.**waːl/ "cell phone", /r**au**dˤa/ "kindergarten", and
- /ai/ in /muk**ai**jif/ "air conditioner", /l**ai**muːn/ "limon",

These diphthongs have never been described in the UHA phonological literature; where most Hejazi phoneticians transcribed the diphthongs /au/ in as [aw] where the /w/ is treated as a consonant (e.g. /fa.rau.la/ ('strawberry') →[fa.raw.la]).

The vowel inventories for the other Arabic dialects mainly include the six monophthongal vowels of MSA: /a, i, u, aː, uː/ and /iː/, with no diphthongs or triphthongs. They all replace the MSA diphthongs /ai/ and /au/ with the monophthong vowels: /ai/→ /eː/ or /ɛː/, and /au/→ /oː/. In most of the dialects, vowels can be affected by the adjacent consonants. For example, in all dialect, except for Egyptian Arabic, the vowel inventories include the back-low /ɑ/ as an allophonic variant of the central-low vowel /a/ near emphatic consonants. Table 3 in Appendix B presents all of the monophthongal and diphthongal vocalic variants across the Arabic dialects. Although Ingham (1971) reported many allophonic variations for UHA vowels, it is important to note that no other research has reported such variants as vowels usually have received very little attention from Arabic linguists. Therefore, it cannot be concluded that adult SHA–speakers still use these vowels in allophonic variations.

2.1.3.3 Phonotactic restrictions

UHA and MSA have identical syllable structures (see Table 2.7). In both forms, the syllable structure pattern is described using the formula: C_1–V–$C_{0\text{–}2}$, where the onset must have only one consonant and the coda may include up to two consonantal segments. The vowel could be either long or short. In some of the Arabic literature, the word shape with a long vowel such as CVːC is written as CVVC, where VV stands for a long vowel (cf. Al-Ani, 1970; Al-Muhanna, 2009; Ryding, 2010) and therefore these linguists view Arabic as having five syllable patterns. However, the long vowel VV is always considered as a monophthong and never consists of a two–vowel sequence (Al-Ani, 1970; Jarrah, 1993). Similarly, regarding the geminated coda, which is allowed in some Arabic dialects such as Levantine, the prolonged consonant is also described as a CC. Khattab and Al-Tamimi (2012) described the long vowels and long consonant (i.e. geminated consonant) in Arabic syllables using the IPA suprasegmental symbol for long phones (e.g. CVːC instead of CVVC, CVCːV(C) instead of CVCCV(C), and CVCː instead of CVCC for the long/geminated consonants). The current study is adopting Khattab and Al-Tamimi's (2012) IPA description (CVːC, CVCː, …).

The most common syllable structures in UHA and MSA is CV, which does not occur in monosyllabic words, and can occur freely in any word position (initial, medial, and final). The least common syllable structures are CVːC and CVCC (Khattab & Al–Tamimi, 2012), where CVCC is the only syllable that is not allowed in word-initial position. Al-Muhanna (2009) reported that the open syllable CVː in UHA appears word medially or initially but never finally. Moreover, Saiegh-Haddad et al. (2020) reported that Arabic words in general are phonologically simple as they mostly consist of $C_1VC_{1–0}$, even in multisyllabic words, which are long but have simple phonological structures.

Table 2.7 *Syllable structures and weight in MSA and UHA*

Syllable weight	Stress position (*in bold*)		Examples (identical words)	Glossary
Light	ˈ**CV**.CV.CV	antepenultimate	/ˈ**ba**.ga.ra/	Cow
Heavy	CV.ˈ**CV**ː.CV	penultimate	/za.ˈ**ra**ː.fa/	Giraffe
	ˈ**CVC**.CVC	penultimate	/ˈ**dˤif**.dˤaʕ/	Frog
Superheavy	CV.ˈ**CV**ː**C**	final	/xa.ˈ**ru**ː**f**/	Sheep
	CVC.CV.ˈ**CV**ː**C**	final	/baɣ.ba.ˈ**ɣa**ː**n**/	Parrot

In the other dialects, the syllable structures of UHA and MSA (i.e. C_1–V–$C_{0–2}$) are almost wholly maintained in the Egyptian dialect (Ragheb & Davis, 2014). In Jordanian, Kuwaiti, and Najdi Arabic the syllable structures are much wider (i.e. $C_{1–2}$–V–$C_{0–2}$), where the vowel could be short or long (Alzaidi, 2014; Mashaqba, Al-Shdifat, Al Huneety, & Alhala, 2019; Watson, 2002).

In terms of word length, disyllabic words are more common in all Arabic dialects than monosyllabic ones, with the CVCV word shape the most commonly found pattern across all Arabic dialects (Khattab & Al-Tamimi, 2012). However, multisyllabic words are also common in all forms of Arabic. For example, Abdoh (2010) found that SHA-speaking children are likely to produce words with multisyllables. Because of 'inflectional suffixes and prefixes[5], the length of Arabic words may get changed according to the morphological structures affixed to words. It is noteworthy that Arabic, like other Semitic languages, is characterised by its 'root–and–pattern morphology' (Watson, 2002, p. 3), where the root is consists of three or (less commonly) four consonants and cannot stand alone, and the pattern consists of one or more vowels inserted to that root is the pattern (e.g. the root k–t–b has the broad lexical sense of 'writing' from which the word for 'book' is /ki.taːb/) (Ryding, 2005). Therefore, multisyllabic words are a common structure in all Arabic forms as bound–morpheme structures lengthening Arabic words up to five or more syllables (Abdoh, 2010).

Syllable weight is determined by vowel length and the presence of a coda (Abu Guba, 2018). In all forms of Arabic, syllable weight comprises three types: light, heavy and superheavy. Light syllables are always

[5] Inflectional suffixes and prefixes are affixed to words to indicate grammatical function, adding extra elements to the word. These elements are related to the word's meaning (e.g. tense, aspect, mood, negation, number, gender) For example, the suffixes /–na/ could be attached to the disyllabic word /ki.ta:b/ 'book'→ /ki.ta:.ba.na/ 'our book' which is a four–syllabic word (Shamsan & Attayib, 2015). In UHA, for example, the stem noun /madrasa/ 'school'→/mad.ra.sa.ti/ 'my school' or /mad.ra.sat.ha/ 'her school' or /mad.ra.sa.tu/ 'his school'. These examples show that the word shape changes according to the morphological structures.

open, heavy syllables are open and closed and superheavy syllables are closed or doubly closed (see Table 2.7 for illustration; Watson, 2011). These types of syllable weights, as well as the syllable position within the word, determine the stress pattern in the word (see the following section).

2.1.3.4 Stress assignment

All Arabic speakers exhibit word stress, which is predictable in terms of syllable structure and is sensitive to syllable weight, syllable position and vowel length (Watson, 2002). A general rule in stress application in UHA, as well as MSA, is that the heavy syllable (containing a long vowel ending in a consonant) is stressed. However, if the final syllable is superheavy and closed (CV:C or CVCC), it receives stress (Holes, 2004). In a word that has no heavy or superheavy syllable, the stressed syllable will be the first syllable (Hellmuth, 2013). These rules are similar to those reported by Jarrah (1993) who described the stress patterns of UHA as it is spoken in Al-Madinah. He summarised the stress location in HA, in general, in the three statements, which are reproduced below.

a) Stress a final superheavy syllable (e.g. /xa.ˈ**ru**ː**f/** ‘sheep’).
b) Otherwise, stress a penultimate heavy syllable (the second to the last syllable of a word, e.g. /za.ˈ**ra**ː.fa/ ‘giraffe’).
c) Otherwise, stress an antepenultimate heavy syllable (the syllable that comes two syllables before the last in a word, e.g. /ˈ**mak.**ta.ba / ‘library’).
d) Otherwise, stress the penultimate or the antepenultimate syllable, whichever is separated from the first preceding heavy syllable or (if there is none) from the beginning of the word by an even number of syllables (e.g. /mu.ˈ**dar.**ri.sa/ ‘female teacher’ or /ˈ**ba**.ga.ra/ ‘cow’).

In other words, in disyllabic words, the superheavy syllable with a longer vowel is the stressed syllable (e.g. CV.ˈCVːC, or ˈCVː.CV), and it is either the first or the second syllable. In words with similar vowel lengths, such as CV.CVC or CV.CV, the stress is applied on the penultimate syllable (Alzaidi, 2014; see Table 2.7).

Because stress in Arabic is predictable and depends on syllable weight, it has similar patterns in the other Arabic dialects, without significant differences among them. For example, the stress patterns in Jordanian Arabic are similar to UHA, where the stress falls on the penultimate syllable if it is a heavy syllable (e.g. /bin.ˈ**sa**ː.miħ/ ‘we forget’). Otherwise stress falls on the antepenultimate syllable (e.g. /ma.ˈ**ʕal.**la.mak/ ‘he did not teach you’). Like UHA, the final syllable bears stress if it has a long vowel or a final CC (e.g. /daˈ**rast**/ ‘I studied’). Egyptian, Syrian and Kuwaiti Arabic share similar patterns of stress to UHA (Abu Guba, 2018; Alqattan, 2015; Watson, 2002). Saudi Najdi is another dialect that shares a stress pattern with UHA, where Al Amro (2019) described similar stress pattern rules for Saudi Najdi to those reported for UHA by Jarrah (1993).

2.1.4 Summary of Arabic phonology

The most important aspects of the Arabic language are summarised in the following points:

- Arabic is a South Semitic language that falls within the Afro-Asiatic language family.
- Arabic has different forms described by Arabic linguists, including CA, MSA, and colloquial forms.
- CA is the form of Arabic based on the Quran, the Islamic holy book, and it is used today as the language of all Islamic speech and activities (e.g. prayers).
- MSA is the form of Arabic, originally a descendant of CA, that maintains its phonology, morphology, and syntax but uses only a subset of the syntactic structure of CA. It is used today as the language of Arabic literacy across the Arab world.
- The dialectal form of Arabic is the spoken form that differs considerably from MSA in terms of phonology, morphology, syntax, and lexicon. Each region in the Arab world speaks a dialect that differs from one to another.
- In SA, there are almost five different spoken dialects, and UHA is the one spoken in the western region of SA; it is the dialect of interest in this study.
- There are around 14 million people who speak UHA in SA, and the dialect children growing up in the western region of SA get exposed to first as their mother language before their exposure to MSA, which begins when they start school.
- The consonantal system of UHA is similar to MSA; both include a complete set of guttural /ʔ, χ, ʁ, ħ, ʕ, h / and /q/ and emphatic consonants /tˁ, dˁ, sˁ, ðˁ/. Their consonantal system differs in that UHA lacks the MSA dental phones /θ/, /ð/, /ðˁ/ where they replaced in UHA by /t/,/d/, /dˁ/, respectively.
- UHA consists of 27 consonant phonemes and 28 consonant phones, whereas MSA has consonants 28 phonemes consonants and 33 phones.
- The vocalic system of UHA includes three short (/i/, /a/, /u/) and five long vowels /iː/, /aː/, /uː/, /eː/, /oː/ with two diphthongs (/ai/, /au/) (i.e. total of 10 vocalic phones), whereas MSA has the three short vowels, and only three long vowels (/iː/, /aː/, /uː/) with the two diphthongs (i.e. eight vocalic phones).
- UHA and MSA allow for three types of consonant co-occurrence: word-final CC, heterosyllabic CC, and geminations. Both prohibit the word-initial CC.
- The syllable structure in UHA and MSA can be described using the formula: C1–V–C2, where CV is the most common syllable structure.
- The stress pattern in UHA is similar to the MSA stress rules, where the final superheavy syllable attracts the stress. If there are no heavy syllables in a word, then the stress falls into another predictable location.

The following section will review the most frequently cited studies on phonological acquisition in monolingual Arabic-speaking children, focusing on the five dialects of Egyptian, Jordanian, Kuwaiti, Syrian and Saudi Najdi Arabic because, as reported earlier, most of the well-cited studies have focused on these dialects.

2.2 Phonological development of Arabic

A review of normative studies of phonological acquisition in monolingual Arabic-speaking children was conducted. Although this current project focuses on the phonological development of children speaking Saudi Hejazi Arabic, the review included studies on children speaking other Arabic dialects because of the limited information available about UHA. Therefore, this review presents the knowledge on different dialects first, followed by reviewing the little information available about the phonological development in UHA. The selected studies comprise a mix of unpublished theses, journal articles, and assessment manuals, with a mainly cross-sectional design addressing the development of Arabic speech sounds. All of these studies covered phonological development in Arabic-speaking children; however, the existence of multiple dialects across the Arab world (see Section 2.1) yields variation in the sequences and trajectories of phonological acquisition in children speaking these different dialects. To date, available research has focused on data from children speaking the following dialects:

- ***Jordanian Arabic*** (Al Khalayah, 1980; Amayreh & Dyson, 1998, 2000; Amayreh, 2003; Dyson & Amayreh, 2000; Hamdan & Amayreh, 2007; Mashaqba, Al-Shdifat, Al Huneety, & Alhala, 2019),
- ***Egyptian Arabic*** (Abou-Elsaad, Afsah, & Rabea, 2019; Abou-Elsaad, Baz, & El–Banna, 2009; Ammar & Morsi, 2006; Omar, 1973; Ragheb & Davis, 2011; Saleh, Shoeib, Hegazi, & Ali, 2007).
- ***Kuwaiti Arabic*** (Alqattan, 2015; Ayyad, 2009)
- ***Saudi Arabic*** (Al-Bader, 2009; Alawwad, 2009; Al-Sabi & Naqawah, 2013; Bahakeem, 2016)
- ***Syrian Arabic*** (Owaida, 2015).
- ***Qatari Arabic*** (Al-buainain, 2012).

Two of the studies cited above were excluded from consideration due to lack of accessibility (i.e. Bahakeem, 2016; Al Khalayah,1980). Consequently, the review included a total of 18 papers, five unpublished theses (PhD and Master's), nine journal articles, one book and two book's chapters, and one assessment manual. Table 2.8 presents a list of all studies included in this review, together with information on their scopes of analysis.

Table 2.8 *Summary of the Arabic studies on phonological development and their scoop of analysis*

Dialects	Focues / Authors	Source	Consonant		Vowel		Consonant cluster	Gemination	Phonological patterns
			Inventory	**PCC**	**Inventory**	**PVC**			
Jordanian	**Amayreh & Dyson, 1998**	**Article**	×	×					
	Amayreh & Dyson, 2000	**Article**	×		×				
	Dyson & Amayreh 2000	**Article**							×
	Amayreh, 2003	**Article**	×						
	Hamadan & Amayreh, 2007	**Article**	×						
	Mashaqba et al., 2019	**Article**							×
Egyptian	**Omar, 1973**	**Book**	×		×		×		
	Ammar & Morsi, 2006;	**Book Chapter**	×						×
	Saleh et al., 2007	**Article**	×						×
	Abou–Elsaad et al., 2019	**Article**							×
	Ragheb & Davis 2014	**Book Chapter**					×		
Kuwaiti	**Ayyad, 2011**	**PhD Thesis**	×				×		×
	Alqattan, 2015	**PhD Thesis**	×	×			×		×
Saudi	**Alawwad, 2009**	**Master Thesis**	×						×
	Al–Bader, 2009	**Master Thesis**							×
	Alsa'bi & Naqawa, 2013	**Assessment Manual**	×						
Syrian	**Owaida, 2015**	**PhD Thesis**	×	×	×		×		×
Qatari	**Al–Buainain, 2012**	**Article**							×

It is clear from Table 2.8 that 13 studies focused their analysis on the consonantal inventory either phonetic or phonemic, while 11 studies provided a description of phonological patterns. In contrast, only three studies addressed vocalic acquisition in their investigation, and no study had calculated the percentage of vowels correct (PVC). The most likely reason for not including vowels in many studies is a general perception that vocalic acquisition occurs at an earlier stage than consonantal acquisition, and that errors in use of vowels are rare among typically developing children (Dodd et al., 2003; Priester et al., 2011). Only five studies investigated CC acquisition in Arabic, and no study has examined the development of geminated consonants development in Arabic-speaking children.

The following section provides a full description of Arabic speech sounds acquisition (i.e. consonants, vowels, and CCs) across different dialects, highlighting the similarities and differences in terms of the phonetic and phonemic inventories of each dialect. Section 2.2.2 provides a review of all Arabic studies that addressed the phonological patterns in the speech of Arabic children, speaking different dialects. Furthermore, the limitations of these studies are described in each section, particularly data analysis and methodological limitations.

2.2.1 Speech sounds acquisition

In order to examine the speech sounds acquisition, Arabic researchers tested around 1,255 children speaking different Arabic dialects[6], aged 1;2 to 8;4 years between the years of 1998 and 2015. Based on a single-word naming assessment (Al-Sabi & Naqawah, 2013; Amayreh & Dyson, 1998; Ammar & Morsi, 2006; Ayyad, 2009; Hamdan & Amayreh, 2007; Owaida, 2015) or CS sample (Alqattan, 2015; Amayreh & Dyson, 2000; Saleh et al., 2007), these cross–sectional studies provide an overview of the consonantal acquisition in Arabic with a focus either on the phonetic aspect (Alawwad, 2009; Amayreh & Dyson, 2000) or phonemic aspect (Al-Sabi & Naqawah, 2013; Alawwad, 2009; Alqattan, 2015; Amayreh & Dyson, 1998, 2000; Ammar & Morsi, 2006; Ayyad, 2009; Hamdan & Amayreh, 2007; Owaida, 2015; Saleh et al., 2007). As reported earlier, the acquisition of vowels and CCs have rarely been reported in the Arabic phonological literature.

Interest in investigating Arabic phonological development began in 1973 with a descriptive study conducted by Omar. This study used a case study technique to investigate the linguistic development of 37 Egyptian Arabic-speaking children aged 6;0–15;0. Her description of the children's phonological system did not follow any criteria. Table 2.9 summarises the methodological and analytical characteristics of these 12 speech sounds acquisition studies, organised according to their publication year.

[6] These dialects are Jordanian (353 children), Egyptian (66 children), Kuwaiti (150 children), Syrian (160 children), and Saudi (526 children).

Table 2.9 *Summary of the normative studies on Arabic speech sound acquisition*

Authors	Year; Design	Dialect	Sample size	Sample age	Sample type (stimuli)	Type of analysis	C position	Phonemes sampled C	V	CC	Criterion
Omar	1973; Case study	Egyptian	Total = 37 typical No age groups	0;6–15;0	CS sample	Phonetic Inventory/ Vowel inventory	Not specified	Not specified			Not reported
Amayreh & Dyson	1998; Cross–sectional	Jordanian	Total= 180 typical Per age group= 20 9 age groups, 10 M, 10 F	2;0–6;4 (6–month intervals)	SW naming test (58 black and white line drawing)	PCC Age of C acquisition (Phonemes) Phonemic inventory	I, M, F	27	0	0	**For age– group level:** • 50% of children produce the sound correct in at least two positions for customary • 75% of children produce the sound correct in all three positions for acquisition • 90% of children produce the sound correct in all three positions for mastery **For each child level**: No data
Amayreh & Dyson	2000; Cross–sectional	Jordanian	Total= 13 typical One age–group 6 M, 7F	1;2–2;0	CS sample	Phonetic inventory (C & V)	I, SFWW, SIWW, F	Not specified			**For age– group level**: • C occur in at least two different words for each position in at least five children. **For each child level:** • C occur in at least three different words in the free sample in any position.
Amayreh	2003; Cross–sectional	Jordanian	Total= 60 typical Per age group= 30 15 M, 15 F in two age groups	6;6–8;4 (6;6–7;4, 7;8–8;4)	SW naming/reading test (a modified version of Amayreh, 1994 (80 words)	PCC Age of C acquisition (Phonemes) Phonemic Inventory	I, SFWW, SIWW, F	11	0	0	**For age–group level:** • 75% of children produce the sound correct in all possible occurrences (any position). **For each child level:** • C produced correctly in at least 75% of its total occurrences (any position).
Ammar & Morsi	2006; Cross–sectional	Colloquia Egyptian	Total= 36 typical Per age–group: Group 1: 5 M, 5 F= 10, Group 2: 13 M, 13 F= 26	3;0–5;0 (12–month intervals)	SW naming test (228 coloured pictures + concrete items)	Age of C acquisition (Phonemes)	I, M, F	25	0	0	**For age–group level:** • C produced correctly in at least 90% of responses (for mastery), or 50–89% of responses (for customary). **For child level:** No data
Saleh et al.	2007; Cross–sectional	Cairene, Egyptian	Total= 30 typical Per age–group=10 Group 1: 5 M, 5 F Group 2: 6 M, 4 F Group 3: 5 M, 5M	1;0–2;6 (6–month intervals)	CS sample	Phonemic inventory	I, M, F	Not specified			**For age–group level:** • C produced by five or more out of ten children in each group (any position). **For each child level:** No data
Hamdan & Amayreh	2007; Cross–sectional	MSA	Total= 100 typical One age–group= 50 M, 50 F	(M=6;4) School–age (1st grade)	SW naming test (a modified version Amayreh, 1994; 65 picture words)	Age of phonemes acquisition	I, M, F	28	0	0	**For age–group level:** • C produced by five or more out of ten children in each group (any position). **For each child level**: No data

Alawwad	2009; Cross–sectional	Najdi Saudi	Total= 20 typical Per age group= 10 Group 1: 3 M, 7 F Group 2: 4 M, 6 F	3;0–4;0 (6–month intervals)	SW naming test 165 words (pictures + objects).	Phonetic Inventory (words level)	I, M, F	27	0	0	**Phonetic:** **For each child level:** • C occurs at least twice in a child sample (any position) **For age–group level:** • C correctly occurs 75% of the time in a group.	**Phonemic:** **For each child level:** • C occurs correctly in all in positions 90% of responses for mastery. • C occurs correctly in all positions in 75–89% of responses for customary. **For age–group level:** • % Of each phoneme added then divided by 10. (% are like child level)
Ayyad	2011; Cross–sectional	Kuwaiti	Total= 80 typical Per age–group= 43 and 37 Two age groups, 38 M, 42 F	3;10–5;2 (M=4.4) (3;10–4;5, 4;6–5;2)	SW naming test (88 coloured pictures)	Whole word matching Phonemic inventory Word shape	I, M, F	32	0	33	**For age–group level:** • 90% of children produce the sound correctly (mastery) (any position) • 75–89% of children produce the sound correctly (customary) (any position) **For child level:** • C produced correctly with no more than two mismatch in a child sample (any position)	
Alsa'bi & Naqawa	2013; Cross–sectional	MSA	Total= 506 typical* 15 age groups, 323 M, 183 F 352 atypical	3;0–8;0 (4–month intervals)	SW naming test (51 words elicited from 38 pictures)	Age of C acquisition (Phonemes)	I, M, F	28	0	0	Not reported	
Alqattan	2015; Cross–sectional	Kuwaiti	Total= 70 typical Per age–group= 10 seven age groups, 5M, 5F	1;4–3;7 (3–month intervals)	CS sample	PCC Age of C acquisition Phonemic inventory	I, M, F	Not specified			**For age–group level:** • 90% of correct responses by >5 children in a group (mastery) (any position) • 75% of correct responses by >5 children in a group (acquired) (any position) • 50% of correct responses by >5 children in a group (customary) (any position). **For child level:** • At least one production of C (any position).	
Owaida	2015; Cross–sectional	Syrian	Total= 160 typical Per age group= 20 Eight age groups, 10M, 10F	2;6–6;5 (6–month intervals)	SW naming test (60 coloured pictures) + spontaneous speech (50 words)	PCC /PVC Age of C /V acquisition (Phonemes)	I, M, F	28	8	7	**For age–group level:** • 90% of children produce the sound correctly in at least two positions for mastery. • 75% children produce the sound correctly at least two positions for acquisition. **For child level:** No data	

Key: **SW**: Single word; ***MSA***: Modern Standard Arabic; **NR**: Not reported in the study; **I, M, F**: Initial, medial, and final position in the word **C**: consonants, **V:** vowels, **CC**: consonant cluster. ***The size of each age–group was not equal in this study. **PCC**: Percentage of consonants correct. **CS**: connected speech. ***‡* Late C:** are the consonants reported to not acquired by children aged 6;4 in the previous study of Amayreh and Dyson (1998). Grey cells show studies that used a CS sample as their method to elicit the speech sample.

It is clear from Table 2.9 that no study investigated the speech sounds acquisition of SHA-speaking children. Only two studies investigated the speech sound acquisition in the Saudi population (Al-Sabi & Naqawah, 2013; Alawwad, 2009), however; Al-Sabi and Naqawah (2013) focused on the MSA forms of consonantal phonemes, while Alawwad (2009) investigated Saudi Najdi Arabic-speaking children. The following section presents the most important findings of these reviewed studies, focusing first on the consonantal acquisition, including the percentage of consonants correct (PCC) and the consonantal co-occurrences inventories. Section 2.2.1.3 is reports the available findings regarding vocalic acquisition.

2.2.1.1 Consonantal acquisition

Three studies investigated early consonantal acquisition in children aged between 1;0 and 2;11 years (Alqattan, 2015; Amayreh & Dyson, 2000; Saleh et al., 2007) using CS to elicit their speech sample. Focusing on Jordanian Arabic, Amayreh and Dyson (2000) found that the phonetic inventory of 13 Jordanian Arabic-speaking children aged between 1;2 and 2;0 included /b, t, d, ʔ, h, m, n, l, w, j/. In addition, two fricatives /ʃ, ħ/ occurred in only five and six children, respectively, in the sample in at least two different words in a particular position. For Egyptians Arabic, Saleh et al. (2007) identified the phonemic inventory of 30 Egyptian Arabic-speaking children aged between 1;0 and 2;6, which included the phonemes that appeared in the speech of five children in this age group, i.e. /b, t, d, ʔ, s, h, m, n, l, w, j/, in agreement with Amayreh and Dyson (2000), except for /s/. No occurrence of any emphatic consonant was reported, except for /dˤ/, which only two children used out of the 30 participants. It is important to note that Saleh and her colleagues did not report any specific criterion for counting the phonemes at the level of the individual child. If produced correctly by only five children in a group, then the phoneme was included in the inventory. Investigating Kuwaiti Arabic, Alqattan (2015) found that Kuwaiti Arabic-speaking children acquired the plosives /b, t, d, k, ʔ/, nasals /m, n/, approximants /w, j/ and lateral approximants /l/, and fricatives /f, s, h/ by the age of 2;4–2;7 at 75% accuracy. Only two emphatics /tˤ/ and /sˤ/ were acquired at 3;0 years old. It is clear from these reviewed studies that Amayreh and Dyson (2000), Saleh et al. (2007), and Alqattan (2015) all agreed that the age of acquisition with 75% of accuracy for /b, t, d, ʔ, h, m, n, l, j, w/ is before the age of 2;6. The minor differences in the occurrence of some sounds (i.e. /k, s, ʃ, ħ/ and the emphatics / tˤ, dˤ, sˤ/) could be due to the differences in cut-off criteria used by each study.

Table 2.10 summarises the reported findings on the phonemic acquisition of Arabic based on a single-word naming tasks from participants speaking one of the five dialects (i.e. Jordanian, Egyptian, Kuwaiti, Saudi, and Syrian) aged between 2;0 and 8;0 to provide comparable results. This summary provides the age of acquisition at 75% accuracy and the age of phoneme mastery with 90% accuracy (for some of the studies). It is important to note that the age range acquisition information reported for Al-Sabi and

Naqawah (2013) reflects the age of acquisition for the MSA speech sounds without considering any dialect. Additionally note that Amayreh and Dyson (1998) presented in their study two ages of acquisition for each consonantal phoneme, one for the MSA phones and one for acceptable dialectal variation. Within Table 2.14, only the age of acquisition for the acceptable form is presented.

Table 2.10 *Age of acquisition for Arabic phonemes according to Arabic normative studies on different dialects*

Author	Amayreh & Dyson (1998)		Ammar & Morsi (2006)		Ayyad (2011)		Alsa'bi & Naqawa (2013)	Owaida (2015)	
Dialects	Jordanian		Egyptian		Kuwaiti		Saudi	Syrian	
Sample size	180		36		80		506	160	
Age range	2;0–6;4		3;0–5;0		3;10–5;2		3;0–8;0	2;6–6;5	
Criteria	75%	90%	50–89%	90%	75%	90%	NT	75%	90%
b	3;0–3;4	4;6–4;10	3;0–4;0	**	<3;10	3;10–4;6	3;0–3;2	<2;6	2;6–2;11
t	2;6–2;10	4;0–4;4	< 3;0	3;0–4;0	<3;10	3;10–4;6	3:0–3;10	<2;6	2;6–2;11
tˤ	8;4*	**	< 3;0	**	3;10–4;6	4;7–5;2	5;8–6;2	3;6–3;11	4;6–4;11
d	3;0–3;4	6;0–6;4	3;0–4;0	**	<3;10	3;10–4;6	3;0–3;1	2;6–2;11	3;0–3;5
dˤ	7;4*	8;4*	4;0–5;0	**	NA	NA	5;2–6;2	4;6–4;11	4;6–4;11
k	2;6–2;10	3;6–3;10	< 3;0	3;0–4;0	<3;10	3;10–4;6	3;7–4;2	3;6–3;11	4;6–4;11
g	2;6–2;10	3;0—3;4	< 3;0	**	<3;10	3;10–4;6	NA	NA	NA
ʔ	2;0– 2;4	3;0–3;4	< 3;0	3;0–4;0	<3;10	3;10–4;6	3;2–3;6	<2;6	2;6–2;11
m	2;0–2;4	3;6–3;10	< 3;0	3;0–4;0	<3;10	3;10–4;6	3;1—3;2	<2;6	2;6–2;11
n	2;6–2;10	2;6–2;10	< 3;0	3;0–4;0	<3;10	3;10–4;6	3;0–3;2	<2;6	2;6–2;11
f	2;6–2;10	3;0–3;4	< 3;0	3;0–4;0	<3;10	3;10–4;6	3;0	<2;6	2;6–2;11
s	5;0–5;4	**	< 3;0	**	4;7–5;2	**	5;7–6;2	3;0–3;5	3;6–3;11
z	6;0–6;4	**	>5;0	**	4;7–5;2	**	6;1–6;5	4;0–4;5	5;0–5;5
sˤ	6;0–6;4	**	3;0–4;0	**	4;7–5;2	**	5;9–6;3	4;0–4;5	5;6–5;11
ʃ	5;0–5;4	**	/ç/ < 3;0	/ç/ 3;0–4;0	3;10–4;6	4;7–5;2	5;10–6;4	3;6–3;11	5;6–5;11
x	4;6–4;10	5;6–5;10	< 3;0	3;0–4;0	3;10–4;6	4;7–5;2	5;3–5;6	3;0–3;5	4;0–4;5
ɣ	6;0–6;4	**	> 5;0	**	3;10–4;6	4;7–5;2	4;9–5;4	4;6–4;11	5;0–5;5
ħ	2;6–2;10	5;0–5;4	< 3;0	3;0–4;0	<3;10	3;10–4;6	3;6–3;8	<2;0	2;6–2;11
ʕ	7;4*	8;4*	3;0–4;0	4;0–5;0	3;10–4;6	**	4;8–5;2	3;6–3;11	4;6–4;11
h	5;0–5;4	**	< 3;0	3;0–4;0	<3;10	3;10–4;6	4;6–5;0	3;0–3;5	3;0–3;5
ʒ	4;0–4;4	**	NA	NA	/dʒ/ 4;7–5;2	/dʒ/ **	/dʒ/ 5;8–6;4	6;0–6;5	**
j	2;6–2;10	3;0–3;4	< 3;0	3;0–4;0	<3;10	3;10–4;6	5;5–6;0	2;6–2;11	3;0–3;5
w	2;0–2;4	2;0–2;4	< 3;0	3;0–4;0	<3;10	3;10–4;6	3;0	<2;0	<2;0
r	5;6–5;10	6;0–6;4	< 3;0	**	4;7–5;2	**	4;10–5;8	4;0–4;5	5;6–5;11
l	3;6–3;10	5;6–5;10	< 3;0	3;0–4;0	<3;10	3;10–4;6	3;0–3;1	<2;6	2;6–2;11
q	7;4*	8;4*	NA	NA	3;10–4;6	4;7–5;2	5;10–6;2	NA	NA
θ	6;0–6;4	**	NA	NA	4;7–5;2	**	5;9–6;4	NA	NA
ð	6;0–6;4	**	NA	NA	3;10–4;6	**	6;6–6;8	NA	NA
ðˤ	6;0–6;4	**	/ zˤ />5;0	/ zˤ / **	3;10–4;6	4;7–5;2	6;0–6;2	>6;5	**
Key	Very early: 2;0–2;11	Early: 3;0–3;11	Middle 4;0–4;11	Late: 5;0–5;11	Very late: > 6;0–6;4 *According to Amayreh (2003)	** sounds were not achieved at the mastery level by oldest age group.		-NA=phone is not within the dialect's inventory. -Phones in red are MSA phones	

It is clear from Table 2.10 that there are many similarities in the rate of acquisition reported by the reviewed studies. For instance, all plosives, nasals, and a few fricatives were reported as being acquired before the age of 4;0. The most important findings from each study in Table 2.14 are reported below.

Jordanian Arabic children acquired with 75% accuracy /t, k, ʔ, f, ħ, m, n, w, j/ by the age of 2;0–2;11. By 3;0–3;11, /b, d, l/ were added to their inventory; however, Amayreh and Dyson (2000) reported an earlier age of acquisition for the plosives /b, d/ and lateral /l/ by the age of 2;6 using a CS sample, whereas the back plosive /k/ was absent from their inventory. The affricate /dʒ/, fricative /θ, s, ʃ, h/ and trill /r/ were all acquired before age 5;11. Interestingly, Amayreh and Dyson (1998) found that all the Arabic emphatic sounds /tˤ, dˤ, sˤ, ðˤ/ were acquired after the age of 6;4, as well as the MSA plosive /q/, the sibilants /θ, ð, s, z, ʃ/, the affricate /dʒ/, and the back fricative /ʕ, h/. Further, they concluded that word-medial consonants were more accurately produced than those in either the word-initial or final position, with no differences found between the word-initial and final word positions. It is important to note that all medial consonants targeted in their study were in SIWW, except for /d/ and /x/ that were targeted in the syllable–final within word position.

In another study (see Table 2.9), Amayreh (2003) investigated the acquisition of these late consonants (i.e. /tˤ, dˤ, q, θ, ð, ðˤ, z, sˤ, ʕ, dʒ/) in 60 typically developing children aged between 6;6 and 8;4 years. He found that /dˤ, q, z / were acquired by the age of 7;4, the plosive emphatic /tˤ/ was acquired by 8;4, while /ðˤ, ð, sˤ, θ, dʒ/ were not acquired by the age of 8;4, which was the upper age limit in his study. Amayreh clarified that these consonants, especially the emphatic one may be too difficult to articulate for children younger than 8;4 years who produced these phonemes without their secondary emphatic feature, which would require a degree of ‘articulatory precision’ (Amayreh, 2003, p. 526). De Castro and Wertzner (2012) declared that the accurate production of a sound depends on the development of oral–motor control, which involves synergy between articulators from the lips to the pharynx and these movements require maturation of both cognitive and motor skills. For this reason, the emphatic consonants have a late age of acquisition among children speaking Arabic, which according to Amayreh (2003) is after the age of 8;0 for Jordanian Arabic. Further, he clarified that the late acquisition of the standard affricate /dʒ/ which continuously changed to the fricative /ʒ/ by children aged ≥7;0, could suggest that these two forms /dʒ, ʒ/ varied freely in children’s production indicating that children at this age do not perceive them as different phones.

For the acquisition of Egyptian Arabic, Ammar and Morsi (2006) reported a customary production (with 50–89% accuracy) of the Egyptian Arabic phonemes /b, d, ʕ, sˤ/ by the age of 3;0–4;0, whereas the phonemes /t, k, ʔ, m, n, w, j, l, f, ç, x, ħ, h/ were acquired before the age of 3;0. Note that they used a one–year age range for an age group; as reported in Chapter 1 Section 1.4.2.1.1, such an age interval does not allow for focused observation of speech sound development among children to discover any new additions to the inventory of that age group, which is especially important during the first three years of speech sound acquisition (Elrefaie, Hegazi, El-Mahallawi, & Khodeir, 2021; Watson & Scukanec, 1997). For the older age group 4;0–5;0, Ammar and Morsi (2006) found that the emphatic /dˤ/ was in customary production, and the back fricative /ʕ/ was the only phoneme added to the inventory of mastered phonemes for children in this age group. The remaining speech sounds were reported to be acquired after the age of 5;0 (i.e. /b, d, g, s, z, ɣ, r, tˤ, dˤ, sˤ, zˤ/). In agreement with Amayreh and Dyson (1998), Ammar and Morsi (2006) found that all the emphatic phonemes did not reach the mastery level of accuracy by the children in their sample. In contrast, they disagreed with Amayreh and Dyson regarding the word position for phonemes correctness. They found that the final word position was the most difficult, and the initial and medial positions do not differ significantly.

Kuwaiti Arabic and Syrian Arabic phonemes acquisition showed a similar trajectory to that reported for Jordanian and Egyptian Arabic regarding almost all plosives, nasals, the fricatives /f, ħ/, and the approximates /w, j/, according to Ayyad (2011) for Kuwaiti and Owaida (2015) for Syrian Arabic. However, Ayyad reported an earlier age of acquisition for the plosive /q/ (i.e. 3;10–4;6) and mastery at 4;7–5;2 as it is used more frequently by adult Kuwaiti Arabic speakers compared to speakers of the other Arabic dialects. Furthermore, the interdental /θ, ð/ and the emphatic /ðˤ/ were all acquired earlier by Kuwaiti Arabic-speaking children compared to the other dialects as these phonemes are part of adult Kuwaitis' phonemic inventory. Ayyad (2011) reported that these interdental phonemes /θ, ð, ðˤ/, as well as /z, sˤ, ʕ, dʒ, r/, did not reach the mastery level of production by the oldest age group she examined (i.e. 5;2 years).

In terms of Saudi studies, Al-Sabi and Naqawah (2013) found that the Saudi Arabic-speaking children acquired /b, t, d, ʔ, m, n, f, w, l/ by the age of 3;0-3;10. According to their data, the children acquired all MSA phonemes by the age of 6;4; unfortunately they did not clarify their criteria for "acquisition". It is important to bear in mind that these data were presented in the JAT[7] test manual.

The other data about the phonological development in Saudi Arabic-speaking children is provided by the master's degree thesis by Alawwad (2009) which focused on the Najdi dialect, as mentioned above. She found that by the age of 3;0–3;6, Saudi Najdi Arabic-speaking children mastered the production (i.e. 90% accuracy) of 17 phonemes /b, t, , k, g, ʔ, m, n, f, ħ, ʕ, h, w, j, l, r, tˤ/ and only six phonemes were produced with 75% accuracy, these are /s, z, sˤ, ʃ, x, ɣ/. Alawwad (2009) reported that Saudi Najdi Arabic-speaking children completed their phonemic inventory by the age of 4;0 where all Najdi phonemes were acquired with 75% accuracy. All the phonemes reached the mastery level of

[7] JAT= JISH Articulation Test, JISH= Jeddah Institute for Speech and Hearing.

accuracy,except for / θ, ð, ðˤ, s, ɣ, dʒ/. There were no significant differences between the three-word positions regarding the phoneme accuracy for both age groups. Alawwad’s results showed an earlier age of acquisition and mastery for most of the Arabic phonemes compared to all other reviewed studies; however, her sample size was limited to 20 participants which may affect the generalizability of her results.

Further, a study conducted by Abdoh (2010) focused on the phonological structure of the first words in 22 SHA-speaking children aged between 1;0 and 1;9. Although she documented the consonantal acquisition in these young children, Abdoh did not set any cut-off criterion for acquisition of Hejazi consonants and reported every single occurrence of the consonantal phoneme. The main focus of her study was describing the shapes and the syllable structures of the early words in SHA-speaking children aged between 1;0 and 1;9. She found that these children acquired the /b, t, d, ʔ, m, n, h, w, j/ consonantal phonemes at the age of 1;0–1;3, and all of the short vowels /a, i, u/. Abdoh (2010) noticed that SHA-speaking children often produced words with geminated consonants at this early age. In contrast, no words with final CCs were found at this age (1;0–1;3). By the age of 1;4–1;6, the consonantal phonemes /s, l/ were added to the inventory of the children and the long vowels /aː, iː, uː/ started to appear. An expansion in the inventory was observed by the age of 1;7–1;9, where the /k, ɡ, ʃ, z, ħ, ʕ/ consonants and /aː, iː, uː, oː eː/ were acquired. The emphatic consonants / tˤ, dˤ, sˤ, zˤ/ were all de–emphasised by children at this age (1;7–1;9), and word-final CCs started to occur in some children’s speech at this age.

Abdoh’s (2010) findings were in agreement with previous reports on the phonemic inventory of children aged 1;0–1;9 in other dialects (i.e. Egyptian, Jordanian, and Kuwaiti Arabic), in particular for /b, t, d, ʔ, m, n, h, l, w, j/. However, Abdoh reported that the acquisition of /s/ similar to that of Egyptian Arabic (Saleh et al., 2007) and Kuwaiti Arabic (Alqattan, 2015), while that of /k/ was similar to Kuwaiti Arabic. It is important to note that Abdoh documented the occurrence of consonantal phonemes among SHA-speaking children based on a single occurrence of each sound by at least nine of the ten children in an age group. Such a limited number of participants in the samples and the lenient cut-off criterion used may explain the inclusion of the phonemes /ɡ, ʃ, z, ħ, ʕ/ in the inventory of SHA-speaking children, unlike studies on the other dialects. These studies (i.e. Alqattan, 2015; Amayreh & Dyson, 2000; Saleh et al., 2007) set more strict criteria for a phoneme to be included in the inventory (e.g. the phoneme needs to occur in at least once in each word–position).

Abdoh (2010) also found that the syllable structure of words produced by SHA-speaking children included the formula $C_1VC_{0–2}$, and most of the words were disyllabic, with almost 15% of their production being multisyllabic words. Her analysis of onset and coda acquisition revealed that a considerable number of words began with a coda consonant in the syllable-final within word position, which indicated the occurrence of heterosyllabic CCs in the production of SHA-speaking children aged

1;7–1;9. However, she reported that many syllable onsets within words were deleted by children due to segmental difficulty (e.g. / ʔabɣa/→[ʔaba]), or they geminated this onset and assimilated the coda of the first syllable to the onset of the second syllable (e.g. /warda/→[wadda]).

It is notable that the age of acquisition varies slightly across these different dialects. For example, Al-Sabi and Naqawah (2013) reported that the sound /ʃ/ was acquired later than that reported in other studies (i.e. between the age of 5;10 and 6;4 years, depending on the gender of the child where girls acquired the sound earlier than boys) (see Table 2.10). Similarly, Amayreh and Dyson (1998) stated, based on data from 160 children, later age of mastery for consonants phonemes /tˤ, dˤ, q, θ, ð, ðˤ, s, z, sˤ, ʃ, ʕ, h, ʒ, ɣ/ which after the age of 6;4, the upper age limit of their study. In contrast, four studies (i.e. Ammar & Morsi, 2006; Ayyad, 2009; Owaida, 2015) reported an earlier age of acquisition for these emphatic sounds /tˤ, dˤ, sˤ/ (i.e. 5;0–5;6) and the pharyngeal fricative /ʕ/ (see Table 2.10). These differences could be due to several factors: first, the assessment tool used, and second, the methodological differences (e.g. sample size and their scoring method). However, dialectal differences should also be acknowledged as this could influence the frequency of the sound in adult speech and the syllable structure in which the sound occurs.

In general, Arabic-speaking children tend to acquire less complex and more salient sounds (e.g. /b, t, d, k, g, ʔ, m, n, ħ, l/) before other consonants (e.g. /tˤ, dˤ, q, θ, ð, ðˤ, s, z, ʒ, ɣ/). Still, the variants in the age of acquisition of /sˤ, ʃ, ʕ, h/ could reflected the influence of different methodology and dialects between reported studies.

To identify the age of acquisition of Arabic consonants from the data reviewed in the above section, the procedure used by Crowe and McLeod (2020) was followed, who reviewed the findings from 15 studies that examined the typical acquisition of English consonants in the United States. Crowe and McLeod (2020) used data from 15 studies to record the age of acquisition (in months) when 50%, 75% and 90% of children at each age group acquired the consonant's production. They extracted the age of consonantal acquisition data in one of two ways:

> "*1) If the data for the percentage of children who had acquired each consonant at each age were available in the study, then this was used to determine the age at which each consonant was acquired by 50%, 75%, and 90% of children.*
> *2) If the data for age of acquisition were presented only for predetermined criteria/criterion in the study, these criteria/criterion were used*" (p. 5).

Moreover, when these data were presented separately for subgroups of participants (e.g. males and females), they recorded the youngest age of acquisition.

Therefore, the mean ages of acquisition of Arabic consonants were identified from five studies, each describing the age of acquisition in a different dialect: Egyptian (Ammar & Morsi, 2006); Jordanian (Amayreh & Dyson, 1998); Kuwaiti (Ayyad, 2009); Saudi (Al-Sabi & Naqawah, 2013); and Syrian (Owaida, 2015). All these studies followed a similar methodology (see Table 2.9) in terms of their sample selection, elicitation task, and age range. Following the inclusion and exclusion criteria reported by McLeod and Crowe (2018), all these studies described Arabic singleton consonant acquisition, and some described the acquisition of consonant clusters (e.g. Ayyad, 2009). They all described the consonant acquisition in more than ten typically developing children, mostly monolinguals. Further, the children's age of acquisition data was reported when 75% of the children had acquired a phoneme and when 90% of the children had mastered a phoneme. Table 2.11 summarises the data on age of acquisition of consonantal phonemes with 75% and 90% accuracy criteria.

Table 2.11 *Mean age of acquisition and mastery for Arabic-speaking children across the five dialects according to age in years (Al-Sabi & Naqawah, 2013; Amayreh & Dyson, 1998; Ammar & Morsi, 2006; Ayyad, 2009; Owaida, 2015)*

Age groups		Arabic phonemic acquisition (75% accuracy)	Arabic phonemic mastery (90% accuracy)
<2;6–2;11 (24–35 months)	Very early	/t, k, g, ʔ, m, n, f, ħ, w/	
3;0–3;5 (36–41 months)	Early	/b, d, j, l, h/	/ ʔ, n, m, f, w/
3;6–3;11 (42–47 months)			/b, t, k, ħ, h, j, l/
4;0–4;5 (48–53 months)	Middle	/ s, ʃ, x, ɣ, ʒ /	/d /
4;6–4;11 (54–59 months)		/ sˤ, ʕ, r /	/ g, s, x, ɣ /
5;0–5;5 (60–65 months)	Late	/ tˤ, dˤ,θ, z/	/ sˤ, ʃ, r/
5;6–5;11 (66–71 months)		/ q, ð, ðˤ/	/ tˤ, dˤ, θ, ðˤ /
6;0–6;5 (72–77 months)	Very late		/q, ð, z, ʒ /

Note: MSA only, shared phones (dialectal and MSA)

Some studies did not sample one or two of the Arabic consonants. Most often omitted were the interdental consonants /θ ð ðˤ/ (Ammar & Morsi, 2006; Owaida, 2015) as these MSA sounds were not used by the adult speakers of the dialects examined in these studies (i.e. Egyptian and Syrian), like UHA. Therefore, the mean age of acquisition for these sounds classes was calculated using data from three studies only (Al-Sabi & Naqawah, 2013; Amayreh & Dyson, 1998; Ayyad, 2011). Although Alsa'bi and Naqawah (2013) collected their normative data from children mainly speaking UHA, they did not consider the dialectal forms of the MSA plosive /q/ or the dental /θ ð ðˤ/. In short, considering the mean age of acquisition and mastery reported by the reviewed studies, it could be concluded that the acquisition of Arabic phoneme consonants is completed by 5;11 years, and mastered by the age of 6;5. The results of the current project will be compared to this data.

2.2.1.1.1 Percentage of consonants correct (PCC)

PCC is another measure reported in Arabic studies to calculate the accuracy of consonantal production. For example, Amayreh and Dyson (1998), Owaida (2015) and Alqattan (2015) all reported the PCC at each age. They found that word-medial consonants were produced more accurately than those in word-initial or final positions. As stated earlier, Amayreh and Dyson (1998) calculated the PCC by considering both the MSA production and acceptable dialectal production. As expected, they found that children's PCC scores were higher for the dialectal form compared to the MSA form with a developmental trend across all age groups. Similarly, Owaida (2015) applied the PCC measure across each age group according to word positions (i.e. PCCWI for word-initial, PCCWM for word-medial, PCCWF for word-final position). She also found a significant increase in PCC across all age groups with age. Further, the pronunciation accuracy was significantly related to word position, where the word-final position was the most accurate (i.e. PCCWF= 88.22%) unlike the result reported by Amayreh and Dyson (1998) who found that the consonants in the medial position were the most accurate phones. Table 2.12 represents the PCC values calculated by Amayreh and Dyson (1998) and Owaida (2015). For Owaida's (2015) findings, the reported PCC is the mean number for the three PCCs (i.e. PPCWI, PCCWM, PCCWF), and for Amayreh and Dyson's (1998), only the overall PCC for the acceptable production is reported to ease the comparison.

Table 2.12 *Percentage of consonants correct (PCC) in Jordanian and Syrian Arabic-speaking children (adopted from Amayreh & Dyson, 1998; Owaida, 2015)*

Age group	Amayreh and Dyson (1998) %	Owaida (2015) % (SD)
2;0–2;4	52	NA
2;6–2;11	62	71.53 (2.95)
3;0–3;5	73	81.33 (3.79)
3;6–3;11	73	84.71 (2.79)
4;0–4;5	80	85.60 (1.69)
4;6–4;11	76	90.94 (2.04)
5;0–5;5	85	91.18 (2.49)
5;6–5;11	82	94.84 (1.67)
6;0–6;5	90	96.31 (1.53)

The overall PCC increased with age, reflecting a developmental trend in the consonantal accuracy. However, the PCC reported by Amayreh and Dyson (1998) had a lower score than that reported by

(2015). It is also evident that neither study documented any further information regarding how this measure was calculated or whether the calculation was done manually or using any software.

2.2.1.2 Consonant co-occurrence acquisition

2.2.1.2.1 Word-final consonant cluster (CC)

For word-final–CC, only two studies in Table 2.9 reported information on CC acquisition and both were done on Kuwaiti Arabic. In the first study, Ayyad (2011) found that word-initial CC words were not acquired by Kuwaiti Arabic-speaking children before the age of 4;0. She further reported that the criterion of 90% of accuracy had been met for the acquisition of only one CC /fl/ in /fluːs/ 'money'. Alqattan (2015) found that Kuwaiti Arabic-speaking children can produce CCs at the age of 2;0, with a higher level of occurrence in the word-final position. However, the percentage of instances of correct production of CCs was higher in the word-initial position. She concluded that children acquiring Kuwaiti Arabic show a dramatic increase in the number of clusters in all three-word positions between the ages of 2;4 and 2;7, which concurs with Ayyad's (2009) findings. It should be noted, however, that these two studies only reported the age of acquisition of CCs based on the complexity of the cluster, without further information about the specific age of acquisition of each CC.

Dyson and Amayreh (2000) observed that for typically developing Jordanian Arabic-speaking children, there are low percentages of final deletion or clusters reduction between the age of two and four years old. This observation indicates that Jordanian Arabic-speaking children can produce words with a word-final CC accurately by the age of 4;0. Omar (1973) also observed that Egyptian children acquire CCs at about the same age as the single consonants are acquired, between 1;5 and 5,0; however, she did not specify the position of the CC in terms of whether it was either word-final CC or heterosyllabic CC. It could be assumed that Egyptian Arabic-speaking children acquired CCs before the age of 5;0 in all possible positions.

Ragheb and Davis (2014) conducted a study on two monolingual Egyptian Arabic-speaking siblings aged 2;8 to examine the acquisition of final CCs. They found that the Egyptian children deleted the first consonant of a true word-final CC at the age of 2;8, and geminated the second, for example /kalb/-[kabː], and this occurred when the CC has falling sonority. However, if the final element in the word-final CC is pharyngeal, then some children tend to delete the pharyngeal consonant and geminate the first consonant in the cluster. Ragheb and Davis (2014) concluded that if the children did not acquire the elements of the word-final CC in target words, they tend to geminate one consonant as a repair strategy for the unacquired CC. It is important to note that they only assessed ten types of word-final CCs in two siblings speaking Egyptian Arabic dialect. Therefore, their results cannot be generalised to other Egyptian Arabic-speaking children, or even to other Arabic-speaking children in general.

2.2.1.2.2 Heterosyllabic CCs

In the Arabic literature, there is a lack of studies on the acquisition of heterosyllabic CCs; however, this is a common structure in Arabic words. As reported in Section 2.1.3.2.2, Amayreh and Dyson (1998) included six words with heterosyllabic CCs in their word list. However, they did not examine the acquisition of such clusters among Jordanian Arabic-speaking children. Their final results showed that children were more accurate at producing medial consonants than initial and final consonants. Amayreh and Dyson examined the consonants in the medial position as a single consonant without considering the syllable structure of such consonants even though they were part of the heterosyllabic CC. Similarly, Alqattan (2015) reported the occurrence of words with the CVCCV(C) structure among her participants; however, she considered CCs in the medial position as a true cluster that occurs as a syllable–initial cluster within the word. Thus, acquisition of heterosyllabic CCs among Arabic-speaking children has never been investigated.

2.2.1.2.3 Gemination

One study reported in Table 2.9 that did examine geminated consonants was conducted by Omar (1970). She observed and analysed the speech of 13 Egyptian Arabic-speaking children, and found that consonant gemination was acquired by the age of 3;6. She reported that the consonants acquired at an early age (e.g. /t, k, m, j/) were produced as geminated consonants more accurately (by the age of 3;0) than consonants acquired at a later age (e.g. /ɣ/), which were pronounced correctly by age 3;6. Similarly, Dyson and Amayreh (2000) found that the geminated feature was maintained even when a consonant was substituted (e.g. /batˤ.tˤiːx/ → [bat.tiːx]). Even in young children (aged 2;0), such a consonant sequence poses a more straightforward task than a sequence of dissimilar consonants (i.e. heterosyllabic CCs). (Dyson & Amayreh, 2000).

Another study conducted by Khattab and Al–Tamimi (2012) reported on the acquisition of gemination in Lebanese Arabic-speaking children. They investigated the developmental pattern of acquiring the shape and structure of early words in Lebanese children. The authors conducted a longitudinal study on five children aged between 1;1 and 1;6 at the first session. The data was rescored once every month until the children fulfilled the criterion of producing 25 different word types with around 50 words in their vocabulary. All children used disyllabic shapes with medial geminated consonants, even if the target words did not contain geminated consonants. In other words, they tended to add phonetic length to more than one segment in a word. For example, a target CVːCV (e.g. /baːba/) can be produced not with a long vowel only but also with a long onset of the second syllable (CVːCːV(C)), e.g. [baːbbah]).

Khattab and Al–Tamimi (2012) explained that because Arabic has prolonged vowel and consonant length, this might increase the salience of contrastive duration for children, leading to their extensive experimentation with segment length and the production of syllables with heavy rhymes and/or codas.

In general, gemination seems to be acquired by Arabic-speaking children at an early age (i.e. before the age of 4;0 at the maximum); however, the results of the reported studies may not be generalisable due to methodological differences. More studies are needed to investigate the acquisition pattern of geminated consonants in Arabic-speaking children.

2.2.1.3 Vowels

In terms of ***vowels acquisition***, Omar (1973) observed the emergence of a vowel inventory in two Egyptian children using a free speech sample. She found that the early vowels, including /a, i, u/, were acquired at the age of 1;5, /e, o/ were acquired at 2;3, and long vowels were acquired after the age of 3;0 (Table 2.13). It is important to note that Omar (1973) did not identify any specific criterion for the acquisition of any phone. Similarly, Amayreh and Dyson (2000) reported that by the age of 2;0, 12 vowels were produced by their participants. These included all six MSA vowels (the short /a/, /i/, /u/, and their long counterparts), as well as /e/, /eː/, /oː/, /ɔ/, /ɑ/, /ɑː/, and /ə/. They found that the open front /a/ was the most common vowel in their sample, whereas the schwa vowel /ə/ occurred least often. It is important to note that this vowel inventory includes a non–Jordanian Arabic vowel (i.e. /ɔ/) because the researchers reported all vowels produced by their participants. In another study, Owaida (2015) reported that Syrian children had acquired vowels with 90% accuracy mainly before the age of 2;6. Table 2.13 summarises vowel acquisition according to Omar (1973) and Owaida (2015) since these are the only studies to report the age of vowel acquisition.

Table 2.13 *Age of acquisition of Arabic vowels (Omar, 1973; Owaida, 2015)*

Vowels	**Omar (1973)**	**Owaida (2015)**	
Criteria	**One occurrence**	**75%**	**90%**
a	1;5	2;6–2;11	2;6–2;11
a:	3;0	2;6–2;11	2;6–2;11
u	1;5	2;6–2;11	2;6–2;11
u:	3;0	2;6–2;11	2;6–2;11
i	1;5	6;0–6;5	>6;5
i:	3;0	2;6–2;11	2;6–2;11
o	2;3	NA	NA
o:	3;0	2;6–2;11	2;6–2;11
e:	3;0	2;6–2;11	2;6–2;11
e	2;3	2;6–2;11	2;6–2;11

NA = The vowels are not in the inventory of this dialect.

It is noticeable that the short front close vowel /i/ has a much higher error rate than other vowels, according to Owaida (2015). These findings concur with that of Amayreh and Dyson (2000) who did not identify the vowel /iː/ among their sample. Owaida (2015) explained that adult speakers of Syrian Arabic replace /i/ with [e] or the schwa /ə/. She also added that in most words used to assess /i/ in her test items, /i/ appeared in weak syllables, which are frequently prone to omissions in younger Syrian children, as in /ħisˤaːn/ becoming [sˤaːn] 'horse', and /miftaːħ/ becoming [taːh] 'a key'. Unlike Omar's (1970) study, Owaida (2015) set a higher criterion for acquisition and mastery of each vowel, and such methodological variants could explain the different results between the two studies, especially in the different age of acquisition of the short vowel /i/. However, it still could be concluded from the reviewed studies that Arabic-speaking children master vowel production by the age of 3;0. More studies are needed to confirm this conclusion and provide more information about the acquisition of the short vowel /i/.

2.2.1.2.3 Percentage of vowels correct (PVC)

The only Arabic study to investigate the PVC among its participants is Owaida's (2015) study on Syrian Arabic-speaking children. She reported the PVC for each vocalic phoneme and calculated the mean of PVC for each age group (see Table 2.14).

Table 2.14 *Percentage of vowels correct for Syrian Arabic speaking children as reported by Owaida (2015)*

	2;6–2;11	**3;0–3;5**	**3;6–3;11**	**4;0–4;5**	**4;6–4;11**	**5;0–5;5**	**5;6–5;11**	**6;0–6;5**
PVC	89.22	90.86	93.17	91.65	94.60	92.11	94.37	98.79

Owaida's results revealed a developmental trend in vowel production accuracy which seems to be already acquired by children aged 2;6. In general, there is a lack of information on PVC scores for children speaking other Arabic dialects.

2.2.2 Phonological patterns

Nine cross–sectional studies included in this review describe phonological error patterns (Abou-Elsaad et al., 2018; Al-Bader, 2009; Al-buainain et al., 2012; Alqattan, 2015; Ammar & Morsi, 2006; Ayyad, 2011; Dyson & Amayreh, 2000; Owaida et al., 2015; Saleh et al., 2007). Table 2.15 summarises the phonological patterns that were identified in the speech of Arabic children aged between 1;0 and 6;5 in different studies.

Table 2.15 *Overview of the normative studies on Arabic phonological error patterns and the identified patterns in each study*

Dialect	**Egyptian**			**Jordanian**		**Qatari**	**Syrian**	**Kuwaiti**		**Saudi Najdi**	
Authors (year)	Ammar & Morsi (2006) *	Saleh et al. (2007) ***	Abou-Elsaad et al. (2019) **	Dyson & Amayreh (2000) **	Mashaqba et al. (2019) ***	Al-buainain et al. (2012) ***	Owaida et al. (2015) **	Ayyad (2011) **	Alqattan (2015) ***	Al-Bader (2009) *	Alawwad (2009)
Age range & sample size	3;0–5;0 N = 36 typical 22 atypical	1;0–2;6 N = 30 typical	2;0–5;0 N = 120 typical 30 atypical	2;0–4;4 N = 50 typical	1;0–3;0 N= 20 typical	1;4–3;7 N = 140 typical	2;6–6;5 N = 160 typical	3;10–5;2 N = 80 typical	1;4–3;7 N = 70 typical	2;0–3;0 N = 20 typical	3;0–4;0 N = 20 typical
Number of PP tested	8	15	16	11	8	13	14	13	15	16	14
Criteria for child–level	Once	Twice	Twice	Once	Once	Once	Occurs five times	Once	Once	PP occurs at least 20% of the time in the child's speech	PP occur at least 25% of the time in the child's speech
Criteria for age–group level	Occurs in 25% of children	No criteria reported	Occurs in >10% of children	Occur among 5% of children	No criteria reported	No criteria reported	PP occurs in >10% of children	No criteria reported	Occurs in two children out of 10 in a group. Age–appropriate errors occur	20 % of occurrences of each PP in all children	25 % of occurrences of each PP in all children
					Substitution						
Glottal stop replacement		X	X			X	X				
Stopping		X	X(fricatives)	X(fricatives)		X (not specified)		X (fricatives)	X (fricatives)		
Fronting		X	X (velar & palatal)	X (/k/ and /q/)		X (not specified)	X (velar)	X (velar)			
Backing			X (fricatives)			X (not specified)	X (not specified)	X (not specified)			
Gliding		X				X	X		X		
Lateralisation		X	X	X (of /r/)	**Not tested**		X		X (of /r/)		
De–emphasising	X			X		X	X	X	X	X	
Devoicing	X		X (post & prevocalic)	X (final)		X	X	X	X (postvocalic)		
/r/ deviation	X		X (lateralisation)			X	X (lateralisation)	X		X	
Dentalisation				X			X	X			
Other	Sibilant deviation			Initial voicing, denasalisation			Deaffrication	Spirantisation	Deaffrication	Deaffrication	
					Syllable structures						
Final C deletion		X		X	X	X	X (not specific)	X	X		
Cluster reductions		X	X	X	X	X		X	X	X	X

Syllable deletion	X (weak)	X	X	X (weak)	X	X (weak)	X
Other	Diminutive, reduplication Stridency deletion	Coalescence, Stridency deletion Gemination of consonant sequence	stridency deletion	Metathesis/Epenthesis Onset deletion Gemination of consonant sequence		Stridency deletion	Epenthesis
				Assimilation			
Regressive non–contiguous	X	X					
Regressive contiguous	X	X					
Progressive contiguous	X	X					
Progressive non–contiguous	X	X					
Total assimilation		X		X	X	X	X
Partial assimilation							
Others							

Keys: *Speech sample elicited by both coloured pictures and concrete items. **Speech sample elicited using picture naming task only. ***Speech sample elicited using connected speech only.

As shown in Table 2.15, a large number of phonological patterns were identified in the speech of monolingual Arabic-speaking children (see Appendix P for the patterns' definitions). However, it is important to note that most studies reported every instance of phonological variation from the adult form found in the speech of children that occurred once or twice. It should be questioned whether these instances of variations should be considered as phonological patterns (McReynolds & Elbert, 1981), or are they occurred by chance (Fox, 2006). Further, in a number of cases it was not specified which sounds were affected by a pattern. For example, Saleh et al. (2007) identified a fronting pattern among their participants, but they did not specify which specific sounds were fronted, either velar of palatal fronting; this limits comparability. It is also challenging to compare all of the Arabic normative studies in terms of phonological patterns because of the universal problem of terminology used to describe these patterns (see Chapter 1, Section 1.3.2.2.2). For example, Dyson and Amayreh (2000) reported that strident consonants are often deleted (e.g. /dˤif.dˤaʕ/→[duː.daʕ] 'frog') or substituted with a non–strident one (e.g. /sa.ma.ka/→[ta.ma.ka] 'fish'). However, the strident substitution was described as stopping by other researchers (e.g. Dodd et al., 2003; McIntosh & Dodd, 2008). For strident deletion, such as in the above example (/dˤif.dˤaʕ/→[duː.daʕ]), sometime it is described as 'compensatory vowel lengthening' (e.g. Grech, 2006) as the vowel being lengthened to substitute for a consonantal segment (Grech, 1998). Although 'stridency deletion' pattern reported in many other Arabic studies (e.g. Alqattan, 2015; Owaida, 2015) in addition to Dyson and Amayreh (2000), the compensatory vowel lengthening pattern was never reported to occur in the speech of Arabic children. Further, Dyson and Amayreh (2000) described fronting (e.g. /k/→ [t] and /q/→ [g, k, d, t]) and dentalisation patterns (e.g. /s/→ [θ]) as two distinctive categories, while in many other studies these two patterns were described as 'fronting' (e.g. Dodd et al., 2003; McIntosh & Dodd, 2008). Such distinctive definitions could lead to the misidentification of similar patterns across studies and hinder the universal comparison.

Assimilation is another pattern that could cause confusion for researchers because of the different definitions for this pattern across studies (e.g. Abou-Elsaad et al., 2019; Dyson & Amayreh, 2000), either because of its definition across studies, such as Abou–Elsaad et al. (2019), or because it is not considered in the analysis of patterns (e.g. Alqattan, 2015; Dyson & Amayreh, 2000). In their studies, Abou–Elsaad et al. (2019) and Saleh et al. (2007) considered Ingram's (1989) definition of assimilation, which refers to instances where a sound becomes like another sound in the word. For example, when a word like /tɑː.kol/ (she eats) becomes [tɑː.tol], then this error pattern is considered 'fronting' according to Dyson and Amayreh (2000) while Saleh et al. (2007) and Abou–Elsaad et al. (2019) identified it as 'assimilation'.

2.2.2.1 Types of phonological patterns

Three common substitution patterns were reported across almost all dialects: stopping of fricatives (except for Syrian), fronting (velar and palatal) and de–voicing, except for the Saudi Najdi. Glottal stop replacement, lateralisation, de–emphasisation and dentalisation were also frequently reported across Arabic dialects (see Table 2.15). For the syllable structure patterns, all studies documented two patterns: final consonant deletion and syllable deletion. CC reduction was also reported by all studies, except the Syrian study, as word-final CCs do not occur in the speech of adult Syrian Arabic-speakers (Owaida, 2015). Assimilation was also reported as a common pattern among all dialects; however, not all reported studies included this pattern in their analysis. It is essential to note that the primary goal of Mashaqba et al.'s (2019) study was to investigate lexical syllable structure and to identify phonological patterns that included this aspect. Because of this, no substitution process was reported in this study. Interestingly, the only systemic substitution patterns identified among Saudi Najdi Arabic-speaking children were de-emphasising, /r/ deviation, and deaffrication, and the only syllable structure pattern was cluster deletion, unlike other dialects. Such limited numbers of patterns could be explained by the strict cut-off criteria used by Alawwad (2009) and Al-Bader (2009) in terms of the frequency of pattern occurrences (20% or 25%of the time in the child's speech) and each group (25% or 20% of occurrences of each pattern in all children).

The only language–specific pattern to be reported for almost all dialects is de–emphasisation. Dyson and Amayreh (2000) explained that the de–emphasis of emphatic consonants is a simplification pattern that reflects a particular difficulty faced by children speaking Arabic. Specifically, they overcome the complex emphatic consonant production by merely ignoring the required secondary articulation, which is the pharyngealised feature. Weak syllable deletion and CC reduction are two other patterns that have been reported by all studies reviewed in this chapter. However, they have been considered insignificant in Najdi Saudi Arabic-speaking children (Al-Bader, 2009).

2.2.2.2 Age of occurrences

Four of the normative studies reported the ages at which each pattern occurred among typically developing Arabic-speaking children. Table 2.16 summarises the findings on the reported patterns and their ages of occurrence for six dialects: Jordanian, Saudi Najdi, Kuwaiti, Egyptian, Syrian, and Qatari, which were collected from eight studies (Abou–Elsaad et al., 2009; Al-Bader, 2009; Al-buainain, 2012; Alawwad, 2009; Alqattan, 2015; H. Ayyad, 2011; A. Dyson & Amayreh, 2000; Owaida, 2015). It is noteworthy that the two Kuwaiti and two Najdi Saudi studies targeted age groups on a continuum to each other; therefore, the results of these studies could be used to identify a wider age range of phonological development in children speaking these dialects, regardless of their methodological differences.

Table 2.16 *Age of occurrence of Arabic phonological patterns*

Phonological patterns		2;0–2;5						2;6–2;11						3;0–3;5						3;6–3;11						4;0–4;5						4;6-4;11						5;0–5;5					
Substitution	GRep					E	Q	J			S	E					S	E	Q																								
	Stopping	J		K		E	Q	J		K		E	Q			K		E	Q			K						K															
	Fronting	J		K	S	E	Q	J		K	S	E	Q				S	E	Q				S	E					S						S						S		
	Backing				S	E	Q				S	E					S	E																									
	Gliding				S	E					S						S						S						S						S						S		
	Lateralisation (including /r/ deviation)	J	N	K	S	E	Q	J	N	K	S	E	Q	J		K	S	E	Q	J		K	S	E		J		K	S	E				K	S	E*					S		
	De-emphasisation	J	N	K	S			J		K	S			J		K	S	E*		J		K	S	E*		J			S	E*		J				E*		J					
	De-voicing	J		K	S	E					S	E					S	E					S	E					S	E						E							
	Dentalization										S			J			S			J								K	S					K									
SS	CC-rd		N	K		E			N	K		E	Q		N*	K		E			N*																						
	FCD	J		K		E		J		K	S					K	S					K	S						S		Q												
	WSD	J			S	E	Q	J			S	E	Q			K	S	E	Q			K	S	E	Q			K	S														
Assmi.	Total assimilation					E	Q					E	Q					E	Q					E	Q					E													
	Partial assimilation					E																																					

Note: SS= syllable structure, Asmi= assimilation, GRep = Glottal replacement, CC-rd = consonant-cluster reduction, FCD = final consonant deletion, WSD weak syllable deletion, J: Jordanian Arabic (Dyson & Amayreh, 2000); N: Najdi Saudi Arabic (Al-Bader, 2009); N*; Najdi Arabic (Alawwad, 2009); K: Kuwaiti Arabic (Ayyad, 2011; Alqattan, 2015); S: Syrian Arabic (Owaida, 2015); E: Egyptian Arabic (Abou-Elsaad et al., 2019); E*: Egyptian Arabic (Ammar & Morsi, 2006); Q: Qatari Arabic (Al-buainain, 2012).See Appendix P for the patterns' definitions.

As shown in Table 2.16, lateralisation pattern was reported in all studies as being maintained until the age of 4;5, except for Najdi Arabic. It is noteworthy that for the other Najdi Saudi study, which targeted an older age group (3;0–4;0 years old), Alawwad (2009) used a higher criterion to identify a pattern among an age group (25%). Such a methodological difference could be the reason for this pattern not being reported among older Najdi Saudi Arabic-speaking children. It should be noted that some reported studies identified an /r/ deviation pattern, which describes any error with /r/ pronunciation, including the substitution of /r/ with [l] (Al-Bader, 2009; Alawwad, 2009; Alqattan, 2015; Ammar & Morsi, 2006), or the substitution of /r/ with [j] or [w] (Al-buainain, 2012; Alqattan, 2015; Ammar & Morsi, 2006). Such variation in definitions may affect the reliability of comparisons between different dialects. However, no study reported whether this pattern included the distortion of /r/ (i.e. substitute trill /r/ with alveolar[ɹ] or retroflex [ɻ]).

De-emphasisation was also reported to be suppressed after the age of 4;5, in children speaking all targeted dialects, except for Najdi Saudi. While Dyson and Amayreh (2000) indicated that this pattern could last longer, its occurrence was not expected after 4;5 years in children speaking Jordanian Arabic. Ammar and Morsi (2006) reported de-emphasisation in Egyptian children, but it occurred in almost 16% to 18% of the children in their two age groups, which are lower levels than their cut-off criterion (25%). However, Saleh et al. (2007) and Abou-Elsaad et al. (2019) did not show this pattern among their Egyptian sample.

Stopping was reported to disappear by the age of 2;11 in Jordanian children but lasted longer in Kuwaiti and Syrian children, where it was suppressed by the age of 4;11. Although stopping was shown to be an insignificant pattern in the Najdi Saudi study (Al-Bader, 2009), the author labeled the stopping pattern as ‘deaffrication’ when she considered the substitution of /dʒ/ by [d] as ‘deaffrication’, while this is a ‘stopping’ error pattern according to Smit (1993) because it would be considered as ‘ deaffrication’ if the result was [ʒ] not [d]. Similarly, Ammar and Morsi (2006) included the stopping pattern under a broader pattern named ‘sibilant deviations’, which includes the stopping of sibilants and fronting of fricatives. The occurrence of ‘sibilant deviations’ was reported to be significant among both age groups in their study (3-4 years and 4-5 years) (Ammar & Morsi, 2006).

Similarly, de-voicing was reported to dissolve early (by the age of 2;5) in Kuwaiti children (Alqattan, 2015), but it lasted up to the age of 5;5 in Egyptian children (Abou-Elsaad et al., 2019). Although Dyson and Amayreh (2000) reported the occurrence of a de-voicing pattern and highlighted that it is more common in word-final obstruents, they clarified that its occurrences was among the least frequent patterns to occur in their sample (occurring between 1% and 24% of the time).

Gliding was shown to be typical only for children speaking Syrian Arabic up to the age of 5;5; however, it was shown to dissolve by the ages of 2;5 and 1;7, in children speaking Egyptian and Kuwaiti Arabic, respectively. Gliding was never reported in the speech of Jordanian speaking children.

Assimilation, which was only shown to be significant in the speech of children speaking Egyptian Arabic, disappeared in the speech of children older than 4;5 years (Abou-Elsaad et al., 2019; Saleh et al., 2007).

In terms of syllable structure patterns, three patterns appeared in children's speech by the age of 2;0-2;6: CC reduction in Najdi Saudi, Kuwaiti, Jordanian and Egyptian, final consonant deletion in the speech of Jordanian, Kuwaiti and Egyptian Arabic speaking children, and weak syllable deletion in Kuwaiti, Jordanian, Syrian, Qatari and Egyptian Arabic. For Najdi Saudi Arabic, CC reduction was mainly reported for initial consonant clusters, while Al-Bader (2009) and Alawwad (2009) highlighted that word-final CCs are produced more accurately by Najdi Arabic -speaking children aged between 2;0-4;0. In contrast, Alqattan (2015) reported that most of the CC reduction patterns occurred in word-final CCs (Kuwaiti) compared to reduction in word-initial CCs. In the Jordanian dialect, CC reduction was reported by Mashaqba et al. (2019), who counted CC occurrences without differentiating between the cluster position within words (i.e. word-initial or word-final CC). Although Mashaqba et al. (2019) reported a dropping in the frequency of occurrence of CC reduction among their age groups, they still noted its occurrence in 10% of the oldest age group they examined (i.e. 2;7–3;0). In agreement with this finding, Dyson and Amayreh (2000) reported the occurrence of a sequence deletion pattern by the age of 2;0 until the age of 6;0 years, which included the occurrence of CC reduction. It is noteworthy that they also included under this label the heterosyllabic reduction pattern (e.g. /mad.ra.sa/→[madasa]). They also noted that the consonant sequence reduction pattern (abutting consonants) was produced less accurately than intervocalic clusters (e.g. word-final CC). However, they reported that they did not differentiate between heterosyllabic CCs and geminated consonants when they counted the occurrence of consonant sequence reduction. Dyson and Amayreh (2000) found that children reduced the heterosyllabic CC pattern even in the older age groups (i.e. 4;5–6;0). Mashaqba et al. (2019) reported that the most preferred word structure among Jordanian Arabic-speaking children was disyllabic words, where the CV(V)C.CV(C) word shape accounted for the majority of their productions. Such a word shape indicates that heterosyllabic CCs are correctly produced by children speaking Arabic, Jordanian in particular, at an early age.

Concerning weak syllable deletion (WSD), a similar problem exists regarding the definition of this pattern. Dyson and Amayreh (2000) described this pattern as 'syllable reduction' without specifying the type of omitted syllable, weak or stressed. Mashaqba et al. (2019) reported the occurrences of WSD in all their age groups (1;0–1;6, 1;7–2;0, 2;1–2;6, 2;7–3;0) and its suppression by the age of 3;0. In Kuwaiti Arabic-speaking children, Alqattan (2015) did not identify this type of pattern at a significant level in her sample. In contrast, Ayyad (2011) reported WSD in her sample, which consisted of children aged older than those sampled by Alqattan (2015). The findings of Ayyad (2011) and Owaida (2015) support each other; both reported the suppression age of WSD pattern at 4;5 years.

2.2.3 Summary of Arabic phonological acquisition studies

This review of normative studies on Arabic phonological development demonstrated a wide array of data available on the development of the phonological system in Arabic-speaking children. The most critical reported data could be summarised in the following points:

- For the speech sound acquisition, the most important findings are:
    - On average, Arabic-speaking children complete the acquisition of Arabic consonantal phonemes by age 5;11 with 75% accuracy and have mastered them by age 6;5 with 90% accuracy.
    - For the vocalic system, it was found that Arabic-speaking children almost completed their acquisition of the Arabic vowels before age 3;0.
    - PCC scores were only reported in two studies: Jordanian (Amayreh & Dyson, 1998) and Syrian (Owaida, 2015). On average, Arabic-speaking children achieved PCC scores of 80% by the age of 3;0-4;0, with a PVC score of 89.22% by the age of 2;6-2;11 (Owaida, 2015).
    - The age of acquisition of consonants co-occurrence was rarely reported in the Arabic phonological developmental studies; however, some studies report the occurrence of word-final CC in children's speech by the age of 4;0, with no precise data about the acquisition of heterosyllabic CC. Geminated consonants were reported to be acquired before age 4;0.
    - The difference between the reported age of acquisition for Arabic phonemes (consonants and vowels) could be explained by methodological, analytical, and dialectal differences.
- For phonological patterns, it was found:
    - Some studies reported the occurrence of universal phonological patterns, such as stopping of fricatives, fronting (velar, palatal, or not specified), devoicing, dentalization, /r/ deviation (gliding, or not specified), assimilation, final consonant deletion, syllable deletion (weak, or not determine), and cluster reduction (word-final CC, word-initial CC, heterosyllabic CC, or not specified).
    - Language-specific patterns were also reported by some studies, including glottal stop replacement, de-emphasisation, and lateralisation of /r/.

  - No dialectal-specific patterns were reported across all Arabic phonological pattern studies.
  - The age of suppression of patterns varies across dialects, which could be due to different cut-off criteria or dialectal differences.
  - On average, most of the patterns reported were suppressed before the age of 5;5.
  - The problems of using different terminologies and cut-off criteria in identifying the phonological patterns challenge the researchers in comparing these studies' results.
- There is some limited knowledge on the phonological acquisition of Saudi Arabic-speaking children, which still could not be used to identify the developmental norms for SHA-speaking children for the following reasons:
  - In the study of Alsa'bi and Naqawa (2013), they did not consider the dialectal form of the phoneme consonants they analysed, and they did not report a clear percentage of accuracy of the age of acquisition in their test manual.
  - Abdoh's (2011) study was focused mainly on the describing the shapes and the syllable structures of the early word in SHA-speaking children.
  - The data provided by Alawwad (2009) and Al- Bader (2009) targeting children speaking Saudi Najdi Arabic cannot be generalised to SHA-speaking children because of some differences in phonological structures between the two dialects (see Section 2.1.3.1).

2.2.4 The theoretical implication of Arabic phonological acquisition studies

The trajectory of phonological development among children speaking Arabic could provide some evidence for assumptions made by the emergentist theory in the acquisition of phonology. The universal and language-specific patterns observed in the phonological acquisition among children speaking different varieties of Arabic (see Chapter 1, Section 1.2) support the emergence approach that has the conceptual scaffolding to integrate nature and nurture more fully within a system-based theory (Davis & Bedore, 2013). Such cross-language differences are well addressed in the emergence approach which reflects broader domains to characterised phonological acquisition, including production, perception, neural cognition, as well as social and cultural development components (see Chapter 1, Section 1.1.3; Davis & Bedore, 2013).

In short, there are still limited data available about the phonological development of children speaking SHA. A wider age range and larger sample size are needed to provide a representative normative sample required to assess the phonological development and, therefore, develop an assessment tool (Kirk & Vigeland, 2014). Some tests are available based on data from children speaking either Jordanian or Egyptian dialects or tools created and then standardised on a group of children. However, no test has considered the international construction criteria needed in such a tool (see Chapter 1, Section 1.4.1).

The following section reviews these available Arabic assessment tools in terms of their appropriateness for diagnosing SSD in SHA-speaking children.

2.3 Assessment of Arabic phonological development

As reported in Chapter 1.4, internationally developed requirements regarding psychometric properties (i.e. normative sample, reliability and validity) should be considered when developing an assessment tool, and the concept of assessment needs to be theoretically grounded. To date, three phonology assessment tools have been used among SLTs to assess speech sounds acquisition in Arabic- speaking children. The Amayreh Articulation test (AAT), which was developed by Amayreh (1994), was created based on a normative sample of Jordanian Arabic-speaking children. Two other tests were developed and then standardised on children. The Mansoura Articulation test (MAT) was normed on children speaking Egyptian Arabic, and the JISH Articulation Test (JAT), which is named after the institution that published it, the "Jeddah Institute for Speech and Hearing" (http://www.jish.org), was also standardised on children living in the Hejaz region of SA. The following subsections briefly describe these Arabic phonological assessment tools and highlight their strengths and weaknesses. Appendix C includes a table with all the criteria used to assess the suitability of assessment tools. These criteria were adopted from Albrecht (2017).

2.3.1 Amayreh Articulation Test (AAT)

The Amayreh Articulation Test developed by Amayreh (1994) is one of the oldest Arabic articulation tests. It consists of 58 pictures designed to elicit all 28 Arabic consonants in three different word positions (28 initial, 28 medial, and 23 final), with 21 words targeting more than one consonantal phoneme at a time. Each consonant is elicited only once in each possible position. Sets of three line–drawn pictures are used to elicit the target sounds. This test targets the standard form of Arabic phonemes as spoken in Jordan, as well as the dialectal acceptable variants being noted and credited. Although Amayreh provided a clear description of the acceptable variants of each standard phone, he did not specify the frequency of occurrences of these acceptable variants across the test items. Furthermore, five final consonants are not tested in this tool because adult Jordanian speakers normally delete them. This could limit the application of this test to other Arabic dilates.

The test was administered to 180 monolingual Jordanian Arabic-speaking children aged between 2;0 and 6;4. Ten percent of the children were selected randomly and retested to examine the reliability of the test; the test-retest reliability (r = 0.832) was particularly good. For content validity, the author ensured that all sounds in Arabic were examined in the test. There is good evidence that pronunciation accuracy increases with age, and this confirms the construct validity. Khoja (2019) reported that some SLTs who work in SA used the AAT to assess and diagnose SSDs among children; however, no information available regarding whether the SLTs make any adaptation on to the AAT score to consider dialectal change.

2.3.2 Mansoura Arabic Articulation Test (MAAT)

This articulation test was developed and published by Abou-Elsaad et al. (2009). Its aim of this test was to provide a criterion for comparing the speech sounds in typically developing and phonologically disordered Egyptian Arabic-speaking children. Like Amayreh's test, it examines all possible consonants of the Egyptian dialect in all possible positions (initial, medial, final). The authors added another word-position termed 'double–consonant' and claimed that such a position could push children's skills to the limit because it is a challenging feature. The test also examines the eight Egyptian Arabic vowels. The total number of stimulus items is 106 coloured pictures. The test consists of two sections: the first section examines all 25 Egyptian Arabic consonants, while second section examines all eight Egyptian Arabic vowels (short and long). The authors reported several criteria used to select culturally appropriate items that are familiar to the participating children. It is important to note that MAAT is never been reported to be used by SLTs in SA (Khoja, 2019).

The normative sample of the test consists of 100 typically developing monolingual Egyptian Arabic-speaking children (48 females and 52 males) aged between 3;6 and 5;8. The children were randomly selected from a kindergarten in Mansoura city.

To determine validity, three judges were asked to review the constructed test with respect to the chosen age group, the appropriateness and clarity of the pictures, suitability of selected words to examine consonants and vowels in each word position, suitability of adding basic vowel words to the test, and the suitability of being able to pronounce the word to the child directly and/or with cues. There was a significant correlation between the three judges on all items in the MAAT questionnaire (r=0.91, P=0.00) and which is suggestive of the test's validity.

The test–retest method was used to examine the test's reliability with the same children being retested five weeks later by the same examiner. The reliability coefficients for correct consonant production and picture identification were 0.75 for the female and 0.66 for the male, respectively, which are lower scores that that reported by McCauley and Swisher (1984).

2.3.3 JISH Articulation Test (JAT)

The JAT [JISH[8] Articulation Test] (Al-Sabi & Naqawah, 2013) is a single-word naming test designed to examine all Arabic singleton speech sounds within a set of 51 items. It is the only test available in SA to assess the articulation skills among Saudi Arabic-speaking children, and 60.4% of SLTs reported

[8] JISH= Jeddah Institute for Speech and Hearing

that they rely on JAT to assess and diagnose SSD according to a survey conducted by Khoja (2019) on SLTs working in SA. In JAT, only consonants are tested at least twice in two different word positions, without including vowels or CCs. One-to three-syllable words are included, with some words having the shape of CVCC (i.e. word-final CC), but they are tested as single phonemes in the word-final position (e.g. /t/ sound in the word-final position is tested using the word /bint/ ('girl') where /nt/ is a CC). Children between 3;0 and 8;0 years were tested with this assessment for norm–referencing and validation. Although Al-Sabi and Naqawah (2013) administered this test to 506 typically developing (323 male, and 183 female) and 352 atypically developing (258 male and 94 female) children for standardisation, the sample's age groups were unequally distributed, with only 18 children in the smallest age group (3;0–3;11) compared to 212 children in the age–group of 6;0–6;11. It is clear that the lowest number of the children in the normative sample were from an age–group in which the majority of the actual acquisition of the consonants universally occurs (Dodd et al., 2005).

JAT scoring is based on the pronunciation of the target phoneme only, which means that if errors occur with other phonemes in the word, these errors are usually not recorded and ignored when determining the speech sounds status. It only records one rendition of each phoneme in each word position. Thus, any variation in the phoneme production is not sampled, which means limited phonological constituents were sampled in this test (James, 2001a).

The JAT manual reported the scores of three measures for reliability: test–retest reliability, the internal consistency, and split–half reliability, where all measures showed a coefficient alpha score above .90. For validity check, the content validity was examined by some expert evaluators to judge the picture–naming agreement, and they had 100% agreement. Another validity measured used was discriminant validity to examine the ability of this tool to discriminating between the articulation of the typically developing children and children with SSDs. The manual states that the statistical analysis for this aspect indicated a high ability of JAT to discriminate between the performances of typical and atypical developed children; however, no further explanation was provided.

According to the authors of JAT, it is a test that needs to be administered by an expert Arabic-speaking SLT, and it takes 20 to 30 minutes to be completed. However, the test manual and scoring sheet both lack IPA transcriptions of test items, which may affect the scoring validity across different examiners due to the lack of an agreed based on which to judge pronunciation accuracy.

2.3.4 Summary of assessment tools used to measure Arabic phonological development

These assessment tools outlined above can help to provide a broad understanding of the phonemic ability of children speaking Arabic in general (i.e. JAT) or some of the Arabic varieties (i.e. AAT for Jordanian, and MAAT for Egyptian). However, all of them lack several valuable criteria for a phonological assessment that follows international test construction guidelines (see Chapter1, Section 1.4.1). Although these tests provide at least one reliability and validity measure, their linguistic criteria are missing several factors needed for a proper phonological assessment tool (see Appendix C).

First, the assessments examined only some Arabic consonants, vowels, or CCs to the recommended frequency. Plus, none of them considered the gemination features in Arabic words. For the Egyptian (MAAT) test, two MSA consonants (/q/ and /θ/) are not tested as they were replaced with the dialectal forms (/ʔ/ and /t/) among Egyptian adult speech. Similarly, the Jordanian test (AAT), /ðˤ/ is tested in a word-medial position in a geminated context (i.e. /naðˤðˤaːra/ 'glasses') only, and all other possible positions for these consonants were ignored because it is not common in the Jordanian adult speech. Additionally, all the tests were designed to assess only consonantal phonemes without including the required criteria to analyse Arabic phonological patterns.

Secondly, although some authors provide some information about items selection criteria, they did not give sufficient details about the familiarity of the test items to the age groups being tested. For example, in JAT there are some words with complex morphological structure (e.g. the plural form in /ma.ħaː.fiðˤ/ ('wallets') which could be difficult for children at 3;0–3;11 age–group).

Thirdly, none of these tests considered the linguistic criteria for the Hejazi Arabic dialect in terms of allophonic variants from MSA and the syllable structures. Although JAT was the only standardised test that targeted children located in Jeddah city, the sample included some bilingual children or children speaking different Arabic dialects. Thus, all dialectal variation was scored as errors, since the test is meant to examine the MSA form of Arabic speech sounds production. As reported earlier, it is the only available test in SA, and SLTs in SA reported that they rely on JAT to assess and diagnose SSD (Khoja, 2019). Al-Sabi (2017) declared that the lack of research on the frequency of Arabic articulation errors and the highly misarticulated sounds in Arabic leads to misdiagnoses of SSD in Arabic-speaking children, which consequently leads to having more children with language/speech disorders in mainstream schools, without proper diagnoses or treatment plans, and struggling with their academic performance.

2.3.5 The theoretical implication of the Arabic phonological development assessment tools

The reported review of the available assessment tools revealed that these tests are all developed based on traditional theoretical approaches of phonological acquisition (e.g. generative phonology and natural phonology), in which the segment is considered the main aspect of phonological development. This theoretical framework emphasised the description of observable behaviours, which precluded the assumptions of abstract mental representation of phonological information. According to Schwartz (1992), there are no differences between phonetics and phonology within the traditional approach, with speech described as a string of segments (consonants and vowels) uttered in a sequence.

In general, the creation of these assessments was not based on the emergence approach, as the selections and analysis of the words proposed for these tests did not consider the principles of such a system (e.g. the role of ambient language and the social-cognitive components of phonological acquisition) (see Section 2.3.4). Moreover, these tests were insufficient for examining phonological patterns in Arabic-speaking children since their word selections needed more opportunity for each phone to occur in each possible position. Such a limitation would limit the options for a phonological pattern sampled by these tests; therefore, they cannot be used to evaluate phonological patterns.

2.4 Literature review conclusion

Overall, the literature review presented in Chapters 1 and 2 has highlighted the procedures used to assess the speech of typically developing monolingual children and the significant characteristics of phonological acquisition in Arabic-speaking children in general. It is clear from the research reviewed in Chapter 1 that in assessing the expressive phonological skills of children, the speech sample could be elicited using either a single-word naming task or CS sample. The single-word naming task is widely used in the literature, especially with large samples. Measures of phonetic inventory, PCC, PVC and PPC, the age of acquisition and mastery of consonant and vowels, and the phonological variants are usually reported to provide a clear picture of the child's phonological abilities. It is also clear that tools used to evaluate phonological skills in monolingual preschool children need to be established on international psychometric criteria. These criteria include the normative sample used to standardise the test, the principles of reliability and validity, and the specific linguistic criteria for the language/dialect under study.

While the reviewed Arabic literature in Chapter 2 has undoubtedly furthered our understanding of the developmental trajectories of phonological skills among Arabic-speaking children, a significant problem continues to affect research in this area, which is that the available Arabic studies do not focus on many phonological features of Arabic. Most Arabic phonological studies have only reported the age of acquisition of consonantal phonemes and concentrated on a particular dialect, with some others investigating phonological patterns. As shown in Chapter 2, Section 2.1.3, Arabic phonology has many phonological features that need to be considered when analysing and assessing the phonological skills of children speaking Arabic.

Despite most of the studies reviewed in Chapter 2, Section 2.2 showing general similarities in Arabic phonological acquisition in children speaking Arabic across different dialects, there is still some dialectal variants between the findings of each study (see Section 2.2). The phonological aspect that showed some anomalies is consonantal phonemes acquisition (see Section 2.2.1.1). However, in terms of phonological patterns, there is also some variation reported among different Arabic dialects (see Section 2.2.2). When creating an assessment tool based on the normative samples used in such studies, these differences need to be considered. McLeod (2012) reported that when designing a tool for assessing the speech of a specific language or dialect, it is essential to understand the phonology of this dialect/language in terms of its phonetic inventory, phonotactic inventory, and phonetic and morphological structure. Therefore, a test designed to assess children speaking a specific dialect/language cannot be used to evaluate children speaking another dialect, which would lead to over–identification of speech errors (Low et al., 2019). The effect of dialect on analysing error patterns was also highlighted by Phoon, Abdullah, & MacLagan (2012) in their study on the speech of Chinese influenced Malaysian English–speaking children aged 3;0–7;11 years. They found that not considering dialectal variants would mislead diagnoses of SSD as dialect influenced speech may lead to a later acquisition of consonants, lower PCC scores, increased the percentage of the occurrence of phonological patterns as well as delaying phonological patterns being overcome (Low et al., 2019; Phoon et al., 2012). Arabic dialectal differences in phonological development are poorly documented, causing difficulties in examining the dialectal effect on the phonological acquisition among children speaking these different dialects. However, the detailed description of phonological differences between these dialects (see Section 2.2) showed that dialectal variation is evident in Arabic. For example, Amayreh and Dyson's (2000) findings showed that the three sounds /ʃ/, /ʕ/ and /h/ were acquired by Jordanian Arabic speaking children between 1;2 and 2;0. In contrast, these sounds were not yet achieved by children speaking Kuwaiti Arabic in a similar age group (1;8–1;11) reported by Alqattan (2015). Ammar and Morsi (2006) and Ayyad (2011) both used single-word naming tests to collect speech samples; however, they reported different ages of acquisition for /d/, for example, whereas Amayreh and Dyson (1998) found that /d/ was reached the mastery level of 90% of accuracy in children speaking Jordanian Arabic by the age of 6;0-6;4, while Ayyad reported that Kuwaiti Arabic speaking children produce /d/ with 90% of accuracy by the age of 4;6. Such differences could be occur due to the studies' methodological differences, but also could be occur due to other factors such as the functional load and the frequency of occurrences of certain sounds in a specific dialect.

Methodological differences, especially in speech sample elicitation and the criteria used to identify the acquisition age, could add to the difficulties in examining the effect of a dialect on phonological development (Alqattan, 2015; Amayreh, 2003). For example, Alqattan (2015) found that two emphatics /tˤ/ and /sˤ/ were acquired by children speaking Kuwaiti Arabic at 3;0 years old, while Ayyad (2011) found that these two emphatic consonants were acquired by the age of 3;10–4;6, and 4;7–5;2, respectively. Not surprisingly, the different data elicitation procedures and criteria used in these two studies reflected the differences in the age of acquisition of sounds where one study used CS sample to elicit the speech sample (i.e. Alqattan, 2015), and the other used single-word naming test (i.e. Ayyad, 2011).

Another issue hindering the comparison of the available data about the phonological development in children speaking different Arabic dialects is the influence of target words used in word lists (i.e. syllable structure and word shape frequency). The selection of words may affect the occurrence of errors or may create a statistical artefact primarily if a specific syllable structure or segment occurs more frequently than others. Although the word list of a single-word naming task may exhibit all possible phonological segments and word structures of a spoken language, some words used to examine a phone in for example geminate structure (e.g. /b/ in /ʃubbak/, or in a cluster (e.g. /b/ in /kalb/). They also examined the medial sounds in two different positions (SFWW and SIWW) interchangeably, ignoring the effect of the sound's position on its acquisition rate. Moreover, no one of these studies reported the familiarity level of their selected words. Such information is obtained through systematic item analysis, which is rarely reported in the Arabic phonological literature.

In short, the available data about the phonological development of children speaking different dialects cannot be used to estimate the phonological trajectory of children speaking Saudi Hejazi Arabic because of the reasons mentioned above. In addition, to assess speech sound development in a group of children, it is not accurate to use a test based on norms targeting another dialect. Therefore, there is a need to develop a test considering the speech characteristics of the Saudi Hejazi Arabic, which has children in the normative sample who speak Saudi Hejazi Arabic. The following section provides the aim, and research questions for this current study.

2.5 The study aims and research questions

As reported in Chapter 1, Section 1.4.2.1, establishing normative data based on international criteria is required to understand the typical and atypical phonological development in monolingual children, and to understand the presence or absence of SSDs (Storkel, 2019). Developmental norms for phonological acquisition give SLTs a more objective view of the discrepancy between typical versus atypical development. Thus, normative data need to be established based on specific criteria to differentiate between children with delayed or disordered speech sound development and children with typical development (Dodd et al., 2003). In order to develop reliable normative data that could be used as a base for comparison by SLTs, it is essential to collect these data using a test material that is created based on psychometric criteria (validity and reliability) and considering the recommended international construction criteria for phonological assessment tools (see Chapter 1, Section 1.4.1). The data collected using single-word naming task are frequently used to establish normative data on the phonetic inventory, age of acquisition of phones, percentage of consonants correct (PCC), and the phonological patterns identified in the speech sample of the participants.

Although some phonological development studies have focused on some Arabic dialects such as Jordanian, Egyptian, Syrian and Kuwaiti (e.g. Alqattan, 2015; Ayyad, 2011; Dyson & Amayreh, 2000; Saleh et al., 2007), their criteria for identifying the typical phonetic and phonological acquisition are not clearly described in the literature. The literature review (see Section 2.2) revealed that the data available are incomplete and sometime where dialectal–specific. Although it was showed that the majority of the Arabic consonants were acquired on average by age 5;6, most of these studies did not examine all Arabic consonants and only focusing on their dialectal phonemes (e.g. Ammar & Morsi, 2006; Owaida, 2015). Further, there is a lack of reliable and valid assessment tools that could be used to evaluate the phonological acquisition in Arabic-speaking children, particularly in SHA-speaking children, and thus identifying children who having SSDs remains particularly challenging (Limbrick, Mc Cormack, & Mcleod, 2013). Most of the previous studies on monolingual Arabic-speaking children were mainly conducted by addressing only one task (i.e. single-word naming task) to examine only one aspect of phonology (i.e. consonantal acquisition). However, it has been proven that observing children's speech development beyond such a single assessment task can shed light on the qualitative and quantitative changes in their speech and highlight the markers for atypical development (Storkel, 2019). Fox (2000) reported that assessing consonants only to evaluate phonological development might provide too limited perspective on speech acquisition. Thus, the available data on the phonological development in Arabic-speaking children, particularly Saudi Hejazi, is insufficient.

Hence, the following main research aims for the present study was set:

1) To create a phonological assessment tool (single-word naming test) for Saudi Hejazi Arabic dialect considering international construction criteria.
2) To describe the typical phonetic and phonological development in Saudi Hejazi Arabic-speaking monolingual children aged 2;6–5;11.

With regard to the first research aim, the following research questions will be addressed:

RQ1: Can a phonological assessment tool (the single-word naming test; SHAPA) be developed based on international construction criteria leading to a clinically valuable tool for assessing the phonetic and phonological development in SHA-speaking children?

Considering the second research aim, the following research questions will be addressed:

RQ2: How does the phonetic acquisition in monolingual SHA-speaking children proceed between the age of 2;6 and 5;11 in terms of the following:

2a) At what age do Saudi Hejazi Arabic-speaking children acquire and master the Arabic (UHA and MSA) phonetic system taking into account different measures such as PCC, PCC-R, PVC, PCCC (three different types of consonant co–occurrences), and age of acquisition of C and CC structures?
2b) At what age are the consonants of UHA and MSA stimulable when elicited via a single word naming task or a stimulability task?
2c) Does the phonetic inventory elicited via a single-word naming task versus a stimulability task differ?

RQ3: Which phonetic and phonological error patterns are found in the speech of Saudi Hejazi Arabic monolingual speaking children aged between 2;6, and 5;11? Particularly in terms of:

3a) To what extent does the frequency of phonological variants (Tokens and pattern Types) alter over time as phonological development progresses?
3b) What is the frequency of infrequent variants across age groups?
3c) What phonological patterns can be observed as developmental at different ages using two distinct cut-off criteria?"

To answer these research questions, the researcher is conducting a cross–sectional study using the newly developed assessment tool (SHAPA). The data collected are analysed by evaluating the applicability of this tool to investigate phonological acquisition among these children. It is also analysed to identify the order and age of phonetic acquisition, the types and number of phonological variants and patterns that occur in children's speech, and their age of occurrences. Descriptive analyses are used to report the results. The next chapter describes the study method in detail.

3 Methods and Procedure

This chapter presents the methods and procedures used to address the aim of the current study (see Chapter 2, Section 2.5). In addition, Section 3.3.1.1 includes a detailed description of the design and rationale for creating the Saudi Hejazi Arabic Phonology Assessment (SHAPA) tool for this project.

3.1 Study design

This study has a descriptive cross–sectional design and the fieldwork took place in Jeddah, Saudi Arabia (SA). Data from single-word naming and phonetic imitation tasks were collected from monolingual preschool SHA-speaking children. Parental questionnaires were also distributed to all parents/carers to gather comprehensive background information on the children's general development, including speech and language acquisition.

This project was ethically approved by the ethics committee in the Department of Human Communication Sciences, University of Sheffield, in line with the university's ethics review procedures (see Appendix D).

3.2 Participants

To reflect the typical population of monolingual SHA-speaking children aged between 2 years, 6 months and 5 years, 11 months, participants were recruited from nurseries and preschools in Jeddah, the largest city in the western region of SA, which is the homeland of the Hejazi dialect (see Chapter 2, Section 2.1.2.1). The estimated population of Jeddah is about 4 million people distributed within 137 districts (The Emirate of Makkah Region, 2020). This city has 54 public preschools under the administration of the Ministry of Education that are located in different parts of the city with a total of 5,797 Saudi children enrolled (Kindergarten Administration, n.d.).

The first step in the recruitment process was to obtain approval for the study from the Ministry of Education in Jeddah, as it is the official representative of public preschools in SA[9]. The researcher visited the manager of the preschool department to obtain permission to enter public kindergartens and provided him with all required documents[10]. A month later, the manager provided the researcher with an approval letter with open permission to conduct the study in any preschool under the Ministry of

[9] All researchers in Saudi Arabia who are interested in recruiting participants from public schools are required to obtain prior approval from the Ministry of Education to be eligible to visit public schools.
[10] These required documents included a translated copy of the ethics application, an ethical approval letter from the University of Sheffield, and all test materials including the information sheet, the consent form, the parent questionnaire, the word list and pictures for all the words.

Education's administration in Jeddah.

The initial targeted sample size was a minimum of 210 participants (and with a maximum of 280) with participants equally divided into seven age groups, each with an average of 15 females and 15 males. It was planned that each age group should have a minimum of 30 children to increase the sample size in comparison to most of the previous studies for the higher potential to capture the actual average of the development while making data collection still feasible within a PhD project. To achieve this aim, the researcher began to visit different kindergartens around Jeddah at random[11]. She visited 24 out of 54 preschools in Jeddah over a period of 109 days, divided into two time periods. The first data collection trip lasted from February 2019 to early April 2019. During this trip, the school principals in the 17 visited schools were asked to distribute 502 envelopes to children in their classes who met the inclusion criteria. Two hundred and forty–two parents returned signed consent forms; however, 59 were excluded (see Appendix F for exclusion reasons), while the rest 183 children were included as their received questionnaires showed that they fulfilled the selection criteria. The second period of data collection took place from September 2019 till the end of October 2019. One hundred and eighty envelopes were distributed by class teachers in seven different schools, and 88 families returned the documents with signed consent forms; 36 of them were excluded (see Table 3.1). Figure 3.1 summarises these two recruitment phases.

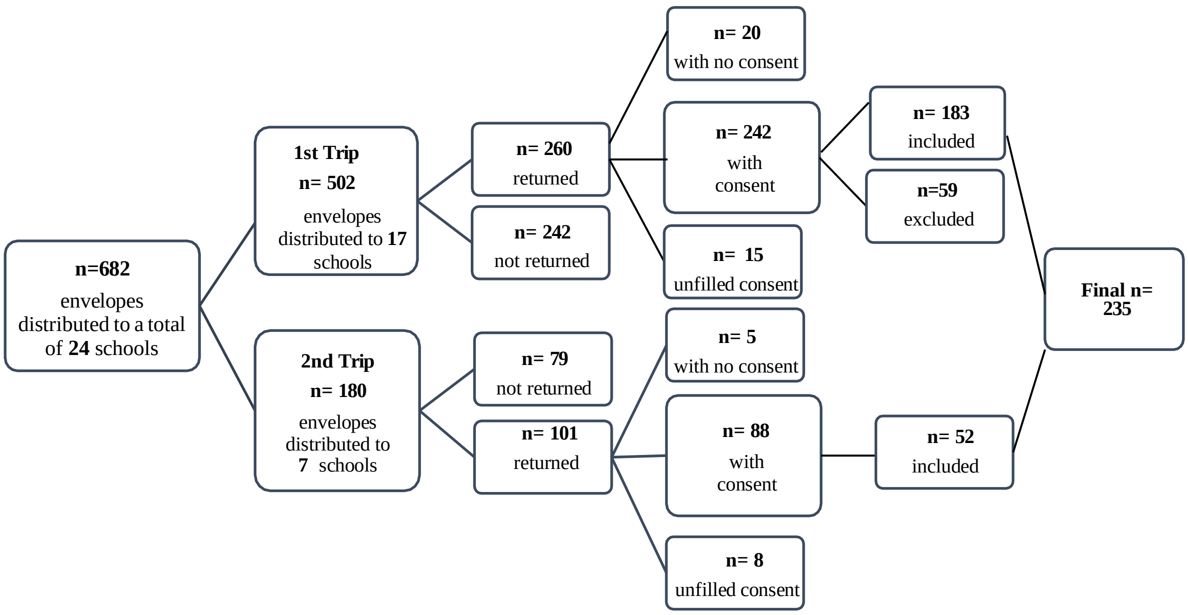


Figure 3.1 *Recruitment process and number of study documents distributed*

During the visits to the schools, the researcher gave a copy of the ministry's letter to the schools' principals; accordingly, they allowed her to conduct the study in their schools. The researcher provided

[11] The researcher selected one of the kindergartens that belonging to the MOE by just drawing one school out of a hat without any criteria for her choice.

the schools' principals with a brief verbal description of the project and its goals, along with all the study documents (information sheet, consent forms, and parent questionnaires) placed in envelopes ready to be distributed to the selected children. The researcher then asked each school director to give these envelopes to the class teachers. These class teachers were asked to distribute the study documents to the children who met the following inclusion criteria.

The child is a **monolingual Saudi Hejazi**, and the main spoken language at home is Arabic, particularly Urban Hejazi Arabic (UHA). Any child whose parents speak another Arabic dialect is excluded.

- The child is aged between 2;6 and 5;11 years.
- The child is typically developing[12], with no history of any medical and/or congenital condition.
- The child has no history of hearing, speech, language, and/or communication disorders or delay.

These inclusion criteria were explained verbally to each class teacher by their school manager. In addition, a three–page questionnaire was sent to the children's families to obtain demographic data on each participant and ensure that all participants met the inclusion criteria (see Appendix E). Moreover, it was used to identify the sociolinguistic situation of participants, especially their exposure to MSA versus UHA; parents of the children were asked to evaluate their child's exposure time to these two forms of Arabic. The second part of the questionnaire had questions about children's early speech and language development skills. As stated earlier, due to the lack of a standardised tool for assessing speech and language skills in SA, detailed questions about speech and language skills were included in this part. Hearing screening was also performed through this questionnaire because pure-tone hearing screening has been reported to be an unreliable and unfeasible tool for children within this age range in the preschool setting (Hall, 2016). Parents answered four questions about ear infections and functional hearing. Information about participants' hearing skills was needed to ensure that all children who participated in the phonological assessment have normal hearing. The last part of this questionnaire was designed to obtain demographic data on the children, which was needed to provide background information on the participants and a more detailed description of the study sample.

Parents who received this survey needed 5 to 10 minutes to complete it with the necessary information. They returned the questionnaire and a signed consent form to the school within one or two weeks. The total number of the returned documents was 361 (from both data collection trips), 330 documents were returned with consent. Any children whose parent/carer had concerns about their speech and language development and/or reported that their children are bilingual were excluded before the tasting day. Other exclusion criteria/reasons included the saturation of the age range by the testing day, lack of the child's

[12] Please note that the selection of 'typically developing' children was based on parental judgment only (see Chapter 1, Section 1.2.2.1.2). Therefore, it was possible that the included children could be typically developing but show SSD to some extent (Clausen & Fox-Boyer, 2022).

motivation during the test, the absence on the testing day or if the child did not complete the test. Appendix F provides more details about the exclusion reasons together with the number of the children excluded for each reason. The total number of excluded children was 59. Therefore, the final sample size of this study consisted of 235 monolingual SHA-speaking children aged 2;6–5;11, divided into seven age groups (2;6–2;11, 3;0–3;5, 3;6–3;11, 4;0–4;5, 4;6–4;11, 5;0–5;5, 5;6–5;11). Detailed information about the gender and age of participants in each age group is shown in Table 3.1.

Table 3.1 *Participants' gender and age distribution (N=235)*

Age groups	N	Female	Male	*M* age (months)	*SD* age (months)	% of sample
2;6–2;11	30	15	15	32.20	1.86	12.77
3;0–3;5	27	14	13	30.00	1.92	11.49
3;6–3;11	38	17	21	45.24	1.79	16.17
4;0–4;5	39	17	22	50.33	1.63	16.60
4;6–4;11	31	16	15	57.19	1.74	13.19
5;0–5;5	37	22	15	62.38	1.53	15.74
5;6–5;11	33	18	15	68.45	1.79	14.04
Total	235	119	116			

In the final sample, the majority of mothers reported having an undergraduate degree (68.09%), or a secondary school level of education (17.45%). A minority had completed either a lower (i.e. diploma degree 6.38% or middle school 2.55%) or higher level of education (i.e. postgraduate degree 4.26%), with 2.55% of the mothers having either no degree or not reporting their level of education.

3.3 Assessment tasks and administration procedures

To investigate phonological developments in SHA-speaking children, participants were assessed on a single-word naming task designed for this research project. This was supplemented by a phone imitation task (stimulability task) aimed at investigating the children's ability to produce the phones of UHA and MSA in isolation.

3.3.1 Single-word naming task

Due to a lack of normative data, as well as the lack of a reliable and well–developed-phonological test, for SHA-speaking children, it was necessary to create a speech assessment tool that met the phonological characteristics of the Hejazi dialect (i.e. UHA). To achieve this aim, the researcher first developed a new set of items that considered all the speech sounds of UHA as well as taking into consideration international construction criteria for speech assessments during the task creation process.

3.3.1.1 Development of the Saudi Hejazi Arabic Phonology Assessment (SHAPA)

When developing the Saudi Hejazi Arabic Phonology Assessment (SHAPA), the researcher initially designed a word list considering linguistic/phonological and cultural criteria. These criteria were based on the international guidelines and linguistic criteria reported in Chapter 1, Section 1.4.1. These included:

1- To contain all consonants of UHA in all possible word/syllable positions: syllable–initial word-initial (SIWI), syllable initial within a word (SIWW), syllable–final within a word (SFWW), and syllable–final word-final (SFWF), with at least three occurrences for each possible position. The two medial positions (SIWW and SFWW) were considered because, as reported by James (2001a) and Amayreh and Dyson (2000), the consonants in this position behave differently relative to initial and final positions. (see Chapter 1, Section 1.4.1)

2- Each consonant should occur in at least two different vowel contexts. For example, the glottal stop /ʔ/ needs to be presented with consideration of various consonant-vowel sequences: /ʔa/, /ʔi/, and /ʔu/ following the model of James (2001a), which was adapted from Stoel-Gammon and Dunn (1985) (see Chapter 1, Section 1.4.1).

3- To include a sufficient number of occurrences of each UHA vowel at least twice in every possible position.

4- To include the MSA consonants that are not part of UHA, (i.e. /θ/, /ð/, and/or /ðˤ/) in at least two different word positions.

5- To include words of different length (monosyllabic to multisyllabic words) As reported in Chapter 2, Section 2.1.3.4, polysyllabic words are common in Arabic children's speech.

6- To make sure that consonants in a cluster are not used to assess single consonant as a singleton (see Chapter 1, Section 1.4.1).

7- To ensure that the targeted speech sounds are in stressed syllable and unstressed syllables (see Chapter 1, Section 1.4.1).

8- To provide sufficient opportunities for production of the most commonly reported phonological patterns in the Arabic literature that occur in the speech of typically developing Arabic-speaking children (see Chapter 2, Section 2.2.2).

9- To use age–appropriate items that are well–known by SHA-speaking children. Due to the lack of normative data on the vocabulary of Saudi Hejazi children aged between 2;6 and 5;11, some educational YouTube channels for Arabic-speaking children and different children's books were used as a reference source to initially find appropriate words for the assessment. The Arabic version of the McArthur-Bates Communication Development

Inventory (JA-CDI; Dashash & Safi, 2014) was used in the second step. Exactly 80% (n=124 out of 155) of the initially chosen items are consistent with items from the Saudi-Arabic adaptation of MB-CDI (JA-CDI) and considered age–appropriate for children aged up to three years. The remaining 20% (N=31) of the SHAPA words were added due to their specific phonetic–phonological features not available in JA-CDI items.

10- To display the items in colourful, culturally sensitive pictures, facilitating spontaneous naming without other cuing than stimuli's elicitation production. As reported in Chapter 1, Section 1.4.1, spontaneous naming is preferred over imitated responses because it is much closer to the daily speech of children.

In order to ensure that each consonant is presented in various phonetic contexts, a word-matrix Excel sheet was created to represent all UHA and MSA consonants, the words used to elicit each consonant in different consonant-vowel contexts, their word shape, and the frequency of each word within the list (see Appendix G) similar to the model of James (2001a) which was adapted from Stoel-Gammon and Dunn (1985).

The initial word list consisted of 155 items which were assessed for age-appropriateness. To achieve this goal, the researcher created a Google Form of these 155 words and distributed it to families and friends who were Hejazi Arabic-speaking parents, kindergarten teachers, and SLTs to identify each word's age–acquisition. A clear written instruction was provided at the top of the form, asking the participant to choose the age range that he or she thinks is appropriate for a child to name the target word. Each word in the list was provided with eight multiple–choice answers that cover different age ranges (2;0–2;5, 2;6–2;11, 3;0–3;5, 3;6–3;11, 4;0–4;5, 4;6–4;11, 5;0–5;5, older than 5;6). The link to the form was sent to the participants, and all the responses were anonymous. A total of 13 Saudi–Hejazi adults completed this form. This step identified some selected words with late acquisition; therefore, four words were excluded as they were ranked by all participants to be acquired after the age of 5;6. For example 'office' /mak.tab/ and 'river' /na.har/ were rated by the majority of the adult participants to be acquired at an age older than 5;6, so they replaced by 'spaghetti' /ma.ka.ruː.na/ and 'coffee' /gah.wa/ which was found to be more familiar to children aged three years according to JA-CDI word list.

Consequently, some other changes were made to replace the deleted words' phonological characteristics on the list, resulting in the second word list of 151 items, which was used to create the pictures. All images used in this study were downloaded from an online picture provider, www.shutterstock.com, which provides a subscription option for users to download different types of pictures with high–quality options available for licensing (see Appendix H for the proof of purchasing). These pictures were child–-friendly, cartoon-style photos and were presented on slides with a white background to limit distractions. There were 144 pictures used to elicit 151 words.

To check the reliability of the picture–word association, eight adults among a group of families and friends participated in a naming–pictures task to examine whether each picture could elicit the target word. Three of the participants correctly named all the images. However, the other five participants misnamed five pictures; accordingly, they were changed to more explicit photos. The researcher also asked the participants to record their voices while naming the words to check the possible acceptable variants in their pronunciation, finalising the phonetic transcription and the stressed syllables of each target word. Appendix I includes all pronunciation variants of the list items. These naming variants assisted the researcher in identifying the acceptable adult forms of each word during data collection.

The final version of the list included mainly nouns (143) with only four verbs in the present tense and four adjectives, as there is some evidence showing that sound acquisition is placed later in syntactic classes other than nouns (James, 2001a). However, verbs and adjectives were included in the list because they contained some consonants in certain word positions that were not easily found in nouns familiar to children in the age range of this study (2;6 to 5;11). The verbs were in the present simple tense, and all were used to examine consonants in the medial position; for example, the verb ('washing') /ti.ɣas.sil/ (feminine) or /ji.ɣas.sil/ (masculine) was used to test the /ɣ/ consonant in the medial position.

Moreover, eight nouns in the list were presented in a broken plural form[13]. All words with plural forms were formed by reducing the length of the singular form of the word, thus changing the disyllabic singular form to a monosyllabic plural form after removing the suffix *–at* (McCarthy & Prince, 1990). For example, the singular form of the word ('egg') /beː.dˤah/ (UHA) or /baydˤatun/ (MSA) becomes ('eggs') /beːdˤ/ (UHA) or /baydˤun/ (MSA) in a broken plural form. These words were chosen in this particular plural form to test their final–word position consonants.

After developing the final draft of the SHAPA, the researcher piloted it to check the assessment tool's applicability and determine the content validity of the single-word naming test.

3.3.1.2 Pilot study

A pilot study was conducted on four typically developing SHA-speaking children. The researcher met the children and their mothers in her house or at their homes. All the children were tested individually in a quiet room with their mothers attending the sessions.

[13] The broken plural in Arabic involves modification of the pattern of the consonant and vowels inside the word (internal plural); for example, /kita:b/ "book" →/kutub/ "books" (for more details, see McCarthy and Prince (1990).

The pictures were presented in a PowerPoint presentation using a MacBook Air laptop. An audio recording was conducted using Audacity® software and a Samson® USB studio condenser microphone. The microphone was placed near each child to ensure that all responses were clearly recorded. The audio files were saved on the researcher's laptop with an anonymous code attached to each child.

At the beginning of the test, the researcher gave each child clear instructions about the test and asked the child to name each picture on the laptop screen. A hierarchy of prompts was used if the image did not elicit a spontaneous name from the child (see Section 3.3.1.4. for a detailed description of the prompt hierarchy).

The children's responses were analysed by item difficulty, where the mode of stimulation used to elicit each item was documented and the spontaneous responses were counted. Table 3.2 represents the number of each cue used with the children and the percentage of spontaneously produced responses.

Table 3.2 *Types of prompts used with the participants in the pilot study and the number and percentage of the responses using each type of prompt (n= 143 items)*

Participant	**P1 (F)**	**P2 (F)**	**P3 (M)**	**P4 (M)**
Age	2;6	4;7	4;9	4;9
Prompt	n (%)	n (%)	n (%)	n (%)
Spontaneous responses	35 (24.48%)	94 (65.73%)	86 (60.14%)	92 (64.34%)
Semantic cues	2 (1.40%)	3 (2.10%)	6 (4.20%)	10 (6.99%)
Alternative questions	9 (6.29%)	34 (23.78%)	19 (13.29%)	18 (12.59%)
Phonetic cues	4 (2.80%)	6 (4.20%)	24 (16.78%)	16 (11.19%)
Imitation	93 (65.03%)	6 (4.20%)	8 (5.59%)	7 (4.90%)

After analysing the participants' responses and their elicitation process, the word list was revised again according to the children's responses, replacing words with a low percentage of spontaneous naming (see Table 3.3).

Table 3.3 *Items with low spontaneous naming rate and their replacement*

Word from piloted test	Glossary	Replacement for final test version	Glossary
/fin.ʒaːn/	Coffee cup	/fa.raː.ʃa/	Butterfly
/ru.baːtˤ/	Shoelace	/xeːtˤ/	Thread
/na.har/	River	/gah.wa/	Coffee
/tuːt/	Blueberry	/ʕan.ka.buːt/	Spider
/ʔa.sad/	Lion	/jad/	Hand
/mu.ka.ʕa.baːt/	Blocks/cubes	/ma.ka.ruː.na/	Spaghetti
/ra.sˤiːf/	Sidewalk	/mu.kaj.jif/	Air conditioning
/da.waː/	Medicine	/ħa.laː.wa/	Candy

Eleven words named by imitation were not replaced by other items due to the lack of appropriate replacements with a similar phonological structure. These words were rated by the Hejazi Arabic-speaking adult to be acquired at an age older than five. For example, /dˤirs/ 'molar (tooth)' and /ðˤarf/ 'envelope' were not replaced due to the limited number of familiar words initiated with /dˤ/ and /ðˤ/. /hu.duːʔ/ 'silence' and /ʔib.ti.daː.ʔiː/ 'primary school' were also not replaced by other words because of the glottal stop /ʔ/.

3.3.1.3 Final version of the Saudi Hejazi Arabic Phonology Assessment (SHAPA)

Having described the steps taken to check the content validity of the created phonology assessment, a clear description of its UHA–related phonological features will now be provided.

3.3.1.3.1 Consonants

The final word list included 437 consonantal phonemes presented in 151 words in four syllable positions, mostly three times for each possible position (see Appendix J for more details). Table 3.4 provides more details about the frequency of each consonantal phoneme in each word position.

Table 3.4 *Frequency of UHA consonantal phonemes in the word list, according to word position*

The Phonemes	SIWI	SFWW	SIWW	SFWF	Total
1. ʔ	7	–	2	3	12
2. b	11	2	12	7	32
3. t	5	–	7	3	15
4. θ*	2	1	2	1	6
5. ʒ	3	2	4	3	12
6. ħ	4	2	3	4	13
7. x	7	2	1	3	13
8. d	6	1	10	3	20
9. ð*	2	2	–	1	5
10. r	5	20	6	9	40
11. z	3	2	1	3	9
12. s	4	4	5	4	17
13. ʃ	7	1	5	3	16
14. sˤ	6	3	3	3	15
15. dˤ	3	2	3	3	11
16. tˤ	4	1	4	2	11
17. ðˤ*	2	–	1	–	3
18. ʕ	5	1	4	3	13
19. ɣ	3	1	4	2	10
20. f	4	2	8	3	17
21. q*	–	–	1	–	1
22. g	4	2	3	3	12
23. k	3	1	3	3	10
24. l	6	2	14	6	28
25. m	23	4	12	7	46
26. n	4	4	5	11	24
27. h	4	1	3	–	8
28. w	3	-	6	–	9
29. j	5	-	3	–	9
Total	145	63	136	92	436

* These sounds are MSA consonantal phones; they were included in the word list for a comparison purpose between acquisition of MSA and UHA speech sounds. *These are the UHA only phones, not part of the MSA phonetic inventory. All the other sounds are shared phones that are part of the MSA phonetic inventory and used in the UHA phonetic inventory.

Not all consonants are represented in all syllable positions or are represented fewer than three times because some consonants do not, or infrequently, occur in certain positions in Arabic and UHA (see Chapter 2, Section 2.1.3.1). For example, two of the consonants /t, w/ are not presented in the syllable-final within word (SFWW) position due to a lack of child–friendly items that represent this position for these two consonants as a single phoneme. Similarly, because the word-medial glottal stop /ʔ/ in UHA is used to be assimilated to the preceding vowel (see Section 2.1.3.1), it does not occur in the SFWW position in SHAPA. It is also evidence that the SHAPA word list included the MSA consonants, where they occur in the cogent words presented in SHAPA; however, they are not represented across all possible syllables positions because they are usually replaced by their dialectal counterparts in SHA-speaking adults (see Table 2.1). Further, Figure 3.2 illustrates the frequency of consonant types in the word list according to the manner of production.

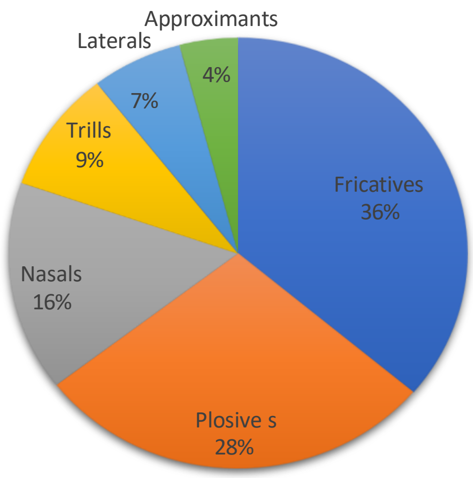


Figure 3.2 *Distribution of UHA consonant types in the SHAPA word list*

Figure 3.2 showed that, and as recommended by James (2001a), fricatives are sampled with a higher frequency than other sounds because these fricatives, as well as liquids, are later- developing sounds and are more prone to misarticulation than plosives, nasals, and glides, which are known to be acquired before other sounds (James, 2001a). However, it could be also noticed that plosives were strongly represented in SHAPA; still their percentage of occurrences was less than the fricatives.

3.3.1.3.2 Consonantal co–occurrences

Although the researcher initially developed the SHAPA test by considering only single consonants, the final version of the test included several instances of consonantal co–occurrence (i.e. word-final CCs, heterosyllabic CCs, and gemination). These types of consonantal co-occurrences were not considered when the SHAPA was initially developed. However, because these features are worth investigating to understand how SHA-speaking children deal with such lexical structures, they were included in the analysis of data collected by SHAPA. The following section describes these features in the SHAPA.

Word-final consonant clusters (CC)

The word-final CCs were included in some of the monosyllabic words on the list, as this feature occurs in UHA only in monosyllabic words. Children's acquisition of word-final CCs and their strategy to deal with this type of syllable structure were included in the analysis of this study, using data from nine items in the SHAPA word list. Table 3.5 lists all the words with a final CC in the SHAPA word list, with a description of the place and manner of articulation of each cluster, as well as the sonority type of each CC (see Chapter 2, Section 2.1.3.2.1).

Table 3.5 *Items with word-final consonant clusters (CCs) in SHAPA*

UHA/MSA	Glossary	Manner of CC	Sonority type
kalb	Dog	L–P	falling
gird	Monkey	T–P	falling
ʃams	Sun	N–F	falling
milħ	Salt	L–F	falling
waʒh	Face	F–F	plateau
dˤirs	Molar	T–F	falling
durʒ	Drawer	T–F	falling
bint	Girl	N–P	falling
zˤarf	Envelop	T–F	falling

Key: L= Lateral, T=Trill, N= Nasal, F= Fricative, P= Plosives (the lateral sounds are the most sonorous and the stop sounds are the least sonorous).

As reported in Chapter 2 (see Section 2.1.3.2.1), MSA nouns with the CVCC structure (e.g. /kalb/) are maintained in UHA if these final CCs have falling or plateaued sonority. Therefore, all the CCs included in the SHAPA word list are falling or plateaued sonorous types. No words–final CCs with rising sonority were incorporated in the SHAPA word list. Although some CCs with rising sonority were reported to be preserved by Hejazi adult speakers in Alfaifi's (2019) study, most of these words with final CCs in his study were either presented in words unfamiliar to the age range of children in this current study–for example, /ʒifn/ 'eyelid'– or were unable to be displayed using child–friendly images for example, /nabdˤ/ 'pulse'. Other MSA words with rising sonority CCs that typically undergo epenthesis by adult SHA–speakers are included in the SHAPA word list (see Table 2.4, Section 2.1.3.2.1). These words were presented in the SHAPA word list as they are pronounced by Saudi Hejazi adults with the vowel being inserted between the CC elements. Thus, the final number of words with word-final CCs in SHAPA word list is limited to only nine words. Such a limitation negatively affects the representativeness of the selected items since they do not reflect the full range of word-final CCs in

UHA adult speech.

Heterosyllabic CCs

The SHAPA word list included 52 words with heterosyllabic consonant clusters (i.e. word-medial CCs across syllables boundaries) either in bisyllabic or trisyllabic words. As reported in Chapter 2 Section 2.1.3.2.2, heterosyllabic CCs occur in UHA, as well as all other varieties of Arabic. Although this feature was not taken into account when SHAPA was developed, it was considered as a feature worth investigating as Mashaqba et al. (2019) reported that the most preferred word structure among Jordanian Arabic-speaking children is bisyllabic words, where the CV(V)C.CV(C) word shape accounts for the majority of their productions. Such a words shape in Mashaqba, and colleagues' study were indicated the production of heterosyllabic CCs by children speaking Arabic, Jordanian in particular, at an early age (e.g. children nagged 1;0 to 3;0). Therefore, considering such a structure in the SHAPA instrument is important for examining how SHA-speaking children acquired this type of consonantal co-occurrences. Appendix K represents all words with heterosyllabic CC in the SHAPA word list.

Gemination

Gemination is another feature that was included in the SHAPA word list. This feature occurs in UHA words, whether disyllabic or multisyllabic, in the word-medial position only. As reported earlier, the items selected for the geminated feature may not represent all possible geminated consonants in Arabic, where there are almost 27 medially geminated sounds, but only eight are sampled in the study's word list.

Table 3.6 represents all words in the SHAPA word list with a geminated middle consonant. The selection of these items was based on choosing initial and final consonant in words that are familiar to SHA-speaking children and which were able to be presented in child–friendly pictures.

Table 3.6 *Items with geminated middle consonants in SHAPA*

UHA	Glossary	Geminated consonant
mu.sal.las/ bal.loːna/ xal.laːtˤ	Triangle/ Balloon/ Blinder	/l/ Alveolar lateral approximant
batˤ. tˤiːx/ batˤ.tˤaːnija	Watermelon/ Blanket	/tˤ/ Emphatic alveolar stop
tuf.faħa	Apple	/f/ Labiodental fricative
nadˤ.dˤaːra	Glasses	/dˤ/ Emphatic alveolar stop
ti.ɣas.sil/ ɣas.saːla	She is washing/ Wash machine	/s/ Alveolar fricative
ħam.mam	Bathroom	/m/ Bilabial nasal
ʃub.baːk	Window	/b/ Bilabial stop
suʒ.ʒaːda	Rug/ mat	/ʒ/ Postalveolar fricative

Notably, in this study, all geminates were transcribed as two identical consonants separated with a syllable boundary mark. Care was taken to ensure that these geminated consonants were not considered as medial consonants (i.e. SFWW or SIWW). For example, in the word /ʃub.baːk/ 'window', the phoneme /b/ was not counted as one for a medial instance of the /b/ sound.

3.3.1.3.3 Vowels

As reported earlier, the word list was created by considering at least two occurrences of each vowel in various positions (see Appendix L for more details). Table 3.7 presents the total number of UHA vowels in the word list.

Table 3.7 *Frequency of vowels and diphthongs in the word list*

Vowels	**/i/**	**/iː/**	**/a/**	**/aː/**	**/ɑ/**	**/ɑː/**	**/u/**	**/uː/**	**/oː/**	**/eː/**	**/au:/**	**/ai:/**	Total
Frequency	43	16	138	38	15	5	29	12	7	6	~~3~~	~~2~~	**314**

It is clear from Table 3.7 that the diphthongs /ai/ and /au/ have only two occurrences in the SHAPA, because these two vowels usually replaced by the long vowels /e:/ and /o:/ in UHA. However, these are still preserved in a few words as in /lai:.mu:n/ and /mu.kai:.jif/, and /zau:.wa:l/ and /tau:.la/, respectively.

3.3.1.3.4 Word length

In terms of word length, the word list consisted of 27 monosyllabic words (17.6% of the list), 86 disyllabic words (57.5%), 31 trisyllabic words (20.3%), and 7 quadrisyllabic words (5%). See Table 3.8 for more details about the percentage for the word–length types, including some examples of their syllabic structures and stress patterns.

Table 3.8 *Frequency of different word lengths in SHAPA*

Word length	Frequency	Example (English)	Hejazi IPA	Syllable structures	Stress patterns
Monosyllabic	27	Dog	kalb	CVCC	–
		Elephant	fiːl	CVːC	–
Disyllabic	86	Sheep	xa.ruːf	CV.**CVːC**	w–S
		Frog	dˤif.dˤaʕ	**CVC**.CVC	S–w
Trisyllabic	31	Spider	ʕan.ka.buːt	CVC.CV.**CVːC**	w–w–S
		Circle	daː.ʔi.ra	**CV**ː.CV.CV	S–w–w
Quadrisyllabic	7	Orange	bur.tu.qaː.la	CVC.CV.**CV**ː.CV	w–w–S–w
		Blanket	batˤ.tˤaː.ni.ja	CVC.**CV**ːCV.CV	w–S–w–w

3.3.1.3.5 Word shapes

All word shapes allowed in UHA were also sampled in SHAPA. The frequency of each shape is presented in Table 3.9.

Table 3.9 *Frequency of word shapes in the list*

Monosyllabic		Disyllabic		Trisyllabic	
CVCC	9 (5.8%)	CV.CV	2 (1.5%)	CV.CV.CV	4 (3%)
CV:C	**12 (7.8%)**	**CV.CVC**	**19 (14.3%)**	**CV.CV:.CV**	**5 (3.8%)**
CVC	5 (3.9%)	CV:.CV:	1 (0.75%)	CV.CV.CVC	2 (1.5%)
		CV.CV:C	**16 (10.5%)**	CVC.CV.CV:C	2 (1.5%)
		CV:.CV	2 (0.75%)	CV:.CV.CV	1 (0.75%)
		CV:.CVC	1 (0.75%)	CV.CV:.CVC	1 (0.75%)
		CV:.CV:C	3 (2.25%)	**CVC.CV.CV**	**5 (3.8%)**
		CVC.CV	9 (7.5%)	CVC.CVC.CV	1 (0.75%)
		CVC.CVC	**13 (9.7%)**	CVC.CV:.CV	2 (1.5%)
		CVC.CV:C	**17 (9.7%)**	CV.CV:.CVC	1 (0.75%)
		CV:C.CV	1 (0.75%)	CV.CVC.CVC	1 (0.75%)
		CVC.CV:	1 (0.75%)	CV.CV:.CV:C	1 (0.75%)
		CV:.CV	2 (0.75%)		

*The bold numbers represent the most frequent shapes in each word–length group.

As predicted by James (2001a), a phonological test that samples words with different word lengths and shapes have a superior item discrimination measure as these complex words may capture some of the subtle features in speech that are not be able to be captured when testing focuses on the aspects of phoneme segments or is dominated by monosyllabic words. Therefore, the SHAPA test was designed to include more complex words to elicit a more representative sample to measure the typical versus atypical speech development.

3.3.1.4 Administration procedure of the single-word naming test

All pictures were presented on PowerPoint slides with one picture per slide with white background to avoid the child becoming distracted. When administering the task, participants were asked to spontaneously name one image at a time. The examiner prompted their responses by posing open and unspecific questions such as ‘What is this?’ or ‘What do you see in this picture?’ However, additional prompting procedures were applied if spontaneous naming was not obtained. This prompting followed a specific hierarchy. First, semantic cues, which include descriptive information, were given to elicit the target word (e.g. the examiner named the function or the location of the object). Second, an alternative question was asked (e.g. the examiner asked the child: ‘Is this X or Y?’, with X always being the target word). This order was intended to inhibit the occurrence of a direct repetition effect. Third, phonemic cues were provided, which include giving the initial sound or CV syllable (first sound imitation). Finally, if the child still could not name the target item, the child was asked to imitate the whole word. Every response was included in the analysis regardless of the level of prompting used to elicit it (Goldstein, Fabiano, & Iglesias, 2004). Each level of elicitation used was recorded in the response record sheet (see Appendix M). Occasionally, children named some of the items in the non-target form or in another acceptable variety of the target word (see Appendix N). If this happened, the examiner accepted this response; and later changed the target word for each child to match his or her response. There were no practice items included in the test since it was expected that all the words would be familiar to all the children in the sample, and the task demand was low because picture–naming is a common activity for all children in this age range (2;6–5;11) (Stackhouse & Wells, 1997).

The time needed for the administrating the naming task varied between 20 and 60 minutes, including a two–minute break if required. Sometimes, the task was divided into multiple sessions depending on the child's age and ability to maintain interest.

3.3.2 Stimulability task: Phone imitation

The stimulability task in this study was designed to include all Arabic speech sounds (for MSA and UHA). It aimed to investigate the children's ability to produce all the 29 Arabic phones (consonants of both UHA and MSA) and the three Arabic long vowels in isolation. Short vowels and diphthongs were not assessed in this task, as MSA has a simple three–vowel quality system (Almbark & Hellmuth, 2015), and the short vowels only differ from the long vowels by the vowel length as the vowel duration is the most important feature that classifies the Arabic vowels, and the UHA vowels in particular (Almurashi et al., 2020). Moreover, the duration of vowel production in ordinary speech is not consistent or organized, and this distinctive length does not occur in adult Arabic speakers (Cowan, 1970). Thus, it is difficult to distinguish the production of long versus short vowels, especially in isolation and by children.

3.3.2.1 Administration procedure of the stimulability task: Phone imitation

Each child was asked to repeat the UHA and MSA sounds once in isolation after the researcher's model. If a child could not imitate a phone correctly after the first trail, the researcher asked him/her one more time to listen carefully and then she repeated the sound again. If the child did not imitate her after the second trail, the researcher asked the child to look at her face and imitate the sound after she reproduced it. If the child could not imitate the correct production of the phone, the researcher recorded the child's actual production on the datasheet (see Appendix O). No sensor prompts were given on how to articulate an unimitated sound. The approximate duration of this task was 3 to 5 minutes.

3.4 Procedures

Data collection for the *single-word naming* and *stimulability* tasks were conducted at the child's school and sometimes with a teacher's attendance. All test sessions took place in a quiet room (e.g. a library or resource room) during school hours. The child's responses were audio–recorded using the Audacity®2.3.3 application. A Samson® microphone (C01U Pro–USB Studio Condenser Microphone) was connected to a Mac laptop to record children's responses from both tasks during testing. After the single-word naming task, and within the same session, the researcher directly conducted the phonetic imitation task. It took approximately 3 to 5 minutes.

During the session, the child was seated on a suitable chair in front of the laptop and microphone, which was placed about ten inches away from the child's mouth to ensure a clear recording. The researcher began the task by familiarising the child with the session content to ensure he or she was willing to participate. The researcher, who is a native speaker of UHA and an experienced SLT, phonetically transcribed all the participants' responses online using broad phonetic transcription (i.e. the International Phonetic Alphabet [revised to 2018] and the Extended International Phonetic Alphabet [Ext–IPA, revised to 2008]). Additionally, the whole session was audio–recorded to later check all the transcriptions and fill in the missing details.

Finally, after stopping the recording, each participant's response was given a code number, and the audio file was saved directly on an encrypted file on the laptop. At the end of each testing day, all these files were downloaded on a password–protected external hard drive.

3.5 Data preparation

3.5.1 Single-word naming task

The first step in preparing the single–word–naming task data was checking all the audio files saved on Audacity® and applying noise reduction features to any unclear/noisy recording. After that, the researcher converted all these Audacity® audio files into MP3 audio files so they could be uploaded into PHON–alpha. Then, she uploaded the MP3 audio files into the PHON® 3.1.0–alpha.6 software. This software is a free program used to analyse phonological data (Hedlund & Rose, 2019; Rose & MacWhinney, 2014). A project labelled 'Arabic' was created on PHON. Within the PHON project's corpus, a folder for each age group was created, where a session for each participant was created and labelled using the child's anonymous code. Before entering the participants' information, the researcher uploaded an Arabic template of IPA transcriptions of the targeted word list of the single-word naming task into PHON. Then, for each participant, a session was created using the saved template where the IPA target tier and IPA actual tier were automatically filled with each word's correct form. After that, each child's audio file was uploaded to be attached to the child's session file. This audio file was then segmented into sections focusing on isolating each child's response to be connected to its target word. After the segmentation process, the researcher edited the actual IPA tier to match the child's pronunciation. By performing this step, each child's session saved included the adult IPA target tier and the child's IPA actual tier matched to the audio segment for a reliability check of the IPA transcriptions (see Figure 3.3).

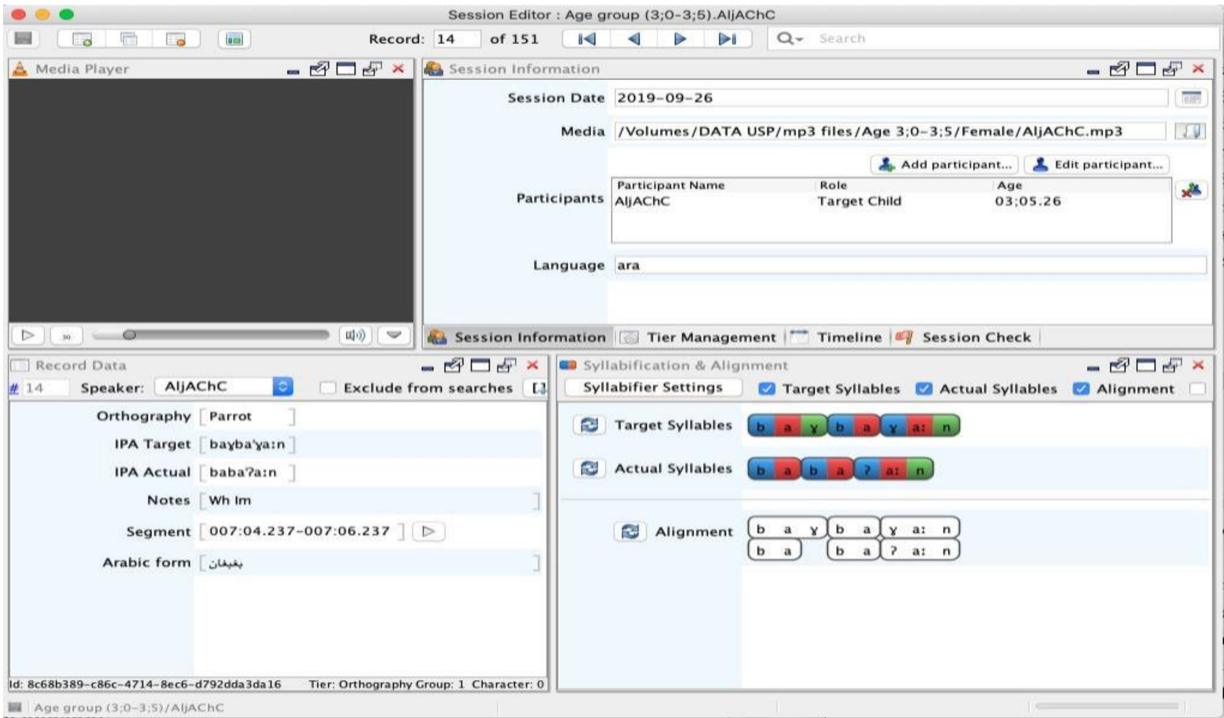


Figure 3.3 *PHON session picture showing the session information, record data, and syllabification and alignment.*

All transcriptions were entered into PHON and cross checked with the recording. The researcher then scored and analysed all the data according to the scoring criteria described in the following section.

3.5.2 Stimulability task: Phone imitation

The researcher tabulated the data collected from all participants into Excel using a separate sheet for each age group. Then, she indicated, on Excel, each child's response as correct, substitution, or null responses.

3.5.3 Missing data

A few children refused to complete the single-word naming task or refused to response during the stimulability/phonetic imitation task. For the single-word naming task, missing data occurred mainly in the youngest age group (2;6–2;11), where ten children did not complete the test. For example, one of the children in this age group only named 55 words; however, these named words were still included in the analysis for several reasons, which concur with those put forward by James (2001a). First, they reflect the range of words that could be named by children at this age. Second, the words named by this age group (2;6–2;11) had varying syllables length and shapes, reflecting that the children's productions were not restricted to simple words. Missing data (unnamed words) were excluded from each child's analysis. However, the children who did not complete the test were excluded from the test duration analysis (see Section 3.6.1). For the phonetic imitation task, 32 children (13.67% of the whole sample) did not complete this task across all the age groups due to different reasons, including the loss of interest in imitating any sounds, time constraints, or fatigue after participating in the lengthy naming task. However, there were still enough data in each age group for each task to be analysed.

3.6 Scoring and analysis

All the responses collected from the 235 SHA-speaking children were scored and analysed separately for each participant. The steps for the scoring and analyses conducted included evaluation the SHAPA test, analysing the phonetic acquisition by calculating the PCC and compiling a phonetic inventory via the two tasks: single-word naming task and stimulability task, and analysing the phonological variants (PhonVar) as well as infrequent variants (InfrVar). Each task is described in detail below.

It is important to note that for scoring and analyses of the phonetic acquisition, all the analyses were carried out twice, once by considering the distorted production of phones that prone to distortion in UHA (i.e. /s, z, ʃ, ʒ, r, l/) as an error, and once by considering them as correct. Wertzner, Sotelo, and Amaro (2005) described the common type of distortion as the phonetic alterations observed in the speech of normally developed children, and the uncommon type as the distortions that may occur in subjects with speech sound disorders. In this study, and according to Wertzner and colleagues, any phonetical alterations characterised by some difficulty in motor abilities involved in speech production were considered as phone distortion. These alterations do not affect the phonological contrasts of the language under study, and therefore do not affect the meaning of the words (Wertzner et al., 2005). Regarding the Arabic language, the trill /r/ and the sibilants /s/, /sˤ/, and /z/ are the most distorted ones (Ammar & Morsi, 2006; Dyson & Amayreh, 2000). The following phonetic variants was considered in the scoring of distortion are:

- Lateralisation and palatalisation of alveolar (e.g. /s/→sˡ], [sʲ]) or postalveolar fricatives (e.g. /ʃ/→ [ʃʲ], [ʃˤ] or [ç]) were considered as phonetic distortions of the target sound.

- The /r/ sound is naturally realised as either a trill /r/ or a flat /ɾ/ in UHA, and both productions were marked as correct. However, the alveolar or retroflex approximant [ɹ] or [ɻ] were considered as phonetic distortions.
- The lateral approximant /l/ was marked as distorted if it was produced as retroflex /ɭ/ or /ɫ/.

All these alterations in the pronunciation of phones prone to distortion were considered as accepted phonetic distortions. Further, some other special considerations that were taken into account when entering the single-word naming task data to PHON.

First, when children changed a word's vowels, especially in the initial syllables, to another vowel in dialectal variation (e.g. /nimir/→[namir] 'tiger'), such a change was scored as correct as long as the changed form was acceptable among Hejazi adults. It is known that vowels tend to vary more than consonants between dialects and accents (Foulkes, 2006). The vowel change was marked as an error only if the child changed the vowel to one not in the UHA vowel inventory (e.g. /ɛ/) or if the changed vowel was an unacceptable variant of the target word among Hejazi adults.

Second, if a child named a word using a different form, especially for cogent words, where UHA words are only differ from MSA form by replacing dental phones, (e.g. UHA /dahab/ verses MSA /ðahab/ 'gold'), such a word or form was used as a target for the specific child, and the child pronunciation was marked as correct production. However, if a child named the MSA form of the unique words, where the UHA words are differ completely from MSA form, (e.g. in UHA / ʃubbaːk/ named as /nafiða/ in MSA 'window'), the researcher tried to instruct the child to name the picture using the form they would use at home (i.e. the UHA dialect form).

Third, using a different form of the target items, such as naming the plural form rather than the singular, or vice versa (e.g. /tuf.faːħa/→ [tuf.faːħ] 'apples', /moːz/→[moːza] 'a banana'), as well as adding some morphemic ending to the word (e.g. [raːsi] 'my head' instead of /raːs/ 'head') were all counted as acceptable variants of the target words.

Fourth, another consideration was given to the addition of the definite article /ʔal/. Such an inclusion was marked as an acceptable pronunciation, and the target word was changed accordingly (see Chapter 2, Section 2.2.1.2.3). The different dialectal patterns were also considered, and they were all scored as correct (see Appendix P for a description of these patterns).

It is important to note that all the acceptable changes to the target words were considered during data entry to PHON. The IPA target on the PHON file was changed to match the child response (IPA actual), so it was not considered an error during the PCC and phonetic inventory calculation.

3.6.1 SHAPA test evaluation

To evaluate the SHAPA test, the researcher assessed features such as test duration, the suitability of the items (item analyses), and inter-and intra-rater reliability. These analyses were considered important because SHAPA is a new test developed for this study (see Chapter 1, Section 1.4.2.3).

First, to evaluate the applicability of SHAPA and whether the test duration is appropriate for use by SLTs during their clinical practice, as well as whether the test is suitable for young children (i.e. for the 2;6–5;11 age rang in this study), the researcher examined the amount of time taken to complete the SHAPA naming task for each participating child, except for ten children where the time required to complete the test was not reported (i.e. giving a total of n=225) (see Section 3.6). Although this characteristic was not considered when the test was initially piloted, it is an important feature that could reduce the test validity, as reported in Chapter 1, Section 1.4.1.

Second, item analyses were also conducted in order to identify how familiar SHA-speaking children aged 2;6–5;11 years were with items in the created test and how suitable they were to elicit suitable responses (see also Chapter 1, Section 1.4.1) (Flipsen & Ogiela, 2015). This analysis included the following evaluations:

1) *Item difficulty measure*. As reported in Chapter 1, Section 1.4.1, the target words of any speech tests should not be difficult for the targeted children and be named spontaneously. To ensure the presence of this feature in SHAPA, the researcher conducted an item difficulty measure in three steps following what is reported in the literature (see Section 1.4.2.3):

    a. First, to identify the number of all items named spontaneously across all age groups, the percentage of children who could spontaneously name an item was calculated for each item. A word was considered spontaneously named even if its form did not match the intended target. For example, if the target word was /moːz/ 'banana', and a child spontaneously named it as [moːza] 'a banana', this word was scored as correct for spontaneously naming. If an item was named spontaneously by an average percentage of 50% or higher of children across all the age groups, this item was considered 'a familiar item'. This 50% criterion was adopted following James (2001a).

b. Second, to calculate the number of items named by whole-word imitation, similar to the spontaneous naming calculation, if an item was named by imitation by an average percentage of 50% or higher of children across all age groups, this item was considered as 'a difficult item'.

c. Third, the number of items named by prompts other than whole-word imitation (i.e. description, alternative questions, and first sound imitation) were also calculated to provide further assessment of the item difficulty measure.

2) *Accuracy of labelling the intended target word*. Each child's responses were evaluated in terms of matching the target words in the test (i.e. the intended target words). If a child named an item using a different vocabulary, this item was scored as inaccurately named; this included, for example, using the singular or plural form of the word (see Appendix I). Nojavan-Pirehyousefan et al. (2021) and James (2001a) reported the accuracy of naming (labelling) the intended target words as one of the measures for test item analysis. They assessed the accuracy of the naming measure by calculating the number of participants who named an item in line with the intended lexical item in the test. For SHAPA, the researcher conducted a similar calculation to check picture-target word naming agreement by counting the number of items named differently from the intended target in each age group for each test item, either spontaneously or by other cueing, but not by whole–word imitation. As mentioned earlier, if a child named an item using a different form, this item was labelled as inaccurately named for this measure only; this included using another acceptable form of the word. For example, if the intended word was /moːz/ 'bananas', but a child spontaneously calls it [moːza] 'a banana', this word was considered spontaneously named but not accurate because it did not match the intended target /moːz/ where /z/ is in the word-final position. Such a measure is helpful to determine the degree to which the participants agree on the name of a given image as there is no database for expressive words in Saudi Hejazi Arabic. According to Nojavan-Pirehyousefan et al. (2021), $\geq$70% accuracy of naming the intended target needs to be met to judge the appropriateness of picture-word agreement. However, if one item had never been named so that it matched the intended target even by children in the oldest age group, then such an item would not be suitable for a phonology test, and either the picture or the item has to be changed. This is because the phonetic-phonological features of the chosen target would not be elicited by the children.

3) *Articulation accuracy of the named items.* Analysing the test items based on the children's pronunciation is necessary for evaluating the articulation accuracy of the named items. The SHAPA test is intended to evaluate the articulation and phonological abilities of children.

Therefore, if any items are named correctly without any error 100% of the time by all children across all age groups, then these items would not be of value for the test. According to James (2001a), items included in a speech test should differentiate children with age-appropriate skills from those who have speech impairment. Therefore, the number of items named with accurate articulation was determined to identify any items with low value for SHAPA.

Inter-rater and intra-rater reliability checks were also applied. Ten percent of the data across all age groups were randomly selected (four participants from the 2;6–2;11, 3;0–3;5, and 3;6–3;11 age groups, and three participants from the 4;0–4;5, 4;6–4;11, 5;0–5;5, and 5;6–5;11 age groups), re-analysed, and re-transcribed by the researcher to check the intra-rater reliability. This included the re-transcription of all named items and the re-analysis of phonetic and phonological variation. A further 10% of the sample data was randomly selected from all age groups to reanalyse the phonological patterns identified by the researcher, considering the patterns identified using ≥4 cut-off criterion. The researcher sent the audio file of 23 participants and a copy of each child's PHON session without the researcher's transcriptions to an experienced SHA-speaking SLT, who received training on using the PHON software by the researcher. The SLT was asked to re-transcribe the words and enter the IPA transcription of each word under the 'IPA actual' using the segmentation recording in PHON. The transcription agreement was calculated through point-by-point analyses of phones (vowels and consonants) for the IPA phonetic transcription. No words were excluded from the analyses.

The phonological patterns were also checked for inter and intra–rater reliability. For inter-rater reliability, the SLT received Excel files for the same 10% of the participants from the researcher and was asked to mark the variation in the children's productions according to the adult targets. Then, the patterns identified by both (the SLT rater and researcher) were reviewed by the researcher and all the agreed patterns were counted (considering the ≥4 cut-off criterion only) to calculate the agreement percentage. For intra-rater reliability, the researcher revaluated 10% of the children's responses, and re-identified the phonological patterns. Similar to the inter-rater reliability calculation, the researcher counted the agreement between the two evaluations and then calculated the percentage of agreement. The results of all these analyses will be presented in Chapter 4, result and discussion: SHAPA test evaluation.

3.6.2 Phonetic acquisition analysis

3.6.2.1 Percentage of consonants, vowels correct: PCC, PCC-R, and PVC

The analysis conducted to measure the speech sound accuracy included: the percentage of consonants correct (PCC), percentage of consonants correct-revised (PCC-R), and percentage of vowels correct (PVC). PCC-R was included in the analysis because it measures all common and uncommon clinical distortions as accepted phonetic alterations to measure the percentage of *phonemic* consonants correct. According to Shriberg et al. (1997a), PCC-R is the most appropriate metric to compare between speakers of diverse age and speech status, which is the case of the children in this current study.

The scoring for these metrics was run through PHON following Shriberg et al. (1997a). Frist, the PCC was calculated through PHON, where all the distorted productions of the phones prone to distortion were counted as errors. To calculate PCC-R, another analysis was run through PHON, and the distortion of all phones was considered as correct. A further comparison was conducted between PCC and PCC-R to examine the difference in the correct percentage score when including all clinical distortions versus excluding them.

3.6.2.2 Phonetic inventory

The phonetic inventory was established twice. First, it was calculated on the basis of the data derived from the single-word naming task, and then from the stimulability task. As reported earlier in the literature review chapter (Chapter 1, Section 1.3.1), compiling a phonetic inventory using both the single-word naming and stimulability tasks could provide detailed information about the phonetic abilities of the typically developing children and may help to predict the occurrence of speech sounds in a child's phonetic–phonological system (de Castro & Wertzner, 2012; Miccio et al., 1999; Powell & Miccio, 1996).

3.6.2.2.1 Compiling the phonetic inventory via the single-word naming task

For compiling the phonetic inventory via the single-word naming task, a phone had to be produced at least twice correctly throughout the speech sample in any word position–whether this position was correct or not to be considered as a part of the phonetic inventory of an age group. There was an exception in compiling the phonetic inventory in counting some phones that have fewer opportunities of occurrences (those that occur twice or less in each word position, see Table 3.5) in the naming task (i.e. aiː, auː, ɑː, plus the MSA phones: q, θ, ð, ðˤ). These were all counted as being part of the inventory if they appeared at least once in a child's single-word naming data. Then, the number of children who produced the target numbers of a phone (at least twice, or at least once for the phones with exceptions) were counted for each phone in each age group. In a further step, the phones found in at least 75% and at least 90% of the children within the same age group were considered as acquired and mastered, respectively, following the criteria used by Fox (2000).

The phonetic inventory compiled via the single-word naming task was generated using PHON. The numbers of phones produced by each child and their frequencies in the speech sample of each child were produced via PHON. Then, the researcher calculated the number of children in each age group who produced each target phone at least twice using Excel. This calculation was carried out twice, as reported earlier, once with distortions counted as errors, and once all the distortions counted as correct. In a second calculation, two criteria were used, following Amayreh and Dyson (1998), which considered a phone to be *mastered* if ≥90% of the children in an age group produced it correctly (see Section 3.6 for scoring). A phone had to be correctly produced by 75–89% of the children in an age group for acquisition production.

3.6.2.2.2 Compiling the phonetic inventory via stimulability task

For each child it was noted and categorised how the child produced each sound in isolation presented by the researcher (correct, phonetic distortion, substitution). If the child was able to imitate the phone in isolation correctly once after the first or the second trail, the sound was marked as correct. Any misarticulation attempt was marked as incorrect. A second calculation was done to include all the distortion sounds as correct for further comparison. The percentage accuracy of phone imitation, where each phone was imitated once by each child in an age group, was calculated by counting each child's correct response for each phone in a certain age group. Then, the total number of each accurate imitated responses was divided by the total number of children in that age group. Since the speech sounds of both UHA and MSA were targeted, the total score for this task was 32.

The two inventories (obtained via the single-word naming task and via the stimulability task) will be compared.

3.6.2.3 Consonantal co-occurrence

This study also analysed the acquisition of the consonantal co-occurrences including word-final CC, heterosyllabic CC, and geminated consonants among SHA-speaking children and their strategy to deal with these syllable structure types. These features were scored correct if:

1. A child produced both elements of the CC (word-final or heterosyllabic) phonetically correctly, without deletion, substitution, epenthesis, or distortions.
2. A child produced both geminated consonants phonetically correctly across the syllable's boundary (SFWW and SIWW) without deletion, substitution, epenthesis, or distortions.

Any epenthesis, substitution, reduction of the cluster, and/or degemination of the geminated consonant was marked as an error.

It was of interest to analyse these features to understand their acquisition among SHA-speaking children despite of their limited frequency in SHAPA. All the analyses were done using PHON:

1) To identify children's word-final CC correct production (PCCC), the data from the nine items in the SHAPA word list (see Table 3.6) were used, where PHON generated all children's responses that included words with a final CC. The researcher than counted every correct production of both elements of a cluster in an Excel file.
2) For the heterosyllabic CC analysis, PHON generated all children's responses that included words with heterosyllabic CC. The researcher also counted the correct production of both consonants across syllables using Excel.
3) The correct productions of geminated consonants were also counted in Excel, after all the children's responses that included words with gemination were generated using PHON.

The criteria used for singleton phonetic mastery and acquisition (90% and 75%) were followed for consonantal co-occurrence (see Section 3.6.2.2). It is important to note that the analysis of consonantal co-occurrences was carried out only once by considering all distortions as correct.

3.6.3 Phonological variants (PhonVar)

The data collected were used to describe the occurrence of phonological variants from adult-like speech. All phonological variants per child were analysed by hand. The number of all phonologically varied productions (Tokens) per child were counted. Then, these variants were differentiated into infrequent variants (InfrVar) and phonological patterns (Types) depending on their frequency of occurrence, where the frequent occurrence of a pattern represents a certain tendency in a child's speech (Dodd et al., 2003). An InfrVar was defined as a single instance of a variant which could occur by chance or due to developmental instability (Dodd, Hua, Crosbie, & Broomfield, 2006). Variants occurring less than four times, and in a second analysis less than six times within the sample, were summarised as InfrVar (Fox-Boyer et al., 2021). The criterion of four or fewer occurrences of a pattern could not be strict enough and causes some non–developmental processes to be identified as developmental (Fox, 2007). However, it was used in this study for a comparison with the available literature. According to Kirk and Vigeland (2014), higher cut-off criteria could be used and be more informative, especially for patterns involving a large number of different consonants. The reason for adopting two types of cut-off criteria to identify phonological patterns is the lack of scientific basis and empirical data for the universal cut-off criteria that used in many studies to determine the phonological patterns (Kirk & Vigeland, 2014).

Previous studies on Arabic had identified a large number of phonological error patterns. Sixteen of those were reported most frequently among Arabic-speaking children. As described in Chapter 2, Section 2.2.2.1, although previous studies used similar labels for some patterns, significant variants can be noted in these patterns' definitions. Therefore, in this current study each pattern was defined in detail. For example, the error pattern of 'fronting' was specified as follows: the fronting of the postalveolar fricatives (ʃ/ʒ), fronting of the alveolar fricatives (s/z), and fronting of the velar stop (k/g). The rationale for such a decision was supported by Kirk and Vigeland (2015), who suggested that some error patterns depend on the phoneme features and positions. Further, each phonological variation deviating from the target word was noted and labelled as detailed as possible (i.e. sounds or word positions affected were noted). Potential patterns emerged from the children's data and were not restricted to a list of patterns found in previous studies.

A pattern was considered to be present in an age group and therefor developmental at that age if it was found in at least 10% of children of this age group (Dodd et al., 2003). The results from the two criteria will be compared. It is important to note that this step of analysis was conducted in Excel, where PHON used only to generate the IPA phonetic transcriptions of each child in each age group. The phonological variation results are reported in Chapter 6.

4 Results and Discussion I – Evaluation of the Saudi Hejazi Arabic Phonology Assessment (SHAPA)

In this chapter, the results of the administration of the SHAPA test will be presented and analysed in order to address the following research question that outlined in Chapter 2, Section 2.5:

> RQ1: Can a phonological assessment tool (single-word naming test; SHAPA, be developed based on international construction criteria leading to a clinically valuable tool for assessing the phonetic and phonological development in SHA-speaking children?

The evaluation of the implementation of SHAPA is considered important because the administrative and psychometric properties of this newly developed test are unknown. This chapter is divided into two sections. First, the result section where the duration of the test administration is analysed and presented, followed by test items analyses, and reliability analysis. The second section documented the discussion of these results.

4.1 SHAPA: Test administration

The 151 words of the SHAPA test were analysed according to participants' responses to measure the test duration. As reported in Chapter 3, Section 3.6.1, the duration of SHAPA administration was analysed to examine whether its administration time is appropriate for use by SLTs, considering their clinical–time constraints, and whether it is suitable for children aged between 2;6 and 5;11. The time required to complete the single-word naming task was calculated for all participants, except for ten children where this information was missing due to the incomplete testing (n=225)[14]. The average time needed to name all the pictures in the test is reported in Table 4.1 below.

Table 4.1 *Descriptive statistics (M, SD, Min, Max) for administration duration of SHAPA across age groups*

	5;6–5;11	5;0–5;5	4;6–4;11	4;5–4;0	3;6–;3;11	3;0–3;5	2;6–2;11	Average across all age groups
Mean	00:20:54	00:20:59	00:25:54	00:28:07	00:35:58	00:39:22	00:40:16	00:30:13
SD	00:05:56	00:04:38	00:08:37	00:08:45	00:10:40	00:08:48	00:08:52	00:08:02
Min	00:14:09	00:14:17	00:13:57	00:12:43	00:17:57	00:17:54	00:24:54	00:12:43
Max	00:39:02	00:35:30	00:44:42	00:50:47	01:01:38	01:04:24	00:54:12	01:04:24

[14] As reported in Chapter 3, Section 3.3.1, some children did not name all the pictures in the naming task; however, their responses were included in the phonetic–phonological analysis but not in the test duration analysis.

The mean duration of the test administration was approximately 30 mins, with a range between 12:43 and 01:04:24. In general, the time taken to complete SHAPA decreased with increasing the age of participants. The three youngest age groups (2;6–2;11, 3;0–3;5, and 3;6–3;11) needed the longest duration to complete the naming task, as these children required frequent reinforcements and praise to maintain their attention and cooperation due to their age. In contrast, the oldest age group needed only half the time required by the youngest age group. However, the mean duration of the SHAPA administration is a feature that could threaten its validity because the recommended administration time for a test intended to measure children’s speech sounds skills is usually between 10-20 mins (cf. Abou-Elsaad et al., 2009; Ceron, Gubiani, de Oliveira, & Keske-Soares, 2020; Chen, Bernhardt, & Stemberger, 2016; Nojavan-Pirehyousefan et al., 2021; Prezas et al., 2014; Stoel-Gammon & Williams, 2013; Tyler & Tolbert, 2002).

4.2 Test items analysis

The researcher applied item analysis checks as a content validity measure to identify the test items difficulty for SHA-speaking children aged 2;6–5;11 years and to examine the children’s abilities on the following criteria:

1) To identify and name the items pictures’ spontaneously versus via imitation (item difficulty).
2) To name the test item as the intended target word (the accuracy of naming the intended word).
3) To pronounce the target word correctly without any error in all age groups (the articulation accuracy).

4.2.1 Spontaneous versus imitated responses

Following James (2001a), the SHAPA words were first coded for spontaneously named responses irrespective of the accuracy of their articulation. Subsequently, the items were allocated to one of ten percentage bands ranging from 0–10% and ranging to 91–100% (see Figure 4.1).

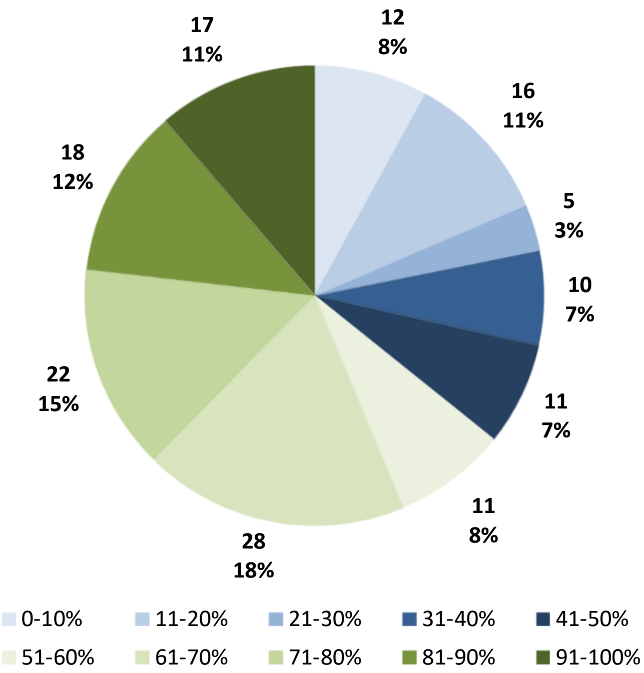


Figure 4.1 *Percentage and numbers of test items named spontaneously***.**

Note: The green–coloured pie sections refer to items spontaneously named by more than 50% of the children. The blue–coloured pie sections refer to those items that could be spontaneously named by fewer than 50% of the participants. The whole pie chart displays the number of items named spontaneously (the upper number) and the percentage (the lower number) of the items in the test per percentage range. See Appendix Q for details

Figure 4.1 shows that in total 64.24% (n=97) of the SHAPA items were named spontaneously by more than 50% of the children. It can be noticed that only eight items were named spontaneously with a percentage range between 91 and 100%, with no one item named spontaneously by all children across all age ranges. The highest percentage of spontaneous naming (96.17%) is accounted for by one item (i.e. /ʒauːˈwaːl/ 'cell phone'). The percentages of all items named spontaneously >90% by more than 50% of children across all the age groups is provided in Appendix R, Table 1. Interestingly, only four items were named 100% spontaneously by the three oldest age groups (5;6–5;11, 5;0–5;5, 4;6–4;11). These are: /beːdˤ/ 'eggs', /samaka/ 'a fish', /saʕa/ 'watch', and /ballona/ 'a balloon'.

A further observation of Figure 4.1 is that on average 35.76% (n=54) of the words were named spontaneously by fewer than 50% of the children. Thirteen items were named spontaneously by a percentage range of 0–10%, while the word with the lowest naming percentage (0.85%) was /ˈjimdˤuɣ/ 'chewing' (see Appendix R, Table 2).

In general, the spontaneous naming rate increased with age, where the number of items named 100%

spontaneously in each age group showed a steady increase across age groups (see Table 4.2).

Table 4.2 *Number of items named 100% spontaneously by all children in each age group*

Age group	2;6-2;11	3;0-3;5	3;6-3;11	4;0-4;5	4;6-4;11	5;0-5;5	5;6-5;11
N (%)	0	0	2 (1.32%)	7 (4.64%)	12 (7.95%)	17 (11.25%)	30 (19.9%)

It is clear that no one item in SHAPA was named spontaneously 100% of the time by all children across age groups, this because no one item was named spontaneously by all children in the two youngest age groups.

In addition, the number of all imitated responses was calculated in a similar way to the calculation performed for spontaneously named items (i.e. the items named with imitation were allocated to one of the ten percentage bands ranging from 0–10% and ranging to 91–100%).

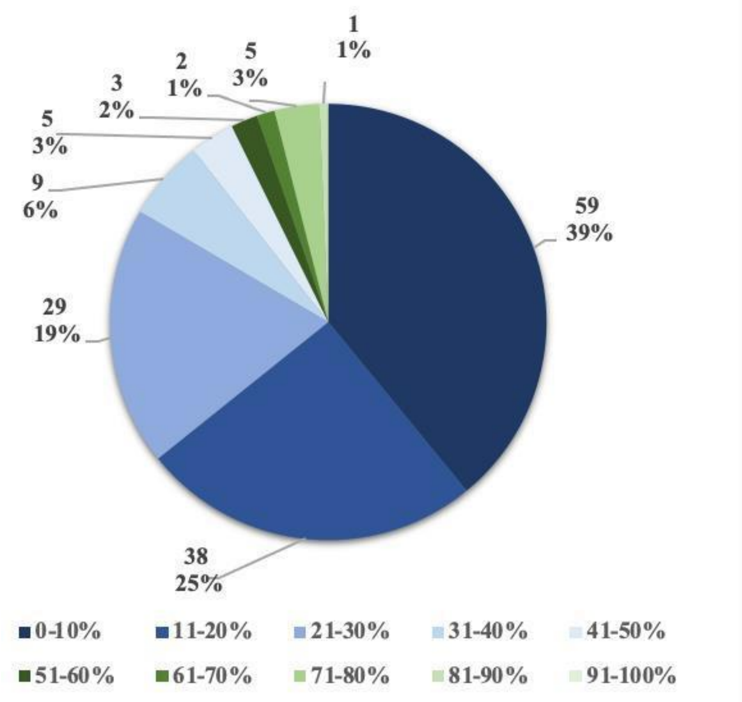


Figure 4.2 *Percentage and number of test items named by imitation*

Note: The green–coloured pie sections refer to items named by imitation by more than 50% of the participants. The blue–coloured pie sections refer to items named by imitation by fewer than 50% of the children.

Figure 4.2 shows that in total 7.28% (n=11) of the test items were named with whole–word imitation cueing by more than 50% of the children across all age groups. These could thus be classified as difficult items. It is also clear from the figure that no one item was named by imitation by a percentage range of 91–100% of all participants. It is worth mentioning that ten of the 11 words that were named with imitation by more than 50% of SHA-speaking children were items that were selected for their specific phonetic–phonological features and they were not included in the JA-CDI Saudi version (see Chapter 3, Section 3.3.1.1). In general, the number of items named spontaneously by more than 50% of children (n=97) is much higher than the number of pictures named via imitation cues (n=11) (see Appendix S for items named by imitation by more than 50% of children).

Considering items named with other cues, 7.95% (n=12) of SHAPA items were named either by description, an alternative question, or by first sound imitation by more than 50% of the children across all age groups (see Figure 4.3).

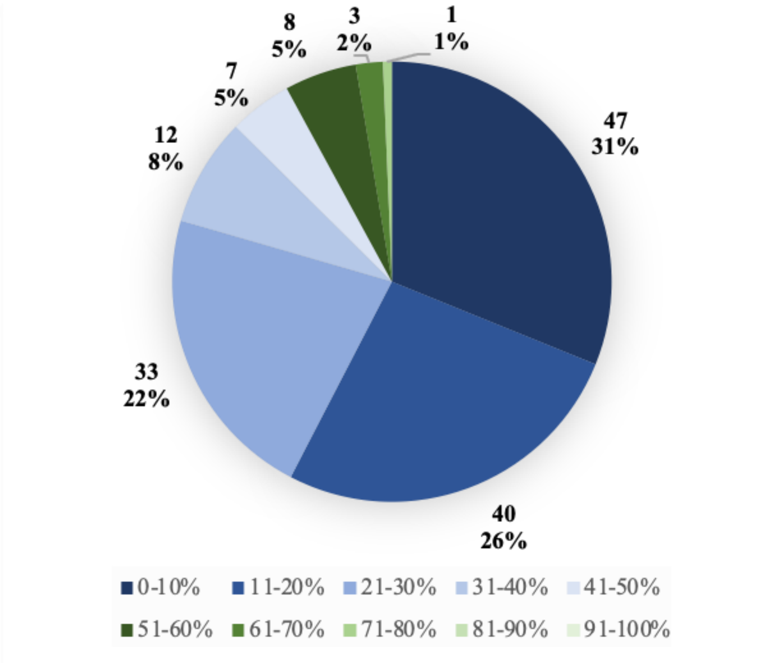


Figure 4.3 *Percentage and number of test items named with other cues (description, alternative question, and first–sound imitation).*

Note: The green–coloured pie sections refer to items named by other cues by more than 50% of the participants. The blue–coloured pie sections refer to items named by other cues by fewer than 50% of the children.

It is evident that 12 items of the SHAPA items were named by other types of prompts by more than 50% of children across all age groups. The word /ˈlaʔ/ ‘no’ was the item that needed a prompt in order to be named by most children across all age groups (73.19%), while the item that needed a prompt by the fewest number of children was /nadˤˈdˤaːra/ ‘glasses’ (i.e. 0.95%).

4.2.2 Accuracy of labelling the intended target word

As reported in the methods chapter, Section 3.6.1, the accuracy of naming the intended target word was measured. Similar to the calculation of spontaneous responses, the naming agreement of an item was allocated to one of the ten percentage bands ranging from 0–10% and ranging to 91–100% (see Figure 4.4).

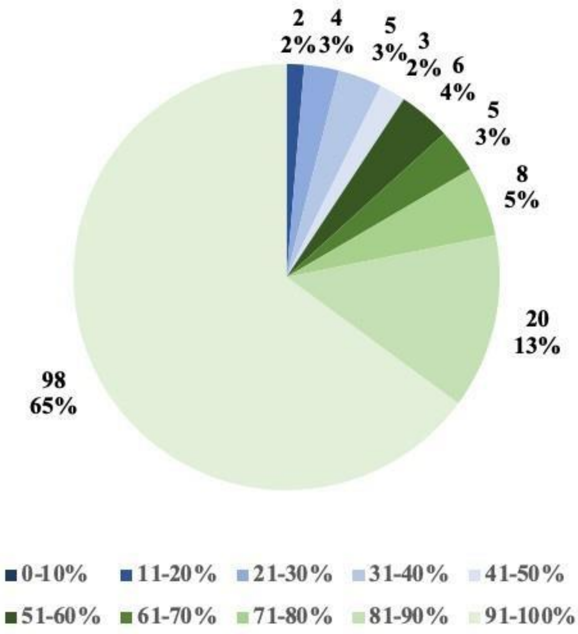


Figure 4.4 *Percentage and number of naming agreement with the intended target lexical items either named spontaneously or by other cueing but not imitation*

Note: The green–coloured pie sections refer to the percentage and number of items named with the intended target by more than 50% of the participants. The blue–coloured pie sections refer to the percentage and number of items named as an intended target by fewer than 50% of children across all age groups.

Most items were shown to be acceptable measures of item suitability in terms of naming agreement, as shown in Figure 4.4. In total, 90.73% (137 out of 151) of the test items were named as intended targets by more than 50% of children across all age groups. Sixty–five per cent of the words (n=98) were named as the intended target by 91–100% of participants; in addition, 13.25% (n=20) of the test items were named as the desired responses by 81–90% of participants. Only 9.27% (n=14) of items were named as the intended target by fewer than 50% of the children across all age groups (see Appendix T), meaning that the majority of the SHAPA items were able to ensure the production of the intended lexical name.

4.2.3 Articulation accuracy of the named items

Another measure applied to check the suitability of items selected for SHAPA was assessing the accuracy of items pronunciation by children across all age groups to eliminate or exchange any items named correctly all the time in all groups. This analysis showed that an average of 62.91% (n=95) items were named with accurate articulation by more than 50% of children across all age groups, with the majority (28%, n=42) items named with accurate articulation by 51–60% of the children (see Figure 4.5).

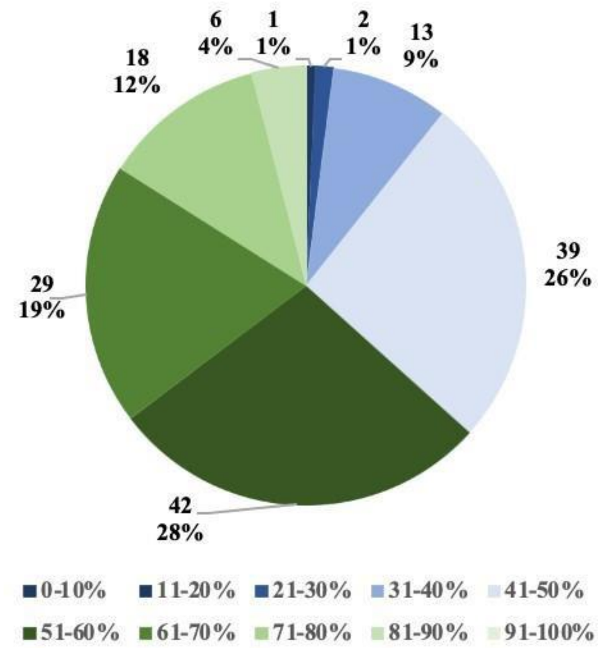


Figure 4.5 *Percentage and number of items named with accurate articulation*

Note: The green–coloured pie sections refer to the percentage and number of items named with accurate articulation by more than 50% of the participants across all age groups. The blue –coloured pie sections refer to the percentage and number of items named with accurate articulation by fewer than 50% of children across all age groups.

It is clear from Figure 4.5 that no item was named with correct pronunciation by 91–100% of children in all age groups. It also shows that only six items were named with accurate articulation by 81–90% of children across all age groups, which are: /kalb/ ‘dog’, /fiːl/ ‘elephant’, /ħaˈliːb/ ‘milk’, /wa.lad/ ‘a boy’, /ˈgalam/ ‘pen’, and /ˈlaʔ/ ‘no’. Similarly, it is evident that no item was named with incorrect articulation by 100% of all children. The item that received the lowest average (i.e. 13.98%) in the correct articulation naming was /ˈmasʒid/ ‘mosque’.

4.3 SHAPA: Reliability

The results of the intra– and inter–reliability analyses are described below and shown in Table 4.2.

4.3.1 Intra–rater reliability

For intra–rater reliability, the point–by–point analyses revealed an agreement of almost 90% for IPA transcription and phonological variation (see Table 4.2). For IPA transcription, the mean percentage of agreement is almost 98.68%, with a narrow range (i.e. 98.14–99.16%). For the phonological patterns identified, the mean agreement percentage is 96.15%, with SD = 4.01 and a slightly wide range (i.e. 91.75–98–90%). Appendix U (Table 1 and 3) presents all the percentage agreement reanalyses of the intra–rater reliability across all age groups for IPA transcription and phonological variation.

4.3.2 Inter–rater reliability

The inter–rater point–by–point agreement on the IPA transcriptions showed a high mean of agreement percentage (i.e. 98.21%) with a narrow range (i.e. 97.24–98.91%) (see Table 4.3). For the phonological variants, the mean agreement score is 89.90%, SD= 0.06, with a slightly wide range (i.e. 84.11–95.58%). Appendix U (Table 2 and 4) presents all the percentage agreement reanalyses of the inter–rater reliability across all age groups.

Table 4.3 *Intra– and inter–rater reliability of the single-word naming task*

Reliability check	IPA transcription		Phonological–pattern analysis	
	M (SD)	Range	*M* (SD)	Range
Intra–rater	98.68% (0.54)	98.14–99.16%	96.15% (4.01)	91.75–98.90%
Inter–rater	98.21% (0.83)	97.24–98.91%	89.90% (0.06)	84.11–95.58%

This result showed that the SHAPA test met two kinds of reliability criteria: inter and intra–rater reliability for both tasks (IPA transcriptions and phonological variants reanalysis) and achieved agreements higher than 90% on average for both tasks.

4.4 Discussion I –Evaluation of the SHAPA test

The SHAPA test was used initially for the first time to collect speech samples from 235 SHA-speaking children to analyse their phonological development. This section discusses the effectiveness of SHAPA at eliciting target words regarding its reliability, administration, and content validity, including the item difficulty measure and picture–naming agreement. Although a pilot study was conducted to evaluate this newly developed tool, it was expected that the larger main study would reveal further information on test items' suitability, which could form the basis for a future revision of SHAPA based on the findings of the current study.

4.4.1 Test administration

The amount of time taken to administer of the single-word naming task to SHA-speaking children was, on average, 30 minutes, with a range between 12 minutes and 01;04;00 (one hour and four minutes). This result, at first glance, shows that the time required to complete naming the 151 items in the SHAPA test by SHA-speaking children is longer than that usually reported in other studies, which is on average 17 minutes (cf. Abou-Elsaad et al., 2009; Ceron, Gubiani, de Oliveira, & Keske-Soares, 2020; Chen, Bernhardt, & Stemberger, 2016; Nojavan-Pirehyousefan et al., 2021; Prezas et al., 2014; Stoel-Gammon & Williams, 2013; Tyler & Tolbert, 2002). However, a review of the number of items reported in these studies (i.e. *M*=76 items) revealed that the mean number is almost double the number of items in SHAPA. Further, reviewing the criteria used in developing these reported tests revealed that the majority of them did not fully include many of the criteria needed for a comprehensive phonological assessment. For example, the MAAT test that developed by Abou-Elsaad et al. (2009) measured many consonants in the word-final position, using words with CCs in the word-final position. Thus, it could be assumed that this test has less test items that make their administration lasting for a shorter time comparing to SHAPA that has 151 test items.

In addition to being longer than tests reported in other studies, the SHAPA evaluation revealed that it is an extended test, which limits its applicability and appropriateness for children. It was noticed that children's performance was negatively affected, especially towards the end of the test, when the rate of spontaneous naming decreased (see below for more discussion). Its length caused many children to not name all test items, especially in the younger age group (2;6–2;11), as they lost motivation to continue participating in the test. It also noticed that the rate of spontaneouse naming decreased among children after they named one–hundred items.

As reported in Chapter 1, Section 1.2.1, the SHAPA word list was chosen considering as many as possible of the linguistic criteria. Therefore, to sample each UHA phoneme in each word position three times (i.e. ideally, a minimum of nine words for each phoneme), considering different word shape and consonant–vowel contexts, it was impossible to create a shorter word list for SHAPA, especially considering the high number of the UHA and MSA consonants (i.e. 26 and 28, respectively) comparing to other languages. To create a shorter word list, Flipsen and Ogiela (2015) reported an alternative approach which could ensure a shorted representative word list by considering the weight of a sound relative to their frequency of occurrence in the language. However, such information is not available for Saudi Hejazi Arabic. Further, as Flipsen and Ogiela (2015) explained, using frequency data that based on adult speech data may not be applicable for children.

Further, the SHAPA word list could be shortened by excluding some items, which need to be performed with caution. According to Eisenberg and Hitchcock (2010), in order to ensure that the time required to administer a test is practical, the number of items included in a test needs to be within the recommended number of a representative sample size (i.e. between 90 and 100 words) to ensure that each sound is tested in more than one word, considering the effect of surrounding phonemes which can influence a child's ability to produce the target sound. However, their review of 11 standardised tests revealed that four of these tests were unpopular due to their length (i.e. ranging from 81 to 141 total words), but the most popular test was GFTA–2, which has only 53 words. Although GFTA-2 is one of the tests showing the fewest number of phonetic control words (for word-initial and final positions), it was frequently used test by SLTs because it is relatively short (Eisenberg & Hitchcock, 2010). However, consonants in GFTA-2 have three opportunities to occur, one in each position, which is why it is not a phonetically controlled test. According to the recommendation of Eisenberg and Hitchcock (2010), a test should include at teat two opportunities for each consonant to be produced in two different words, considering two different word positions. A balance should be made between the practicality and comprehensives examination of the phonology (see Chapter 1, Section 1.4.1). Therefore, shortening the SHAPA test needs to be done carefully, considering the results of the item analysis conducted, as well as to be in line with the recommendation from the literature, which are discussed below.

4.4.2 Test item analysis

As previously mentioned, the creation of SHAPA was based on the international constructional principles for selecting lexical items which should be representative and familiar to the target population (James, 2001a; Marklund, Lacerda, Persson, & Lohmander, 2018; Stoel-Gammon & Williams, 2013). It is essential to ensure that the stimuli consistently elicit the target words to provide sufficient opportunities for children to produce all phonemes in every possible syllable and word position (Ceron et al., 2020). Hence, based on the results of the test items analysis, this section discusses three main findings regarding the appropriateness of selected items:

1) The rate of spontaneously named test items across all age groups

2) The rate of items named using whole imitation and other cues
3) The accuracy of naming the intended target
4) The accuracy of correctly pronouncing the target word.

Each of the following sections addressing the general performance of children on these measures, including comparisons with existing literature.

4.4.2.1 Rate of spontaneous naming

The analysis of the 151 SHAPA test items revealed that only 63.58% (n=96) of the SHAPA pictures were recognised and named spontaneously by more than 50% of children across all seven age groups. No one item named spontaneously by all children across all age groups; however, the spontaneous naming rate increased with age (see Appendix Q). This result aligns with that of Stoel-Gammon and Williams (2013), who found that the number of words named spontaneously increased with age, while the time required for test administration decreased, which is similar to the findings of this study. Specifically, Stoel-Gammon and Williams found that just under half of their words were produced spontaneously (40.39%) at the age of 17–19 months compared to 65% spontaneous production by the oldest age group (35–37 months; 2;9-3;1 year-old).

Although the average spontaneous naming rate for SHAPA was similar to that reported by Stoel-Gammon and Williams, it is still considered lower than that written in other literature, including data from older age groups (3;0–5;0) (e.g. Ceron et al., 2020; James, 2001a), and this brings into question the adequacy of the items and pictures selected for SHAPA. It also raised a question about the level of lexical development in SHA-speaking children, as there is growing evidence confirming that vocabulary size is related to the acquisition of phonological knowledge (Sosa & Stoel-Gammon, 2012). Although the majority of SHAPA words were chosen from JA-CDI, the age of acquisition of the SHAPA word list, or even the lexicon in the JA-CDI word list, is unknown due to the lack of such information in the Arabic literature, as well as in the JA-CDI's test manual. It was unknown if the words included in the JA-CDI word list were presented in ≥75% of the normative sample or ≥50% of children (cf. Marklund et al., 2018), nor their age of acquisition. Thus, the familiarity of the SHAPA words is still questionable. For these reasons, the SHAPA test needs further editing to improve its suitability for SHA-speaking children and increase its spontaneous naming rate, which is recommended for such a single-word naming task (see Chapter 1, Section 1.4.1).

Although the ideal number of opportunities for the production of each phone in a phonological test is recommended to be three occurrences of each phone in each possible position (Eisenberg & Hitchcock, 2010; James, 2001a), applying such a criterion to SHAPA resulted in a very long test that negatively affected the children's performance and led to a reduction in their spontaneous naming rate. Therefore, the new version of SHAPA should include the minimum number of opportunities for phone occurrence in a test, that is, two occurrences in each possible position in two different words for each position (Stokes et al., 2005).

Considering the spontaneous naming results, a future revision of SHAPA could involve reviewing each item score. Then, following the classification system suggested by Ceron et al. (2020) follows, but with some modifications to suit the target age of this study (see Chapter 1, Section 1.4.2.3):

a) Items identified and named spontaneously by more than 50% of children should retained in SHAPA (i.e. n=97 items) (see Appendix R, Table 1).
b) Items named spontaneously by fewer than 50% of children should be reviewed for their phonetic-phonological feature before deciding whether to remove them from the SHAPA word list or keep them and modifying either their pictures or elicitation instructions (i.e. n=54 items) (see Appendix R, Table 2).

In general, removing an item from SHAPA test needs to be done with careful considerations by finding a suitable replacement for the deleted words in a way not affecting the presentation of some phones in some positions.

Further look at the score of each item received in the spontaneous naming analysis revealed that 22 out of the 54 items that named spontaneously by fewer than 50% of children could be totally removed with no need to find a replacement item for them because their deletion should not affect the presentation of their phones in each position. On the other hand, nine out of the 56 items could be removed with a need to find an appropriate replacement to fulfil the linguistic criteria of SHAPA by providing at least two opportunities for each phone in each possible position (Stokes et al., 2005). For example, a phone such as /ʃ/ would require a replacement item to represent it in the word-final position because all items including this phone in the final position were scored as difficult (i.e. /tˤar.buːʃ/ /ʕeːʃ/ /ʕuʃ/). On the other hand, some items such as /ʕɑn.ka.ˈbuːt/ /mu'.kaiː.jif / should be retained in the test as their structure and phonetic features (e.g. trisyllabic words, the heterosyllabic CC /nk/, the diphthongs /aiː/) may not be found in other words familiar to children in the target age (see Appendix R, Table 2 for a full list of words that need to be deleted or amended their elicitation mood by adding clearer elicitation instructions).

In addition, two of the items that need to be removed are items with word-final CCs (i.e. /dˤirs/ 'molar tooth' and /ðˤarf/ 'envelope'). This word structure (word-final CC), in general, was not presented in the SHAPA word list in an adequate way, as CCs were not considered during the creation of SHAPA.

In addition to these two items that received a low spontaneous naming rate, another two words (i.e. /waʒh/ 'face', and /bint/ 'girl') were named frequently with a cluster reduction pattern (i.e. [waʃ] for /waʒh/ 'face' and [bit] for /bint/ 'girl') and these changes were scored as correct because they are examples of the acceptable dialectal changes. Thus, the frequency of word-final CCs reduced due to the word selection. For these reasons, the items representing word-final CCs required further editing to expand their representation by sampling more words with this word structure targeting a wider range of cluster elements after investigating the nature of word-final CCs development in SHA-speaking children (see Chapter 7, Section 7.1.1).

4.4.2.2 Rate of naming by imitation and other cues

To consider the participants' productions elicited with cues other than direct imitation, and to identify the difficult items that were only named by direct imitation by the majority of children, the researcher conducted two calculations: first, to count the number of imitated responses across all the children's age groups; second, to count the number of items named with other types of cues. The results showed that only 7.28% of the test items (n=11) were named with whole-word imitation by more than 50% of the children, with no item named by imitation by 100% of children across all age groups. These 11 words were almost all words that did no reported in the CDI word list, as 10 of them were not included in the Saudi version of the MC- CDI (JA-CDI; Dashash & Safi, 2014), and were chosen because of their phonetic context that could not be found in other items more familiar to children within the target age of this study.

It is evident that these 11 items were also part of the items named spontaneously by fewer than 50% of children across all age groups. For example, the word /hi.laːl/ 'crescent' received a low spontaneous naming rate across all age groups, and at the same time is one of the most frequently named items by imitation, as well as not being included in the CDI word list. Therefore, these items need to either be removed from SHAPA or changing their elicitation pictures or instructions need to be changed if their phonetic–phonological features are non–replaceable.

Further, the items named by children by using other types of prompts (i.e. descriptions, alternative questions, and first–sound imitation) were also calculated to identify items whose mode of elicitation needed to be changed to facilitate spontaneous naming. The result revealed that 12 items were named by a prompt other than whole–word imitation by more than 50% of children. These items also received low spontaneous naming scores as they are part of the items named spontaneously by fewer than 50% of the children across all age groups. The mode of elicitation of the items that named with a prompt by more than 50% of children should be changed to facilitate the spontaneous naming of these words. However, similar to the imitated items, these items should be revised carefully for their phonetic-phonological features, and they should be deleted if their features are replaceable or not needed after applying the new criterion of including only two occurrences of each phone in every possible position rather than three occurrences.

Although the SHAPA test targeted children between the ages of 2;6, and 5;11, the results obtained from SHA-speaking children revealed that the words which are not on the CDI list are the words with the highest difficulty measure, even for children older than 3;0 years, such as /sˤa.ruːx/ 'rocket', and /mas.baħ/ 'swimming pool'. At the same time, some other words which were selected from the CDI word list were still named by imitation be more than 50% of children, such as /ki.taːb/ 'book', and /laʔ/ 'no'. There may be several reasons for this. First, these words may have been still absent from the children's vocabulary as there is no research or database for the age of acquisition of words from JA-CDI. This missing data limited the word selection for SHAPA and, therefore, affected the quantity of items named spontaneously. Second, it is also possible that some of these words were difficult to illustrate using computerised cartoon drawings in a way that children could easily recognise. For example, one of these words is a verb (/ˈjimdˤuɣ/ 'chewing') whose picture often elicited the word [jaːkol] 'eating'. Another word is a noun (/hilaːl/ 'crescent'), and its picture elicited [gamar] 'moon'. A possible solution would be to use computerised animation to demonstrate these difficult–to–name verbs or adjectives, as recommended by McLeod (2012).

In general, comparing the rate of spontaneous naming to the rate of whole–word imitation naming, it could be noticed that the imitation rate was too low (7.28%, n=11). According to Stoel-Gammon and Williams (2013), this comparison could indicate that even though not all children have all test items in their productive lexicon, they are likely part of their receptive vocabulary. They were elicited using other types of cues (i.e. description, alternative questions, and first–sound imitation).

4.4.2.3 Accuracy of naming the intended target

In order to ensure that the selected pictures for SHAPA elicited the intended target words, the researcher counted the number of the words named as an intended target across all age groups. The result revealed that 90.73% (n=137) of SHAPA items resulted in the intended target being named by more than 50% of the participants, with only one item named as the target by all children across all age groups (i.e. /ħalaːwa/ 'candy'). Nojavan-Pirehyousefan et al. (2021) reported that a higher than 70% naming agreement with the intended target that is essential for the conceptualisation step of designing a speech sounds assessment tool, in particular, the purpose and scope of the test, which is related to the selection of the test items (McLeod & Verdon, 2014). For SHAPA, only 14 items did not result in the intended target being named by 50% of the participants; however, these items affected the phonetic features that needed to be elicited. For example, 67 children across all age groups named the target /moːz/ as [moːza], and this affected the production of /z/ in the word-final position. Therefore, this item needs to be removed because the intended phonetic feature was not elicited by this item. In general, six of these 14 items should be deleted according to the scores of the test item's difficulty measure (spontaneous naming) if their phonetic–phonological features could still be presented if they were deleted. Otherwise, the images of these items, or their elicitation instructions, need to be changed to improve the picture–naming agreement score for SHAPA.

A closer examination of the items named differently from the intended target showed that the item that received the highest percentage (84.26%) of naming disagreement is /beːt/ ‘house’, which is one of the words named spontaneously by the majority of children (i.e. by 87.52%). Most children tend to add the definite article /ʔal/ at the beginning of the word /ʔal beːt/ ‘the house’, which changes the phonetic context of the word by adding a syllable and changing the position of /b/ from word-initial to be as syllable–initial within the word. However, this word may be retained in the SHAPA word list because of the phone /t/ in the final position as there are only two words of /t/ as a final consonant, one of these other two should be removed due to its high difficulty score. Similarly, the word /ʒauːˈwaːl/ was named by adding the definite article by 60.43% of children on average; this affects the phonetic context of the word, which was chosen to test /ʒ/ in the word-initial position. Still, this word should retain in SHAPA because of the diphthong /auː/.

The second highest percentage (80.43%) for naming disagreement was for the word /dˤirs/ ‘molar tooth’, which is one of the words with a low spontaneous naming rate; this was changed to /sin/ ‘tooth’ or /ʔasnaːn/ ‘teeth’, which also changes the phonetic target of the intended item (e.g. /dˤ/ in the wordinitial position). The word /ˈjimdˤuɣ/ ‘chewing’, which is one of the items with a high imitated naming rate, was also changed by 65.96% of children to be named as /jakul/ ‘eating’, and this change affects the chosen phonetic context, which includes testing the emphatic /dˤ/ in the syllable–initial within–word position, and the velar /ɣ/ in the word-final position.

The third highest percentage (69.79%) for naming disagreement was for the item /fus.taːn/ ‘a dress’, which was changed in its pronunciation: the majority of children named it as /fis.taːn/ which is an acceptable dialectal change. However, such a change affects the consonant–vowel context of the word that was chosen to test the /f/+/u/ context. For this reason, all words that were chosen for a feature that could be lost if an acceptable dialectal change occurs should be deleted if they did not represent other phonological features.

4.4.2.4 Accuracy of articulation

The result for the accuracy of articulation revealed that only 55 items were named with inaccurate articulation by fewer than 50% of children, while 95 items were named with correct articulation without direct imitation by more than 50% of children. This finding indicates that the SHAPA words showed an adequate level of difficulty to measure articulation, as no item was named with incorrect articulation, or with accurate articulation by all children across all age groups in this study sample. This means that no items should be removed from SHAPA because of the result of this measure.

The articulation accuracy assessment result revealed that SHAPA has an adequate content validity measure. This finding is consistent with the findings of Ceron et al. (2020), James (2001), and Nojavan-Pirehyousefan et al. (2021), who investigated the content validity of their tests and analysed the correct production of their test items. They used the recognition rate (picture-naming agreement) and the target words' production accuracy to confirm their tests' content validity (articulation accuracy). James (2001) and Nojavan-Pirehyousefan et al. (2021) directly compared the accuracy production of the normal (typically developing) group to the cohort with speech impairment. Their item discrimination differential scores fall within the required range (at least 15%). In contrast, the articulation accuracy measure for SHAPA was generated for all children in the study with no specific comparison between the results of children with superior phonological performance (e.g., low InfrVar mean score) and children with low performance (e.g. who had high InfVar mean score). Therefore, the diagnostic validity of SHAPA could not be confirmed. The preliminary consistency between the results of this investigation and the other studies indicates a widespread validation of the SHAPA test. Further steps must be taken to improve the validity of SHAPA (see Chapter 7, Section 7.3).

In short, the item analysis results showed that the items selected for SHAPA were, on average, appropriate to elicit the target words using visual and verbal cues. This could indicate that the content validity of SHAPA was partly confirmed. Only the spontaneous naming was questionable because only 64.24% of items were named spontaneously by more than 50% of children. This percentage is lower than the cut-off criterion followed in this study (i.e. 70% of items were named spontaneously by 50% of children, James, 2001a). Considering the results of other measures, the item difficulty of SHAPA was acceptable for this study, thus providing some content validity evidence for this tool. However, further editing of its word list may increase SHAPA's validity evidence. This editing should include identifying items with low spontaneous naming, high imitation naming, or low picture–naming agreement scores. Their phonetic–phonological context can then be reviewed to decide whether or not their removal will affect the linguistic criteria of SHAPA. Finally, a replacement item should be found wherever it is needed. Otherwise, the item should be deleted, as one of the main reasons for editing SHAPA is to reduce its length to improve its suitability for children.

4.4.3 Reliability

The results of the two measures of inter- and intra-rater agreement showed that SHAPA has satisfactory inter- and intra-rater reliability (see Table 4.2). A 90% or higher of point–to–point agreement for both tasks (IPA phonetic transcriptions and identifying phonological patterns) was achieved between the scoring of the researcher and the qualified SLT rater for the inter–rater agreement, and for the researcher with herself when she reanalysed the data for intra-rater agreement.

The results of the inter- and intra-rater reliability for SHAPA were in line with literature recommendations, where the point–to–point agreement for both measures showed a score higher than 90% agreement for both tasks (i.e. IPA phonetic transcription and phonological patterns identified). Kirk and Vigeland (2014) reported that the required score for the point–to–point percentage of agreement is at least 90% and that inter-rater reliability information should be reported for all the age groups of the normative sample. The inter- and intra-rater reliability percentage agreements reported for SHAPA were in line with Kirk and Vigeland's recommendations.

As reported in Chapter 1, Section 1.4.2.2, it is recommended that multiple measures are needed to report the reliability. In this study, two measures of reliability were calculated: inter- and intra-rater agreement, by having one qualified SLTs reanalyse the IPA phonetic transcription and phonological pattern data. Although McCauley and Swisher (1984) reported that the evidence of inter–rater reliability is met if a correlation coefficient of .90 or better was reported, many researchers have reported inter- and intra-rater reliability based on a point–to–point agreement index (e.g. McLeod & Verdon, 2014; Prezas et al., 2014; van Haaften et al., 2020).

4.4.4 Summary of discussion I

In general, the evaluation of the test administration, item difficulty and reliability of SHAPA revealed some strengths of incorporating some of the construct criteria, including the following:

- The test's normative sample was large (n=235) and was described in enough detail, including information about gender, geographic information, and SES.
- It guaranteed sufficient opportunities for all UHA and MSA consonants and vowels to occur in different word/syllable positions (i.e. SIWI, SIWW, SFWW, SFWF), which ensured sufficient opportunities for each phonological pattern tested in this study to occur.
- It included words with different word shapes and lengths (e.g. multisyllabic words) and phonotactic structures (e.g. word-final CCs, heterosyllabic CCs, and gemination). Although their occurrence may not be representative; their analysis provided preliminary data for acquiring such features.

- Item–analysis procedure revealed a decent score of spontaneous naming rate among SHA-speaking children, where 64.24% of items were named spontaneously by more than 50% of children.
- It included items that are culturally appropriate for the targeted children, as well as words familiar to the age range understudy, where the majority of the test items were selected from JA-CDI. The picture–naming agreement analysis has proven such familiarity, where 90% of SHAPA items were named as the intended target by more than 50% of children.
- The items selected for SHAPA were appropriate to test the phonological development among targeted children where no item was very easy to be named correctly by all children across the age group; on the other hand, no item was very difficult to be named correctly by all children across age groups. Such a feature was evident by analysing the articulation accuracy, which revealed that 61.68% of the items were named with correct articulation without direct imitation by more than 50%of children.
- The inter– and intra–rater reliability showed a high score of the point–to–point agreement for both tasks of SHAPA analysis: IPA phonetic transcription and identifying phonological patterns. This indicates that reliable results could be expected over multiple performances of SHAPA.

Despite these strengths, this study showed some limitations in SHAPA that should be considered before any generalisation. First, the length of SHAPA had negatively affected the children's performance, particularly the rate of spontaneous naming, which showed a reduction toward the end of the word list even for children in the oldest age group. Second, the low score of the average spontaneous naming rate across all age groups indicated that the item difficulty score was high, which also affected the applicability of the test. Third, the items that represent the consonant co–occurrence feature, the word-final CC and gemination, in particular, need to be evaluated and may be expanded to include a more representative sample.

In general, considering the construction criteria in creating a phonological assessment tool for SHA-speaking children could not lead to creating a clinically appropriate assessment tool. Besides, a test maker must have a clear explanation of the test purpose and its targeted population, analyse the appropriateness of the test items, and define the type of words included as well as the procedures used to choose the target words (Ceron et al., 2020). SHAPA was designed with these characteristics in consideration; nonetheless, the spontaneous naming rate was lower than what was stated in the literature, suggesting that the target words were not in the children's lexicon. It is well known that lexical development is linked to the acquisition of the phonological inventory (Sosa & Stoel-Gammon, 2012). Therefore, SHAPA needs further editing to improve its theoretical and practical features in terms of test duration and item number, as well as items difficulty score, which needs to be reduced to improve the spontaneous naming rate (see Chapter 7, Section 7.3).

5 Results and Discussion II – Phonetic acquisition in Saudi Hejazi Arabic-speaking children

In this chapter, the results of the single-word naming task and stimulability task using SHAPA will be presented. The chapter is divided into three subsections. First, the findings of the quantitative measures obtained via the single-word naming task are reported. Second, the findings from both tasks, the single-word naming task and stimulability task, will be presented to describe the phonetic inventory of SHA-speaking children. Third, the results of the single-word naming task will be further analysed to calculate the percentage of consonantal co-occurrences correct, with 75% and 90% accuracy criteria, including word-final consonant clusters (CCs), heterosyllabic CCs, and gemination. All results are presented as descriptive data only, with no statistical analyses. These results will then be discussed to address the second research question and its three sub–questions outlined in Chapter 2, Section 2.5, reproduced as follow:

> RQ2: How does the phonetic acquisition in monolingual SHA-speaking children proceed between the age of 2;6 and 5;11 in terms of the following:
>
> > *2a) At what age do Saudi Hejazi Arabic-speaking children acquire and master the Arabic (UHA and MSA) phonetic system taking into account different measures such as PCC, PCC-R, PVC, PCCC (three different types of consonant co–occurrences), and age of acquisition of C and CC structures?*
> > *2b) At what age are the consonants of UHA and MSA stimulable when elicited via a single word naming task or a stimulability task?*
> > *2c) Does the phonetic inventory elicited via a single-word naming task versus a stimulability task differ?*

5.1 Results II– Phonetic acquisition

5.1.1 Percentage of consonants, vowels and correct: PCC, PCC-R, PVC

To address research question 2a, children's PCC, PCC-R, and PVC scores were calculated from the single-word naming data using PHON. As reported in Chapter 3, Section 3.6.2.1, the analysis of PCC, PCC-R, and PVC was conducted following the rules in Shriberg et al. (1997a). Tables 5.1, 5.2, and 5.3 display the average percentage consonants and vowels correct per age group, as well as data on the standard deviations and range (min–max).

Table 5.1 *Average scores, SD, and range of PCC and PCC-R across the age groups*

Age groups	N (children)	PCC		PCC-R	
		M (SD)	Range	M (SD)	Range
2;6–2;11	30	55.62 (12.76)	31.37–83.00	61.61 (13.40)	38.21–84.80
3;0–3;5	27	67.83 (15.55)	40.21–92.96	71.53 (15.34)	41.06–94.16
3;6–3;11	38	74.28 (12.51)	42.13–94.04	78.70 (11.66)	46.06–95.63
4;0–4;5	39	81.95 (9.21)	61.65–94.31	85.51 (7.64)	65.46–96.42
4;6–4;11	31	86.64 (7.08)	65.07–95.04	89.22 (6.35)	67.47–95.81
5;0–5;5	37	89.80 (8.55)	60.61–98.58	92.36 (7.12)	66.26–98.78
5;6–5;11	33	89.64 (6.67)	69.09–98.58	91.56 (5.87)	71.92–99.19

Key: N= the number of children in an age group, M= the mean (average), SD= standard deviation, PCC= percentage of consonants correct, PCC-R=PCC-Revised.

Table 5.1 shows an apparent increase in the average percentage of consonants correct across all age groups for both measures (PCC and PCC-R) across the age groups. It also shows a high standard deviation and range across all age groups, which both showing a steady reduction trend across all the age groups for PCC and PCC-R measures. It is clear that there is a slight reduction in the PCC and PCC-R mean scores for the 5;6–5;11 age group. In addition, the average scores of PCC and PCC-R appeared to be stable after age 4;0.

Comparing the PCC and PCC-R, the PCC-R measure resulted in a slightly higher percentage of consonants correct across all age groups. After considering all the distorted productions as correct (for PCC-R), some children achieved a mean score of >95% consonants correct, and the mean average of the two oldest age groups (i.e. 5;0–5;5 and 5;6–5;11) approached scores above 90%, with some children achieving a score of 99% (see the maximum range of these two groups in Table 5.1). The gap between the mean scores for both measures reduced with increasing age; the differences between the two calculations almost completely diminished after the age of 5;0.

Table 5.2 presents the average scores of the percentage of vowels correct (PVC). The result showed that the children already achieved a high PVC score (>80%) at an early age (i.e. 2;6–2;11), indicating that their vocalic acquisition was completed before the age of 3;0.Similar to PCC, there is a gradual increase in PVC mean scores across all age groups, with stagnation and the score approaching a ceiling level from the age of 4;6–4;11 (see Table 5.2).

Table 5.2 *Average scores, SD and range of PVC across the age groups*

Age groups	N (children)	PVC	
		M(SD)	Range
2;6–2;11	30	80.11 (9.17)	52.84–94.74
3;0–3;5	27	85.70 (7.41)	67.97–95.17
3;6–3;11	38	91.30 (4.01)	79.23–96.94
4;0–4;5	39	93.75 (3.69)	81.13–100.00
4;6–4;11	31	95.66 (2.14)	88.99–98.46
5;0–5;5	37	95.92 (2.32)	90.71–99.06
5;6–5;11	33	96.48 (2.12)	90.34–100.00

Table 5.2 shows that there are large ranges of the PVC scores across all age groups, but with low SDs that decreased with increasing age. It also appeared that SHA-speaking children reached the competence level of PVC by age 3;6 where the score become higher than 90%. Further, some children in some age groups (i.e. 4;0–4;5 4;6–4;11, 5;0–5;5and 5;6–5;11) reached, or at least approached, a PVC score of 100%.

5.1.2 Phonetic inventories

To address research question 2b, phonetic inventories were compiled and analysed. The data from the single-word naming task and stimulability task were analysed based on the criteria described in the methods chapter (see Chapter 3, Section 3.6.2.2 for the scoring and analysis criteria used in these two tasks).

5.1.2.1 Phonetic inventory via the single-word naming task

Table 5.3 presents the single-word naming task results for the age at which a phone in UHA and MSA can be considered acquired (75% correct productions) or mastered (90% correct productions). Two analyses were implemented: counting all the distorted productions of specific phones as errors and accepting all distorted pronunciations as correct.

Table 5.3 Arabic phonetic acquisition and mastery according to the 75% and 90% accuracy criteria from the single-word naming task

Age group	**Acquisition (75%)**			**Mastery (90%)**		
	Consonants		**Vowels**	**Consonants**		**Vowels**
	With distortions counted as errors	With distortions counted as correct		With distortions counted as errors	With distortions counted as correct	
2;6–2;11	/k/ /ðˤ/ /z, ʃ/ /ɾ, r/	/k/ /ðˤ/ /ʃ, ʒ*/ /ɾ, r/	/ɑː/	/b, t, d, **g**, ʔ/ /m ,n/ /f, **ð**, s, ħ, h/ /l ,w ,j/ /ɾ/	/b, t, d, **g**, ʔ/ /m ,n/ /f, **ð**, s, z*, ħ, h/ /l, w, j/ /ɾ, r*/	/i, iː, u, uː, eː, oː, a, aː, ɑ/
3;0–3;5	/tˤ, dˤ/ / sˤ/	/tˤ/, dˤ/ /sˤ/		/k, ʕ/ **/ðˤ/**	/k, ʕ/ **/ðˤ**, sˤ*/	/a͡uː/
3;6–3;11	/x, ɣ/	/x, ɣ/	/a͡iː/	/tˤ/	/tˤ/	
4;0–4;5	**/q/** **/θ,** ʒ/	**/q/** **/θ/**		/dˤ/ /z, sˤ, x, ɣ/ /r/	/dˤ/ /x, ɣ/	/ɑː/
4;6–4;11				/ʃ/	/ʃ/	/a͡iː/
5;0–5;5				/ʒ/	/ʒ/	
5;6–5;11				**/q/**	**/q/**	

Key: Shared phones, UHA only, MSA only

* Sounds that showed an earlier age of acquisition or mastery when the distorted productions were counted as correct.

The first calculation of the phonetic inventory was conducting using PHON with all phonetic distortions were counted as errors (see the grey columns on Table 5.3). The results revealed mastery (with 90% of accuracy) of almost 16 Arabic phones (UHA and MSA consonants and vowels) by the youngest age group (2;6–2;11). These sounds are: /b, t, d, g, ʔ, f, **ð,** s, ħ, h, m, n, l, w, j, ɾ /, including:

1) The 13 shared consonantal phones (in black font), that are: /b, t, d, ʔ, f, s, ħ, h, m, n, l, w, j/
2) the one and only UHA phone (in green) /g/.
3) Two MSA phones (in red font): / **ð,** ɾ /.

In terms of the shared UHA and MSA phones, 15 consonantal phones were already mastered by children aged 2;6–2;11, including a number of shared sounds: /b, t, d, ʔ, f, s, ħ, h, m, n, l, w, j/. It is important to note that the acquisition of these shared consonants should occur before age 2;6, the youngest age in the current study. This is the reason for not presenting an acquisition age for these consonants in Table 5.5.

The results show that SHA-speaking children achieved an acquisition level of accuracy for all shared consonants by age 3;11, and a mastery level by age 4;5, except for undistorted postalveolar /ʃ/ and /ʒ/. They reached the mastery level of accuracy for these two phones /ʃ/ and /ʒ/ by ages 4;11 and 5;5, respectively, which were the latest shared phones to be mastered. (see Appendix V, Table 1 for detailed information on the acquisition of shared UHA and MSA consonants across the age groups).

Table 5.3 shows that the children reached the mastery level with ≥90% accuracy for the production of most monophthongs production by an early age (i.e. 2;6–2;11), except for the vowels: /ɑː/, which the children mastered after the age of 4;0. In contrast, the two diphthongs /a͡iː/ and /a͡uː/ had a late age of acquisition and mastery (i.e. 3;6 for the acquisition of /a͡iː/ and 4;6 for its mastery, and 3;0 for the mastery of /a͡uː/) (see Appendix V, Table 2 for detailed information on the acquisition of shared UHA and MSA vowels across the age groups).

For the only UHA phone /g/, the results showed an early age of mastery, where the children mastered the production of this velar plosive by the age of 2;11.

The data show slight variations in the acquisition and mastery of MSA phones among the children. They acquired and mastered the two dental fricatives /ð/ and /ðˤ/ before age 3;5, while they acquired /q/ and /θ/ by age 4;0–4;5, and mastered /q/ by the age of 5;11. However, they never reached the mastery level for the dental fricative /θ/ even by the oldest age group (5;6–5;11) (see Table 5.5). Appendix V, Table 1, presents the age of mastery and acquisitions of the shared, UHA, and MSA consonantal phones across all age groups resulting from the first analysis.

Since some shared UHA and MSA fricatives (i.e., /s, sˤ, z, ʃ, ʒ/), the lateral approximant /l/, and the alveolar trill /r/ have been reported to be highly prone to phonetic distortions, a second analysis was conducted using PHON (see Chapter 3, Section 3.6). On this time, specific distortions were accepted as correct. The results shows that the children achieved an earlier age of acquisition and mastery when the distorted productions were counted as correct. Specifically, they had an earlier age of mastery for /r/, /z/ and /sˤ/, when their distorted productions were counted as correct (i.e. 2;6–2;11 and 3;0–3;5) compared to their age of mastery when the distortions were counted as errors (i.e. 4;0–4;5). In contrast, the age of acquisition and mastery of the alveo–lateral /l/ and the two postalveolar fricatives /ʃ/ and /ʒ/ did not change even after counting distorted productions as correct. In both calculations, the alveo–lateral /l/ was mastered by children aged 2;6–2;11, while the two postalveolar fricatives /ʃ/ and /ʒ/ achieved the mastery level by age 5;0–5;5. Appendix V, Table 3 shows results from the second analysis on the age of mastery and acquisitions of the shared distorted phones across all age.

Overall, the single-word naming task results show that SHA-speaking children completed the phonetic inventory of all Arabic consonants and vowels by age 5;11; this involved considering all distorted productions as errors or correct productions except for /θ/.

5.1.2.2 Phonetic inventory via the stimulability/ phone imitation task

In terms of the age of mastery and acquisition, Table 5.4 summarises the phonetic inventory for each age group using data from the stimulability task, considering both analyses (with distortions as errors and with distortions as correct). The exact age of acquisition and mastery of each phone (consonants and vowels) across the age groups are presented in Appendix W, Tables 1 and 2.

Table 5.4 *Phonetic acquisition and mastery according to the 75% and 90%accuracy criteria from the stimulability task*

Age groups	Acquisition (75%)			Mastery (90%)		
	Consonants		Vowels	Consonants		Vowels
	Without distortions counted as errors	With distortions counted as correct		Without distortions counted as errors	With distortions counted as correct	
2;6–2;11	/k, ʔ/	/k, ʔ/		/b, t/ /m, n/ /f, h/ /l,w, j/	/b, t/ /m, n/ /f, s*, h/ /l, w, j/	/iː, uː, aː/
3;0–3;5	/d, tˤ, **g/** /x, ʕ/ **/ɾ/**	/d, tˤ, **g/** /sˤ* z,* x, ʕ/ /**ɾ**/		/k, ʔ/ /ħ/	/k, ʔ/ /ħ/	
3;6–3;11	/dˤ/ /s, sˤ, z, ʃ, ɣ/	/dˤ/ /ʃ, ɣ/		/d/	/d/	
4;0–4;5	**/q/**	**/q/**		/tˤ, **g/** /ɣ/	/tˤ, **g/** /ɣ/	
4;6–4;11		/ʒ*/ /r̝ /		/x, ʕ/	/ʃ* x, ʕ/ /**ɾ**/	
5;0–5;5	**/θ, ð, ðˤ,** ʒ/	**/θ, ð, ðˤ/**		/dˤ/	/dˤ/ /z*/	
5;6–5;11	/r/			**/q/** **/θ, ð, ðˤ/** /**ɾ**/	**/q/** **/θ, ð, ðˤ,** sˤ*/ /r̝/	

Key: Shared phones, UHA only, MSA only

* Sounds that showed an earlier age of acquisition or mastery when the distorted productions were counted as correct.

In general, Table 5.4 shows that children in the youngest age group (2;6–2;11) were able to mastered the production of the following nine shared consonant phones /b, t, m, n, f, h, l, w, j/ with 90% of accuracy via imitation, which they should have acquire before age 2;6.

All of the shared consonants were acquired by children aged 3;11 or younger, except for the undistorted production of the trill /r/ and postalveolar /ʒ/. For these two phones /r/ and /ʒ/, the children acquired their production by age 5;5 and 5;11, respectively. In contrast, the mastery level of accuracy for the production of all shared consonants was never reached in this stimulability task when distorted productions were counted as errors, even for children in the oldest age group. Five fricatives of the shared consonants /s/, /sˤ/, /z/, /ʃ/, and /ʒ/, as well as the trill /r/, were found to be difficult to imitate in isolation by SHA-speaking children at a mastery level even by children in the oldest age group (see Appendix W, Table 1).

For the shared vowels, the children showed a mastery level for the long monophthong vowels /iː, uː, aː/ by age 2;6–2;11, which were the only vowels presented in this task (see Chapter 3, Section 3.3.2). Therefore, the result of vocalic phones' imitation is incomplete in this task (see Appendix W, Table 2).

For the UHA phone /g/, the children imitated it at the level of acquisition by age 3;5 and at the level of mastery by age 4;5.

The results show that the children in this study acquired the production of all MSA phones between the ages of 4;0 and 5;5, and mastered them by age 5;11. The earliest MSA phone to be acquired was the uvular /q/ (by age 4;5), and the final phones to be acquired were the three dental fricatives: /θ/, /ð/, and /ðˤ/ (by age 5;5). In terms of the mastery level of production, all the MSA consonants were mastered by the ages of 5;6–5;11. In general, the acquisition and mastery of the MSA phones via the stimulability task occurred at a later age than that of the shared phones (except /r/).

SHA-speaking children showed a higher level of acquisition and mastery for phones that are prone to distortion (i.e. the fricatives /s, sˤ, z, ʃ/, the alveolar trill /r/, and the alveolar lateral /l/) when all their distorted productions counted as correct (see Appendix W, Table 3). Interestingly, when the children's productions of the distorted /s/ were counted as correct, they reached the mastery level at an earlier stage by age 2;6–2;11 compared to when distortions were counted as errors, in which they never reached the mastery level even for the oldest age group. In general, the children's output for all of these phones showed a mastery level before age 5;11 when their distortions were counted as correct (see Table 5.4), except for the postalveolar /ʒ/, the oldest age group (5;6–5;11) did not reach the mastery level for this phone even after the distorted productions were considered correct.

Overall, SHA-speaking children completed the phonetic inventory of all Arabic consonants and vowels via the stimulability task by age 5;6–5;11. Counting the distortion outputs as correct, the children mastered the production of all Arabic phones except the postalveolar fricative /ʒ/. In contrast, the phonetic inventory via stimulability task was not completed when all distorted productions of phones were counted as errors.

5.1.2.3 Comparison of the phonetic inventory results

To answer research question 2b, the researcher compared the phonetic inventory results for both tasks: the single-word naming task and stimulability task (considering all distorted phones as correct). In general, it is evident that the single-word naming task showed an earlier age of acquisition and mastery (with 75% and 90% accuracy criterion, respectively) for shared, UHA and MSA phones than that found via the stimulability task. Table 5.5 shows the acquired and mastered consonantal phones for each age group in each task, with considering all distorted productions of phones prone to distortion across the age groups. It is important to note that vocalic acquisition is not included in this comparison due to the incomplete vowel inventory imitated during the stimulability task, which included only the long monophthong vowels /iː, uː, aː/, and revealed an early age of mastery (2;6–2;11) in both tasks.

Table 5.5 *Phonetic inventory of SHA-speaking children per age group for both tasks (single-word naming and stimulability task), considering distorted productions as correct*

Age groups	**Stimulability task**		**Single-word naming task**	
	Acquired ≥75%	**Mastered ≥90%**	**Acquired ≥75%**	**Mastered ≥90%**
2;6–2;11	/k, ʔ/	/b, t/ /m, n/ /f, s, h/ /l, w, j/	/k/ / **ðˤ**, z, ʃ/ /**ɾ**, r/	/b, t, **d, g, ʔ**/ /m, n/ /f, **ð,** s, **z, ħ,** h/ /l, w, j/ **/ɾ, r/**
3;0–3;5	/d, tˤ, **g**/ /sˤ, z, x, ʕ/ **/ɾ/**	/k, ʔ/ /ħ/	/dˤ, tˤ/	/k/ **/ðˤ, sˤ, ʕ/**
3;6–3;11	/dˤ/ /ʃ, ɣ/	/d/	/ʒ, x, ɣ/	**/tˤ /**
4;0–4;5	**/q/**	/tˤ, **g**/ /ɣ/	**/q/** **/θ/**	**/dˤ/** **/x,** ɣ/
4;6–4;11	/ʒ / /r/	/ʃ, x, ʕ/ **/ɾ/**		/ʃ/
5;0–5;5	**/θ, ð, ðˤ/**	/dˤ/ /z /		/ʒ /
5;6–5;11		**/q/** **/θ, ð, ðˤ,** sˤ/ /r/		**/q/**

Key: Shared phones, UHA only, MSA only

Phones in blue indicates an earlier age of mastery (≥90%) than in the other task.

Comparing the phonetic inventories elicited by both tasks after counting distorted productions as correct, it is evident that there is a similar trend in the age of acquisition and mastery of shared consonants, with a few differences. Postalveolar fricative /ʒ/is the only shared consonant that children did not master via the stimulability task, while children aged 5;5 reached the mastery level for this consonant via the single-word naming task. Further, the shared trill /r/ and tap /ɾ/ showed a 100% mastery level as early as 2;11 in the single-word naming task, while the stimulability task results revealed a later age of mastery for the tap /ɾ/ by age 4;5 and for the trill /r/ by age 5;11. Interestingly, all the shared emphatic consonants showed an earlier age of mastery via the single-word naming task (i.e. /tˤ/, /dˤ/, and /sˤ/) than that shown by the stimulability task (see Table 5.5).

Further, the children showed an earlier age of mastery for the only UHA phone /g/ (i.e. 2;6–2;11) through the single-word naming task compared to the stimulability task; in the latter, they did not master its production until age 4;5.

Interestingly, the children showed an earlier age of mastery for the MSA phones via the single-word naming task than the stimulability task (see Table 5.5). However, they only reached the mastery level for imitating the MSA interdental fricative /θ/ by age 5;6–5;11 (the oldest age group) in the stimulability task and did not master this phone in the single-word naming task result. Similar to the shared emphatic consonants, the MSA emphatic /ðˤ/ showed an earlier age of mastery via the single-word naming task than the stimulability task.

5.1.3 Consonantal co-occurrences

To answer research question 2c, further analysis was conducted on the acquisition of consonantal co–occurrence either at the end of the word position (i.e. word-final CCs) or in the middle of the word (i.e. heterosyllabic CCs and gemination).

5.1.3.1 Word-final consonant clusters (CCs)

Chapter 3 (see Section 3.3.1.3.2. under consonant clusters) reported that the single-word naming task included only nine words with word-final CCs. Each of these clusters occurred only once. Therefore, a preliminary analysis was conducted on this syllable structure type in UHA. PHON was used to generate the children's responses for this specific feature. Then, two calculations were done. First, the average percentage of correct productions of all nine word-final CCs (i.e. PCCC) was calculated for each age group (see Table 5.6). Second, the percentage of the correct productions of each cluster was calculated for each age group to identify the age of acquisition and mastery of each cluster (see Appendix X, Table 1). To be included in the calculation, both elements of the word-final CCs needed to be produced phonetically correctly (see Chapter 3, Section 3.6.2.3 for the criteria used to calculate the accurate productions of word-final CC). Table 5.6 displays the average percentage scores for the acquisition of correctly articulated word-final CCs for each age group, as well as the standard deviations and range.

Table 5.6 *Average scores of percentages of word-final consonant–clusters correct (PCCC) across the age groups*

Age groups	N (children)	Mean (SD)	Range	
			Min	Max
2;6–2;11	30	31.85 (22.17)	0.00	77.78
3;0–3;5	27	43.62 (24.25)	0.00	77.78
3;6–3;11	38	55.85 (20.74)	11.11	88.89
4;0–4;5	39	69.23 (17.38)	33.33	100.00
4;6–4;11	31	75.99 (19.27)	22.22	100.00
5;0–5;5	37	80.18 (15.52)	44.44	100.00
5;6–5;11	33	81.48 (16.82)	44.44	100.00

The most prominent observation from Table 5.6 is the steady increase in the PCCC with increasing age, but with a large range across all age groups. It is clear that the youngest age group (2;6–2;11) showed a low PCCC score for word-final CCs, with some children never correctly producing words with word-final CCs. From the age of 4;0–4;5, some children reached 100% PCCC for word-final CCs, although the average PCCC score was still below 85% even for the oldest age groups. It could be concluded that by the age of 4;6–4;11, SHA-speaking children achieved an average PCCC score of 75% for word-final CCs.

In terms of the word-final CC inventory, the children in the youngest age group (2;6–2;11) appeared to be unable to produce a final CC syllable structure, even those with elements acquired as singletons. The majority of the targeted word-final CCs (i.e. 7 out of 9) were acquired by age 5;5. Two unacquired clusters are those containing the postalveolar fricative /ʒ/ as one of its elements (/ʒh/ and /rʒ/). For the mastery criterion, the earliest cluster to be mastered was /lb/ (i.e. 4;0–4;5), followed by /nt/, /ms/ and /rd/ which were mastered by before age 5;11 (see Table 5.7).

To summarise the previous result, Table 5.7 below presents the age of acquisition (with 75% accuracy criterion) and mastery (with 90% accuracy criterion) for each cluster.

Table 5.7 *Age of acquisition and mastery of the UHA word– final CCs in the single-word naming task*

Age groups	Acquired (75% criterion)	Mastered (90% criterion)
2;6–2;11	—	—
3;0–3;5	–lb	—
3;6–3;11	–ms	—
4;0–4;5	–nt/ –lħ	–lb
4;6–4;11	–rd	—
5;0–5;5	–rf/ –rs	–nt
5;6–5;11	—	–ms/ –rd

It is evident that SHA-speaking children did not master the production of the word-final CCs /lħ/, /rs/, /rf/, /rʒ/, and /ʒh/ by age 5;11. They neither acquired, nor mastered two of the word-final CCs in SHAPA /rʒ/ and /ʒh/.

5.1.3.2 Heterosyllabic consonant clusters (CCs)

The single-word naming task included 43 heterosyllabic CCs, consisting of the consonants on syllables boundaries (e.g. the coda of the first syllable and the onset of the second syllable). Five of these 43 clusters were presented in the single-word naming task with more than one occurrence (two to three occurrences).

To better understand how SHA-speaking children acquired this type of CCs, a preliminary analysis of all heterosyllabic CCs in the single-word naming task was conducted. PHON was used to generate children's responses that included heterosyllabic CCs. Similar to the calculations done for word-final CCs, the heterosyllabic CCs were also counted in two steps. First, the average percentage correct scores were calculated for the productions of all heterosyllabic CCs across all age groups, including the distorted pronunciations of any element of the clusters (see Table 5.8). Second, the percentage of the correct productions of each cluster was calculated for each age group to identify the age of acquisition and mastery of each cluster (see Appendix X, Table 2). Table 5.8 represents the average percentage correct scores of the heterosyllabic CCs for each age group, together with the standard deviations and range.

Table 5.8 *Average percentage correct scores of UHA heterosyllabic CCs (PCCC) across the age groups*

Age groups	N (children)	Mean (SD)	Range	
			Min	Max
2;6–2;11	30	28.78 (21.49)	2.04	79.59
3;0–3;5	27	39.98 (22.14)	4.08	75.51
3;6–3;11	38	54.89 (19.59)	8.16	85.71
4;0–4;5	39	69.02 (15.02)	26.53	91.84
4;6–4;11	31	73.73 (10.67)	40.82	89.80
5;0–5;5	37	81.08 (12.35)	40.82	95.92
5;6–5;11	33	79.22 (9.74)	51.02	95.92

As expected, the average number of correctly articulated heterosyllabic CCs is increased with ascending age. Table 5.8 shows that there is a high variation in children's performance in each age group, as the ranges for each age group are large. For the youngest age group (2;6–2;11), the average PCCC score for heterosyllabic CCs was low, with an extensive range; however, some children approached a PCCC score of 80% for heterosyllabic CCs. For the two oldest age groups, the average PCCC scores were approached 80%, with some children in these two groups reaching a PCCC score of 95% for heterosyllabic CCs. In general, the production of heterosyllabic CCs, as a linguistic feature, was only acquired (with 75–89% accuracy criteria) and not mastered (with ≥90% criterion) by the oldest age group in this study sample (see Appendix X, Table 2).

Specifically, Appendix X, Table 2, shows that by the age of 2;6, most of SHA-speaking children could not acquire the production of any heterosyllabic CC, with only a few clusters emerging by this age (i.e. seven heterosyllabic CCs). By the age of 5;6–5;11, 11 out of the 43 heterosyllabic CCs were never produced to the level of the acquisition criterion by children in this study, where they only reach a customary level of accuracy (emergent with <74% criterion). One heterosyllabic CC /sʒ/ (as in /masʒid/ 'mosque') was never produced at even the customary level of accuracy by the oldest age group (5;6–5;11). In general, 22 out of the 43 heterosyllabic clusters targeted in this study (i.e. 51.2%) were mastered by SHA-speaking children by age 5;11.

5.1.3.3 Comparing of the acquisition of word-final and heterosyllabic CCs.

Even though there are more words with heterosyllabic CCs in SHAPA than those with word-final CCs, a comparison between the developing tendency of both cluster types was conducted based on what has been analysed thus far. Table 5.9 compares the average scores, standard deviation, and range of the PCCC scores for both types of clusters across age groups.

Table 5.9 *A comparison between the average percentage correct scores, SD and range for word-final CCs and heterosyllabic CCs across the age groups*

Age groups	N (children)	Mean (SD)		Range			
				WFCC		Hetero–CC	
		WFCC	Hetero–CC	Min	Max	Min	Max
2;6–2;11	30	31.85 (22.17)	28.78 (21.49)	0.00	77.78	2.04	79.59
3;0–3;5	27	43.62 (24.25)	39.98 (22.14)	0.00	77.78	4.08	75.51
3;6–3;11	38	55.85 (20.74)	54.89 (19.59)	11.11	88.89	8.16	85.71
4;0–4;5	39	69.23 (17.38)	69.02 (15.02)	33.33	100.00	26.53	91.84
4;6–4;11	31	75.99 (19.27)	73.73 (10.67)	22.22	100.00	40.82	89.80
5;0–5;5	37	80.18 (15.52)	81.08 (12.35)	44.44	100.00	40.82	95.92
5;6–5;11	33	81.48 (16.82)	79.22 (9.74)	44.44	100.00	51.02	95.92

WFCC= word-final consonant cluster, Hetero CC= heterosyllabic consonant cluster

As seen in Table 5.9, the acquisition of both cluster types showed an increasing trend with ascending age. Both types began to appear (at a level of >50% to <75%) in the speech of SHA-speaking children by age 3;5 for word-final CCs and by age 3;11 for heterosyllabic CCs, with high SDs and ranges. In general, the average acquisition rate looks similar for both clusters, especially after the age of 4;0 where some children mastered a PCCC score of >90% for both types of clusters.

5.1.3.4 Gemination

The third type of consonantal co–occurrence analysed for this current study is gemination. The single-word naming task targeted nine geminated consonants presented in 12 words (see Chapter 3, Section 3.3.1.3.2). Therefore, a preliminary analysis of consonantal gemination was conducted in this study using data generated from PHON. Average scores for the percentage of geminated middle consonants correct (PCCC) across all age groups were calculated (see Table 5.10). The percentage of the production of each geminated type was then calculated for each age group (see Appendix X, Table 3).

Table 5.10 *Average percentage correct score of geminated consonants (PCCC) across the age groups in the single-word naming task*

Age groups	N (children)	Mean (SD)	Range	
			Min	Max
2;6–2;11	30	48.89 (24.34)	0.00	100.00
3;0–3;5	27	62.65 (21.23)	25.00	100.00
3;6–3;11	38	76.32 (15.92)	41.67	100.00
4;0–4;5	39	85.04 (10.50)	66.67	100.00
4;6–4;11	31	81.99 (12.19)	50.00	100.00

5;0–5;5	37	88.51 (11.84)	50.00	100.00

Even in the youngest age group (2;6–2;11), mean PCCC score of 50% of geminates was achieved, but with a high SD and range (SD=24.34, range=0.00;100.00). Interestingly, some children produced all geminates correctly in all age groups, while some children aged 2;6-2;11 never produced this feature. Germinates can be considered to be acquired at the age of 3;6–3;11 years (the mean PCCC score was 78.95%) and mastered by 5;6–5;11 years (the mean PCCC score was 92.86%). However, in each age group, some children reached a PCCC score of 100% for geminates (see the ranges in Table 5.10).

In general, SHA-speaking children tended to achieve consonantal gemination earlier (with 90% accuracy criterion) than other types of consonantal co-occurrence (word-final CCs and heterosyllabic CCs).

5.1.4 Summary of results II

The results of the phonetic acquisition of SHA-speaking children could be summarised in the following points:

- The phonetic acquisition of SHA-speaking children revealed a developmental trend.
- SHA-speaking children scored>80% for PCC and PCC-R by age 4;0.
- PCC and PCC-R revealed close scores across all age groups, where both scores became much closer with increasing age.
- For PVC, the results show that SHA-speaking children had almost mastered the production of all Arabic vowels by age 3;6.
- SHA-speaking children had already acquired all shared, UHA, and MSA phonetic consonants by the age of 4;5, and their phonetic inventory was mastered by the age of 5;11 when considering the single-word naming task (distortion as correct) except for the dental MSA phone /θ/, which could be added to SHA-speaking inventory after age 5;11.
- Stimulability task showed a later age of acquisition (i.e. after the age of 4;5) of some phones (/ʒ, r, θ, ð, ðˤ/), considering distorted productions as correct, where the mastery age for shared, UHA and MSA found to be 5;11, except for /ʒ/.
- Consonantal co-occurrences analysis revealed that SHA-speaking children showed a similar acquisition rate for word-final CCs and heterosyllabic CCs, where they achieved a score of PCCC >80% by age 5;0.
- They obtained a >80% PCCC score for gemination by age 4;0.

5.2 Discussion II – Phonetic acquisition

One aim of this study was to investigate the phonetic acquisition of 235 SHA-speaking children aged between 2;6 and 5;11 years. The overall phone accuracy was calculated using PCC, PCC-R, and PVC scores across the age groups, while the phonetic inventory of the children was generated using two modes of elicitation: a single-word naming task and a stimulability task. The results of the phonetic inventory and the accuracy of the production of consonants and vowels are discussed in this section.

5.2.1 The rate of acquisition of the Arabic phones

The general hypothesis about phonological development is that children's speech becomes more accurate as they get older (Dodd et al., 2003). The general results of this study support this hypothesis, where the analysis of the single-word naming task revealed that the children's phonetic acquisition rate increased with age. The following section will provide a detailed discussion for each core finding of this study, as well as comparisons with results reported in other studies.

5.2.1.1 Percentage of consonants and vowels correct (PCC, PCC-R, PVC)

As reported in the results section, the children's PCC and PCC-R scores increased with age, demonstrating gradual progress in the child's ability to speak SHA adequately. It was found that their PCC and PCC-R scores nearly reaching the score of 90% for the oldest age group, 5;6–5;11.

It was also noticed that there was a slight reduction in the mean score of PCC and PCC-R for the oldest age group (5;6-5;11). Such a reduction has been described in the psychological literature as 'U-shaped developmental functions' (Werker Hall, & Fais, 2004). This phenomenon explains the transient reorganisation of children's phonological system, or the emergence of an organised system triggered by acquiring new competencies (Vihman, 2014; Werker et al., 2004). Although the reduction in the score of PCC and PCC-R found in the SHA-speaking children's speech was minor, a U-shaped change in performance could explain it as children at this age (5;6-5;11) might begin to construe the input information differently because the exposure to MSA increased (Weker et al., 2004). A similar minor reduction was also observed in the PCC score reported by Amayreh and Dyson (1998) for the similar age group (5;6-5;11), which could support the assumption about the effect of the exposure to the MSA and how the emergence of this new competency caused children to reorganize their phonological system. However, it is important to consider that this minor reduction could also have resulted from the different number of the sample sizes of each age group (the N of 5;0-5;5 age group= 37, and the N of 5;6-5;11 age group was 33). In general, the difference between the mean scores of these two groups was 0.16 and 0.80, respectively; therefore, this difference could have more logically resulted from the difference in the sample size of each age group rather than reflecting the U-shaped change in the phonological development of SHA-speaking children.

The results also suggested differences between the mean PCC and PCC-R scores, as some children in each age group achieved a PCC-R score of >95%. This was not the case with the PCC scores, which shows that the children’s PCC rate was highly influenced by distortions which caused the lower PCC level, not phoneme errors. The differences between the two measures decreased with age and diminished after age 5;0, which indicates that the number of distorted phones in children’s speech decreases with increasing age (Wertzner et al., 2005). A comparison of the differences between PCC and PCC-R with other studies is not possible due to the lack of studies that calculated both scores. Most studies either reported PCC scores (e.g. Aalto, Saaristo–Helin, & Stolt, 2020; Dodd et al., 2003) or PCC-R scores (e.g. Ceron, Gubiani, de Oliveira, & Keske–Soares, 2017; van Haaften et al., 2020). However, Shriberg et al. (1997a) reported that the differences between PCC and PCC-R indicate the percentage of all distortion errors that are common and uncommon clinical distortions.

Compared with the findings from other Arabic studies, the present research revealed overall similar PCC scores across the age groups to those reported in the previous Arabic studies (see Table 5.11). For example, the 4;0–4;5 age group of SHA-speaking children achieved a PCC score of 81.95%, which is very similar to the PCC score reported for Jordanian Arabic-speaking children from the same age (i.e. 80%) and for the Syrian Arabic-speaking children too (i.e. 82%) (Amayreh & Dyson, 1998; Owaida, 2015).

Table 5.11 *Comparison of the percentage of consonant correct (PCC) score observed in this research with those reported for children speaking other Arabic dialects (Jordanian and Syrian)*

Age group	**Amayreh and Dyson (1998) ***	**Owaida (2015) % (SD)**	**Current study % (SD)**	
	PCC	**PCC**	**PCC**	**PCC-R**
2;0–2;4	52%	NA	NA	NA
2;6–2;11	62%	71.53 (2.95)	55.62 (12.76)	61.61 (13.40)
3;0–3;5	73%	81.33 (3.79)	67.83 (15.55)	71.53 (15.34)
3;6–3;11	73%	84.71 (2.79)	74.28 (12.51)	78.70 (11.66)
4;0–4;5	80%	85.60 (1.69)	81.95 (9.21)	85.51 (7.64)
4;6–4;11	76%	90.94 (2.04)	86.64 (7.08)	89.22 (6.35)
5;0–5;5	85%	91.18 (2.49)	89.80 (8.55)	92.36 (7.12)
5;6–5;11	82%	94.84 (1.67)	89.64 (6.67)	91.56 (5.87)
6;0–6;5	90%	96.31 (1.53)	NA	NA

* Amayreh and Dyson’s study did not report the SDs across their age groups.

Table 5.11 showed that the PCC scores for SHA-speaking children are higher than those reported for Jordanian Arabic-speaking children (Amayreh & Dyson, 1998), and lower that those reported for Syrian Arabic-speaking children (Owaida, 2015), considering that Amayreh and Dyson (1998) were based their result on the production of MSA consonants, while Owaida (2015) only thought the dialectal Syrian consonants production. For this current study, both forms of consonants (dialectal and MSA) were counted as correct, where the target words were changed accordingly; thus, the PCC calculation would not be affected by the use of dialectal (i.e. UHA) consonants, and this could explain the higher scores of PCC for SHA-speaking children comparing to those reported for Jordanian–speaking children by Amayreh and Dyson (1998).

It is clear from Table 5.11 that the reported SDs for the Syrian Arabic-speaking children across all age groups are narrower than those reported in the current study. This could indicate that the performance of SHA-speaking children within an age group showed greater variability than children speaking Syrian Arabic. Similarly, the differences in the mean scores of PCC could be related to methodological variants between both studies: in the current study the researcher collected the speech sample using SHAPA, a lengthy assessment tool with 151 pictures, where achieving higher scores may not be possible even for adult speakers. In comparison, 62 pictures were used to elicit the speech sample from the Syrian Arabic-speaking children (Owaida, 2015), and 58 pictures in the AAT were used to elicit the speech sample from the Jordanian Arabic-speaking children (Amayreh & Dyson, 1998). Furthermore, the differences in scores could be attributed to the SHAPA construction criteria, which include more multisyllabic words in the test than those included in the other word lists used in the other studies, as reported by James et al. (2016).

Universally, the PCC score for SHA-speaking children was found to be very close to what was reported by McLeod and Crowe (2018), who found that, on average, children achieved a PCC score of 63.50 by age 2;0, and a PCC score of 93.80 by age 5;0, based on data from 12 different languages including Arabic. Comparing the score of PCC-R in the current study to other literature showed that the PCC-R score obtained by SHA-speaking children is lower than the PCC-R score of children speaking different languages, e.g. Brazilian Portuguese (Ceron et al., 2017) and Dutch (van Haaften et al., 2020). Ceron et al. (2017) and van Haaften et al. (2020) reported the score of PCC-R, and they did not calculate the PCC score for their sample. Thus, comparing the variation in the result of PCC and PCC-R found in the current study to those reported in the literature could not be carried outdue the lack of such studies. However, Shriberg et al. (1997b) reported both scores from their records data of 836 typically developing children and those with disordered speech. They found that the mean score for the typically developing children aged 3;0 was 79.4% on PCC, and their PCC-R score was 92.8%, with distortion accounting for 13.4 % points for the typically developing group's lowered PCC score. However, the difference between PCC and PCC-R for the 8-year-old group was 1.7%, where their mean PCC was 95.8%, and PCC-R was 97.5%, which showed a reduction in distortion occurrences as the age increased. Shriberg et al. (1997b) explained that the difference in measures is that the PCC reflects allophonic mastery, and the PCC-R reflects phonemic mastery. Similar findings were noted for SHA-speaking children in the current study, where the distortion was found to lower the PCC score for the 3-year-old group's children by 3.7% points and for the 5;11-year-old group by 1.92%. Therefore, the result of the current study was in line with Shriberg et al. (1997b) findings and showed a reduction in the differences between PCC and PCC-R scores as children's age increased.

Similar to PCC and PCC-R, vowel accuracy (PVC) increased with age with a nearly-ceiling level from age of 4;6. van Haaften et al. (2020) explained that consonantal production requires more precise speech motor skills than does the production of vowels, and this explains the superior acquisition of vowels over consonants. However, James (2001a) argued that vowel acquisition continues after the age of 3;0. The results of the current study show that the children in the youngest age group achieved a PVC score of 80%, and reaching a score of 90% by age 3;11. Considering the age of vowel acquisition, it was found that only one vowel did not acquired by children at 2;6–2;11 age group (i.e./a͡iː/ which is acquired by age 3;11 and mastered by age 4;11) (see Section 5.1.2.1). The late of age of acquisition for this vowel could be responsible for having a PVC score of 80% in the 2;6–2;11 age group.

The PVC measure has been reported by only one Arabic study, which was conducted by Owaida (2015) on Syrian Arabic-speaking children; her results showed a very similar rate of vowel accuracy to the findings in the current study (see Table 5.12). The PVC scores for the children speaking both dialects showed a steady developmental trend across all age groups, with a high PVC score (>80%) at an early age (2;6–2;11).

Table 5.12 *Comparison of the percentage of vowels correct (PVC) scores observed in this research with those reported for children speaking Syrian Arabic dialects (Owaida, 2015)*

Age group	Owaida (2015)	Current study
	%	% (SD)
2;6–2;11	89.22	80.11 (9.17)
3;0–3;5	90.86	85.70 (7.41)
3;6–3;11	93.17	91.30 (4.01)
4;0–4;5	91.65	93.75 (3.69)
4;6–4;11	94.60	95.66 (2.14)
5;0–5;5	92.11	95.92 (2.32)
5;6–5;11	94.37	96.48 (2.12)
6;0–6;5	98.79	—

Comparing the findings of the current study with those from other Arabic dialect studies, it is clear that the scores for consonant and vowel accuracy differ slightly in these three Arabic dialects (i.e. Jordanian, Syrian and Hejazi). Accepting such differences in dialectal features affects the overall accuracy of children's production (Cole & Taylor, 1990; Phoon, Abdullah, & MacLagan, 2012). It is also essential to consider such variabilities among different dialects when judging phones accuracy. Ignoring such features may lead to misdiagnosing a child with phonological disorders if their dialect was not used as a benchmark for evaluating accuracy (Phoon et al., 2012).

In comparison with PVC scores reported for other languages, SHA-speaking children are less accurate in their vowel production compared to children with other language backgrounds (e.g. English (Dodd et al, 2005), German (Fox, 2000), and Danish (Clausen & Fox-Boyer, 2017)), particularly for children aged 2;6-2;11. Although the vocalic system of UHA is limited to eight vowels only, this results could indicate the effect of Arabic language especially because the findings of this study were in line with the only study calculated the PVC (Owaida, 2015).

5.2.1.2 Phonetic inventory

The phonetic inventories resulting in this study from both tasks (single-word naming task and stimulability task) supported the PCC, PCC-R, and PVC findings in that, as expected, the older children were able to produce more consonants and vowels correctly than younger children.

In terms of the inventory from the single-word naming task, the UHA and MSA shared consonant inventory was almost completed at age 5;5, regardless of distorted productions. For the MSA consonants, the phonetic inventory was completed by age 5;11, except for /θ/, which only acquired with 75% of accuracy by the 4;5 and could be mastered after age 5;11. It could be concluded that Arabic consonants (shared, UHA, and MSA phones) can be divided into three groups based on the findings of the single-word naming task data (90% criterion accuracy) with distortions counted as errors:

- Early sounds (2;6–3;5), which include /b, t, d, k, g, ʔ, f, ð, ðˤ, s, ħ, ʕ, h, m, n, l, w, j, ɾ/.
- Middle sounds (3;6–4;5), which include / tˤ, dˤ, sˤ, z, x, ɣ, r/.
- Late sounds include (4;6–5;11) /q, ʃ, ʒ/.

In terms of the inventory from the stimulability task, the shared consonants inventory was only acquired, except for /r/ and /ʒ/, by children aged 3;11 when all distortions were counted as correct. When the distortions were considered as errors, the phonetic inventory was never completed in the stimulability task because of the inability of SHA-speaking children to correctly imitate the fricatives prone to distortion (i.e. /s, z, sˤ, ʃ, ʒ/) correctly with 90% of accuracy. Further, all MSA phones were added to the inventory of SHA-speaking children with a mastery level at age 5;6–5;11. Therefore, the early, middle, and late mastered sounds (with 90% of accuracy), with distortions counted as errors, according to the results of the stimulability task are as follows:

- Early sounds (2;6–3;5), which include /b, t, k, ʔ, f, ħ, h, m, n, l, w, j/ .
- Middle sounds (3;6–4;5), which include / tˤ, d, g, ɣ, ɾ/.
- Late sounds (4;6–5;11), which include / dˤ, q, θ, ð, ðˤ, x, ʕ/.

It could be noted that MSA phones were acquired and mastered after the acquisition and mastery of UHA phones. SHA-speaking children completed their UHA phonetic inventory by age 5;5, which is earlier than the age of completion of the MSA inventory (i.e. after 5;11) in terms of mastery level of accuracy (with 90%). The late mastery of MSA phones corroborated with the theoretical foundation presented in Chapter 2, Section 2.1.2, about SHA-speaking children's late exposure to MSA, which usually started after school entry. Although some MSA phones, such as the emphatic dental /ðˤ/, had been acquired and mastered early by SHA-speaking children (i.e. by age 2;6–2;11) than other MSA phones, it is essential to acknowledge that these phones are parts of different Saudi dialects (e.g. Saudi Najdi), and this early acquisition of these sounds could indicate the influence of the exposure to these dialects, not because the exposure to MSA. Such findings support previous studies that have found that frequently occurring sounds are produced with higher accuracy than less frequent sounds (Kirk & Demuth, 2002). The more often a child is exposed to sounds of a language, the more opportunities this child could have to establish and enhance his/her phonological representations and to practice the oral motor processing skills to support the emergence of the ambient language's speech sounds (Davis & Bedore, 2013). Because MSA phones were less frequent in the speech of SHA-speaking adults, SHA-speaking children showed a late acquisition and mastery for them, unlike the shared phones that are acquired early because they are more frequent in adult speech.

As reported in the results section, five fricatives were still missing in the children's phonetic inventory obtained via the stimulability task (i.e. /s, z, sˤ, ʃ, ʒ/) and were only mastered when their distorted productions were considered as correct. In general, the phonetic inventory of SHA-speaking children showed a superior rate of acquisition via the single-word naming task compared to the stimulability task. The differences between the inventories of both tasks could be related to two different reasons. First, the results could be due to the application procedure of the stimulability task, where the task included only phone imitation in isolation. During the administration of the stimulability task, each child was provided with a maximum of two trials to imitate each sound in isolation only, with brief cueing applied in the second trial where the examiner asked the child to listen carefully and to look at her face. No sensory cues were provided (cf. Wertzner et al., 2005). Such a procedure could limit the results of the stimulability task because not all phones could be easily imitated in isolation, and maybe some of these sounds need more sensory cues to be correctly imitated. Second, these differences in the phonetic inventories obtained via both tasks could be related to the struggle children may face in differentiating between some sounds in isolation. Therefore, many children in the 2;6– 2;11 age group could not imitate the emphatic sounds /sˤ, tˤ, dˤ/ because they were very difficult to produce in isolation, which may cause that children to fail to differentiate these emphatic sounds from their non-emphatic counterparts. Moreover, many children refused to imitate the trill /r/ in isolation; however, they were able to correctly produce it during the single-word naming task. This finding supports other research's assumption that the speech task can influence the phonetic inventory. Yeh and Liu (2021), for example, found that the phonemic inventory obtained from CS differs from that obtained via single-word naming task due to salience and avoidance effect, given that children tend to avoid producing unfamiliar phones and using words that contain elements already exist in their phonetic inventory. Similarly, some children in this study avoided imitating complex sounds in isolation, while they could produce them within word productions.

Furthermore, it is necessary to consider whether applying a complete cueing hierarchy (Kubaschk, Fox-Boyer, & Klann, 2015) or including syllable imitation with different vowel contexts (Tyler & Macrae, 2010) may result in more reliable findings. Unlike the stimulability task, the single-word naming task in the current study included a complete hierarchy of cueing (see Chapter 3, Section 3.3.1.4), in addition to multiple opportunities for each phone to be produced by children in different word positions in different word contexts,). In contrast, the stimulability task for the current study involved no instruction other than visual and auditory stimuli. Tyler and Macrae (2010) reported that children's performance in a stimulability task could be affected by many factors, including the stimulus and context or degree of scaffolding. They found that stimulability was associated with the visibility of speech sound production, with more visible sounds being imitated with tremendous success. On the other hand, Kubaschk et al. (2015) stated that it is still unknown which cues can best stimulate sounds. They found that indirect cues (e.g. sound gestures) successfully lead to correct sound imitation in isolation. Still, they reported that there is no guarantee that cues would help children who did not spontaneously imitate a sound correctly, as only 47% of children responded to the cues. Therefore, lacking a cues hierarchy may not be a reasonable explanation for the SHA-speaking children's performance in the stimulability task; however, it still could be considered a limitation in applying this task in this study.

It is also important to highlight that an inability to imitate phones in isolation correctly was not necessarily linked with atypical or delayed phonological development (Kubaschk et al., 2015). Kubaschk et al. (2015) reported that typical and atypical German-speaking children who only produced age-appropriate phonological development patterns were only stimulable for some phones. Some children with atypical phonological patterns could imitate all German phones. Then, for SHA-speaking children, the inability to imitate some UHA and MSA phones could not indicate atypical development for their phonetic-phonological system.

Moreover, the children's attention could be another factor (Powell & Miccio, 1996), as the stimulability task was administered after the single-word naming task, which was a long task that took approximately 30 minutes. Consequently, the children's attention and motivation to imitate the sounds correctly could have been negatively affected during the test, resulting in lower performance in this stimulability task compared with the naming task.

Comparing the phonetic inventory results to those of other Arabic studies, the current study revealed that SHA-speaking children have an earlier age of acquisition and mastery for most Arabic phones compared to those reported in the previous Arabic studies in general. It is important to bear in mind that the current study implemented a phonetic approach where the sound included in the inventory if it occurred two times in a child's speech regardless to its position in the word, unlike most of the earlier reported Arabic studies which have described the phonemic repertoire[15] rather than phonetic inventories.

The findings of the current study can be compared to two studies reported phonetic inventory in Arabic-speaking children: one study conducted on Jordanian Arabic- (Amayreh & Dyson, 2000) and the other study on Saudi Najdi Arabic-speaking children (Alawwad, 2009). Alawwad (2009) reported that the phonetic inventory of Saudi Najdi Arabic-speaking children was completed by age 4;0; however, no criteria were noted for this completion if it was at mastery (90%) or acquisition (75%) of accuracy. Although the limited number of participants compared to the current study, Alawwad's findings demonstrated an agreement with this study where SHA-speaking children expressed their ability to produce the emphatic plosive phones /dˤ, tˤ/ with 75% of accuracy by age 3;0–3;5, in line with Saudi Najdi Arabic-speaking children (Alawwad, 2009). In contrast, the consonant phonemes /dˤ, tˤ/ were acquired by children speaking other Arabic dialects on average by age 5;0–5;5. Although these sounds /dˤ tˤ/ have the emphatic feature, which created some production difficulty for children even at a later age (<6;4) (Amayreh, 2003), the result revealed that SHA-speaking children had acquired these two emphatic phonetic consonants, as well as the two others /ðˤ, sˤ/, before age 3;11. This finding may reveal that Arabic children acquired the emphatic feature by age 3;11, producing all the emphatic phones, but they can use them phonemically correct by or after age 5;0–5;5. Therefore, the factor of 'articulatory precision' that was reported by Amayreh (2003) to justify the late acquisition of the emphatic consonant phonemes could not be applied to SHA-speaking children as they expressed their ability to pronounce the emphatic

[15]The phonemic description in this current study was provided by describing the typical developmental patterns in SHA-speaking children (see Chapter 7).

consonant phones. However, this assumption needs further examination for the phonetic inventory across Arabic children speaking different dialects.

The current study's findings coincided with Amayreh and Dyson's (2000) reported phonetic inventory of children aged 1;2 to 2;0. It was found that by the age of two, SHA-speaking children had mastered the production of all the phones stated to be acquired in the phonetic inventory of Jordanian Arabic-speaking children (i.e. /b, t, d,, h,, m, n, l,, j, w/). This discovery suggests that SHA-speaking children may acquire these phones before the age of 2;6. However, Amayreh and Dyson did not identify three phones that SHA-speaking children mastered by 2;11 (i.e. /f, ð, s/), therefore their age of acquisition might be before or after age 2;6. As a result, data from SHA-speaking children under the age of 2;6 may be necessary. This distinction may indicate dialect distinctions between Hejazi and Jordanian and the impact of methodological discrepancies. Although dental fricatives are not part of the UHA phones, its early mastery by SHA-speaking children may indicate the influence of other Saudi dialects that have this phone in their inventory (e.g. Saudi Najdi), as SHA-speaking children are usually exposed to MSA after the age of 3;5 when they begin to learn the Alphabetic in their kindergartens (see Chapter 2, Section 2.1.2).

Previous Arabic research, as aforementioned, reported the consonant phonemic age of acquisition and mastering. They considered consonant phonemes to be 'acquired' if they were produced correctly in two distinct word positions by 75% of the children, and 'mastered' if they were produced correctly in all three-word positions (i.e. initial, medial, and final) by 90% of the children. Although these Arabic studies claim to have explored the acquisition of Arabic consonant phonemes according to Shriberg et al. they did not describe inventories where the sounds 'had been attested as contrastive' (1997a, p.710). Their definition of phonemic consonant acquisition did not contain information concerning realised sound contrasts, where children may distinguish between and among different words (Stokes et al., 2005). As a result, their definition of sound acquisition and mastery may be interpreted as a description of phonetic acquisition and mastery while word positions of sounds are taken into account.

Table 5.13 displays a comparison between the result of the current study and the findings of the other Arabic studies. The mean age of acquisition and mastery was counted using data from four studies targeting children speaking four different dialects (i.e. Jordanian, Egyptian, Kuwaiti, and Syrian, see Chapter 2, Table 2.10 and Table 2.11) following the method conducted by Crowe and McLeod (2020) (see Chapter 2, Section 2.2.1.1).

Table 5.13 *Comparison of phonetic acquisition and mastery in SHA-speaking children with phonemic acquisition in children speaking other Arabic dialects*

Age groups	Phonetic acquisition* (75% accuracy)	Phonetic mastery* (90% accuracy)	Phonemic acquisition+ (75% accuracy)	Phonemic mastery+ (90% accuracy)
<2;6–2;11	/k, ðˤ, z, ʃ, r/	/b, t, d, g, ʔ, m, n, f, ð, s, ħ, h, l, w, j/	/t, k, g, ʔ, m, n, f, ħ, w/	
3;0–3;5	/tˤ, dˤ, sˤ/	/k, ðˤ, ʕ/	/b, d, j/	/n, m, ʔ, f, w/
3;6–3;11	/x, ɣ/	/tˤ/	/ʕ, h/	/b, t, k, ħ, j, l/
4;0–4;5	/q, θ, ʒ/	/dˤ, sˤ, z, x, ɣ, r/	/s, ʃ, dʒ/	/d/
4;6–4;11		/ʃ/	/sˤ, z/	/g, s, x, ɣ, ʕ/
5;0–5;5		/ʒ/	/tˤ, dˤ, θ/	/sˤ, z, ʃ, r/
>5;6–5;11		/q/	/q, ð, ðˤ, ʒ/	/tˤ, dˤ, θ, dʒ/

Key: Shared phones (dialectal and MSA), MSA only. Note: * Information on phonetic acquisition and mastery is based on the results of the single-word naming task with all distortions counted as errors in the current study for SHA-speaking children. + Information on phonemic acquisition and mastery is based on the average age reported in the previous studies for children speaking the other four different dialects (Amayreh & Dyson, 1998; Ammar & Morsi, 2006; Ayyad, 2011; Owaida, 2015).

As described in Chapter 2, Section 2.2.1.1, these studies (Amayreh & Dyson, 1998; Ammar & Morsi, 2006; Ayyad, 2011; Owaida, 2015) used different approach from that used in this current study. Still, it could be notice that there is a similar order of acquisition occurs across both approaches, with different age of acquisition for individual phones. This finding generally supports Dodd et al.'s (2003) assumption about the phonetic acquisition that is expected to occur prior to phonemic mastery. These differences in the age of acquisition could be due to the phoneme repertoire being derived from the correct production of sounds in all word positions; however, the phonetic inventory includes sounds articulated in any context: correctly, as a substitute for another sound, imitated in a CV syllable or was stimulable in isolation (Holm et al., 2021). Dodd et al. (2003) explained that when children are first exposed to a word, they may imitate it correctly (e.g. chicken), but they may produce it incorrectly when they named it as a lexical item, where they may use a system–level sound substitution (e.g. ‘chicken’ is pronounced /tɪʔən/). Phonetic acquisition of /tʃ/ has occurred but not phonemic mastery (Dodd et al., 2003, p. 637). This suggests that the child can produce sounds, use the phonological representation to map a phoneme, and has knowledge of particular words but has not yet developed abstract phonological categories (phoneme sequences) for production (de Castro & Wertzner, 2012; Vihman, 2007). The differences between the findings of the current study and those of previous Arabic studies are related to the different ways of analysing the phones development process according to the phonetic versus phonemic description, which would be expected to yield differences (Holm et al., 2021).

It is worth noting that some researchers did not consider specific MSA phones absent from the dialect under study (e.g. Ammar & Morsi, 2006; Ayyad, 2011; Owaida, 2015). Although the results of this study, which were consistent with Amayreh and Dyson (1998), demonstrated a late acquisition and mastering of MSA only phone, this corroborated with the theoretical foundation presented in Chapter 2, Section 2.1.2 about SHA-speaking children's late exposure to MSA. Even with the late introduction to MSA phones, Arabic researchers should not neglect them as long as these phones are required to build Arabic literacy skills. As a result, failing to record the age of acquisition and mastery of these MSA-only phones would lead to missing valuable information about these children's phonological development.

Comparing the results to languages that differ from Arabic, the phonetic acquisition observed in this study has a similar developmental trend to that reported for children speaking other languages such as British English (Dodd et al., 2003), German (Fox, 2000), and Maltese (Grech, 2006). Nasals /m, n/, plosives /b, t, d, k, ɡ/ including the glottal stop /ʔ/, glide /w, j/, pharyngeals /ħ, ʕ/ as well as the glottal/h/, and some of the front fricatives /f, s/ were acquired by SHA-speaking children earlier than the postalveolars /ʃ, ʒ/, back fricatives /x, ɣ/, and liquid /r/. This acquisition pattern agrees with the universal patterns proposed by Jakobson (1941/1968) as reported in Stoel-Gammon (1985), whereby children speaking different languages are viewed as showing more or less a similar sequence of acquisition. In line with these universal expectations, it is assumed that children acquire sounds that are common in all languages earlier than sounds that are less common across languages (Maphalala et al., 2014). This was proven in the results of the current study, where the children acquired nasals, plosives, and glides, which are common in many languages, before the language-specific phones (e.g. emphatic phones). The findings of this study also confirmed the results of the recent studies which proved that most consonants were acquired before age 5;0, with few exceptions (e.g. /θ/, /r/) (Crowe & McLeod, 2020; McLeod & Crowe, 2018).

At the same time, the language–specific trends reported by Grech (2006) and Amayreh and Dyson (1998) were also observed in the results of this study. SHA-speaking children were shown to have acquired some UHA back fricatives consonants such as /ħ, ʕ/, which are not common in many languages, earlier than other common fricatives such as /z, ʃ/. Such a finding supports the notion of the influence of language exposure on the development of a child's speech. According to Ingram (1989), the common sounds in the ambient language are often acquired first, regardless of their articulatory complexity. The current research supported Ingram's assumption as almost all the UHA phones were acquired before MSA (e.g. the emphatic UHA /dˤ/ acquired before the interdental MSA /θ/). Further, the result of this study aligns with those reported in McLeod and Crowe (2018) regarding the early acquisition of the pharyngeals /ɣ, ħ, x, ʕ/ across languages, which SHA-speaking children acquire by age 3;0–3;11.

5.2.1.3 Consonantal co–occurrences

In terms of the rate of acquisition of consonantal co-occurrences and the percentage of consonant clusters correct (PCCC), a general developmental trend across age groups was also observed in the accuracy percentages as well as the age of acquisition and mastery of all types of consonant cooccurrences: final CCs, heterosyllabic CCs, and gemination. These results are discussed below.

For word-final CCs, it is important to bear in mind that the number of word-final CCs in SHAPA was limited to only nine words with a single occurrence of each CCs, which did not represent a full range of all possible word-final CCs allowed in SHA. The results from analysing this limited number of word-final CCs may not represent the complete picture of the acquisition of such a word structure among SHA-speaking children. However, the results showed that the youngest age group (2;6–2;11) achieved a low mean PCCC score, with a high range, and some children never correctly produced this word structure. Still, in this age group, some children produced word-final CCs with >75% accuracy. It is important to note that the PCCC score calculation was based on counting both cluster elements as correct without any segmental errors. By the age of 4;11, the average PCCC score showed some improvement and reached 75% accuracy.

In terms of the inventory, the word-final CC inventory presented in SHAPA was never completed by the children in this study. However, seven out of the nine word-final CCs presented in SHAPA were acquired before age 5;5 and mastered by age 5;11. Two clusters were never acquired by children /rʒ/ and /ʒh/. These unacquired clusters include the postalveolar fricative /ʒ/ as one of their elements, which is one of the late consonants acquired as a singleton by SHA-speaking children, and it was acquired by age 5;5 and never mastered by age 5;11.

In terms of children speaking other Arabic dialects, a limited number of studies have provided data on the acquiring of word-final CCs. Although multiple studies are available on the typical speech sound development of Arabic-speaking children in different Arabic dialects, no recent studies have focused on the percentage of CCs correct nor the CC inventory of Arabic-speaking children. The only study to report a detailed presentation of final CC acquisition was the study of Ragheb and Davis (2014); however, they did not provide data about the PCCC of the word-final CCs they examined. In contrast, it was found that Kuwaiti Arabic-speaking children acquired with 75% accuracy word shape with final CCs by age 4;0 according to Ayyad (2011). This finding showed some similarity to the current study where SHA-speaking children expressed their ability to produce word-final CCs by age 4;0. (see Table 5.7). Alqattan (2015) also found that the final CC is the most frequently targeted cluster in the speech of Kuwaiti children aged between 1;4–3;7, and they appeared earlier than other types of clusters (i.e. initial and medial) and that in line with universal acquisition of CC (McLeod, Van Doorn, & Reed, 2001b), where 53% of all clusters in their speech occurred as coda cluster (i.e. word-final CC). In general, the result of this current study showed an agreement with what was reported in these previous Arabic studies in terms that children as young as 2;6–year–old developed the ability to produce word shapes with final CC; however, the acquisition level of accuracy of these CC may not be guaranteed till age 4;0.

Undertaking a comparison of the development of CCs is a challenging task in general since, according to McLeod, Van Doorn, and Reed (2001b), there is an insufficient number of studies investigating the development of CCs worldwide. McLeod and colleagues concluded from their literature review that between age of two and three, children speaking English, German, and Spanish, were able to produce CCs, mostly word-initial CCs, in their early word production. For English, where CCs occur in both positions, word-final CCs are generally reported to appear earlier than word-initial CCs (McLeod et al., 2001b). For example, Danish-speaking children were found to show earlier mastery of several word-final CCs than word-initial CCs (i.e. by age 2;6–2;11)(Clausen & Fox-Boyer, 2017). McLeod et al. (2001b) explained this early production as the maturation of the children's motor speech mechanism and continued anatomical development of the oro–musculature. In their investigation of 16 typically developing Anglo-Australian two-year-old children, McLeod, Van Doorn, and Reed's (2001a) findings supported the trend of the gradual development of CCs (for both word-initial and word-final position clusters) among their participants, with an increasing score for the percentage of CCs correct (PCCC) observed among all groups over time, similar to the result of this study.

In general, it is evident that Arabic-speaking children, including SHA-speaking children in the current study, demonstrated a later age of acquisition for word-final CCs (i.e. almost at age 4;0) than that reported for other languages such as English, Maltese and German (i.e. almost at age 3;0), although Maltese is the language that considered as a derived dialect of Arabic. Grech (1998) reported that there are too many words with CC in Maltese either in the initial word position or final position. Thus, the possible explanation for such a difference between the result of this study and performance of Maltese children in Grech's study (1998) could be related to the limited occurrences of words with final CC in general in SHA-adult's speech, particularly in the child-directed input. Most of the words with word-final CCs in UHA are not familiar to children (i.e. MSA words) and adult-speakers used to break the clusters in these MSA words (see Chapter 2, Section 2.1.3.2.1). Therefore, the number of items with word-final CC in SHAPA was limited and did not reflect the real performance of SHA-speaking children regarding this phonetic aspect. This is unlike English, for example, where one-third of monosyllabic words in the wordlist used by McLeod and her colleagues begin with a CC and 18% of monosyllables end with a CC (McLeod et al., 2001b).

Similar to the results for word-final CCs, the mean score for heterosyllabic PCCC was low for the youngest age group (2;6–2;11) in the current study. It approached a mean score of 80% accuracy for the 5;0–5;5 and 5;6–5;11 age groups; however, from the age of 4;0, some children achieved a PCCC score of >90% accuracy. The inventory of heterosyllabic CCs also reflected a developmental trend, where only seven CCs started to emerge by the age group 2;6–2;11. By 5;6–5;11, 22 of the 43 heterosyllabic CCs presented in SHAPA had been mastered. However, the inventory of the heterosyllabic CCs was not completed by 5;6–5;11; some children never produced 11 CCs at an acquisition level, and they were just emergent with <74% of accuracy.

Two heterosyllabic CCs that were not acquired (i.e. /dˤr/ and /sʒ/) have consonants in their elements that were either mastered or acquired as a phonetic singleton only by the oldest age group (5;6–5;11) of SHA-speaking children. It has been reported that the production of such late consonants in a complex structure as heterosyllabic CC causes children some difficulties in pronouncing them correctly. Stockman and Stephenson (1981) clarified that the production of such a cluster requires the ability to adjust articulatory gestures to meet the demands of a rigid timing system for speech. Similarly, Dyson and Amayreh (2000) reported that such a structure is more taxing to the articulatory mechanism than a single consonant. They noticed among their participants that the most marked segment of the cluster, usually acquired later, was typically deleted by children aged between 2;0 and 4;0 in their sample (e.g. /d.r/→[d]). Similarly, Kuwaiti Arabic speaking children aged 4;6–5;2 had acquired the disyllabic word shape CVCCVC and CVCCV, while children aged 3;0-4;0 did not show competence for producing these word shapes with heterosyllabic CCs (Alqattan, 2015; Ayyad, 2011). It could be concluded that the general development of the heterosyllabic CCs among different Arabic dialects showed the same trend found in the current study.

Cross-linguistically, few studies have focused solely on such a type of consonant cluster (heterosyllabic CCs or consonant sequences across syllable boundaries). Most have described the acquisition of the medial consonant singletons without considering the whole sequence of consonants across syllable boundaries (e.g. Amayreh & Dyson, 1998; Petinou & Theodorou, 2016). Grech (2006) reported her observation about on how Maltese-speaking children aged 2;0–3;0 deal with heterosyllabic CCs. Grech noticed that they consistently mismatched these consonantal sequences. They tended to lengthen one element of the consonant sequence and delete the other consistently, known as the gemination of a consonant in sequence pattern. This pattern was found in the speech of Maltese-speaking children, as well as SHA-speaking children, and continue until age 3;6 (see Chapter 6, Section 6.1.2), which may indicated that the heterosyllabic CC could be mastered or acquired after 3;6. Similarly, Topbas (1997) reported that Turkish-speaking children demonstrated difficulties in pronunciation of consonants in –C,C– adjacent contexts; however, she did not reported the age of acquisition of such a word shape. Topbas also noted that consonants on the syllable’s boundaries (i.e. heterosyllabic CCs) were subjected to deletion by most of the children in her sample due to a coarticulation effect. Such a pattern was also observed among SHA-speaking children (see Chapter 6, Section 6.1.2).

In general, the acquisition rate for both types of clusters revealed a similar trend, especially after age

4;0, where some children scored >90% of PCCC. The number of acquired word-final and heterosyllabic CCs increased with age; however, the full mastery of the CC inventory was not necessarily gained by children in the oldest age group in this study (i.e. 5;6–5;11). It is evident that by age 5;6–5;11, 77.78 % of word-final CCs (i.e. 7 out of 9) were produced with a mastery level of accuracy compared to 51% of heterosyllabic CCs (i.e. 22 out of 43). This shows that SHA-speaking children tend to maintain the production of word-final CCs earlier than heterosyllabic CCs.

Gemination is another type of consonantal co–occurrence analysed in this study. It is the only type of consonantal co–occurrence that showed an age of acquisition as early as 3;6, where the mean PCCC was >78% of accuracy. SHA-speaking children achieved a mastery level of accuracy by age 5;6–5;11. Comparing the results with other studies, this developmental trend is in agreement with that reported by Dyson and Amayreh (2000) and Alqattan (2015) regarding the early acquisition of the gemination feature among Jordanian– and Kuwaiti–Arabic speaking children. Further, the Khattab and Al– Tamimi's (2012) conclusion's aligns with the finding of this study, where SHA-speaking children demonstrated acquisition of geminated consonants by the age of 3;6, and the mastery by the age of 5;6. It was found that the acquisition of the geminated feature among other languages reflects a developmental trend with a fluctuating (or U–shape) rate in children before the age of 3;0 (Khattab & Al-tamimi, 2015). Lebanese Arabic-speaking children may not accurately acquire the actual acquisition of the singleton–geminate contrast until a later stage of development (i.e. after the age of 3;0). However, this rate of acquisition could not be observed in the current study, as the youngest age group included in this study was 2;6, which, according to Khattab and Al–Tamimi (2012), is the age following the early word production period where accurate production is already expected.

Savinainen–Makkonen (2007) found that a child aged 1;1, who was acquiring Finnish, produced many words with a syllable structure of the (C)VCːV shape, which included a geminated consonant. She concluded that the medial geminate template is the most salient part of a word for Finnish children, and children use this word shape as an anchor to practice new words. This result, as well as the early acquisition of gemination observed among children in the current study, could indicated that the process of acquiring gemination starts at a younger age (i.e. before age 2;6) for SHA-speaking children. Similarly, Grech (2006) found that Maltese-speaking children aged 2;0–3;6 used to geminate the abutting consonants (consonantal sequencing) consistently, and a similar pattern observed in children in this current study (see Chapter 6, Section 6.1.2). This process could be explained by Khattab and Al-Tamimi's (2012) agreement regarding the generalisation of the geminate consonant template over the early words of children speaking a language with the geminated feature, especially before the age of 3;0.

5.2.2 Summary of discussion II

The discussion of the results of the phonetic acquisition for SHA-speaking children could be summarised in the following points:

- The phonetic acquisition for SHA-speaking children matched the universal phonetic development trend and reflected the language-specific trend.
- PCC and PVC scores obtained by SHA-speaking children were similar to those reported in other Arabic studies (Amayreh & Dyson, 1998; Owaida, 2015).
- Universally, the PCC score for SHA-speaking children aged 5;0 was similar to the PCC score reported by McLeod and Crowe (2018) for 12 different languages.
- The differences between PCC and PCC-R scores found for SHA-speaking children were similar to those found by Ceron et al. (2017), reflecting a reduction in the distortions occurrences with age.
- Compared with PVC scores reported for other languages, SHA-speaking children are less accurate in their vowel production compared to children with different languages background, but similar to children speaking Syrian Arabic, as the only study reported PVC scores.
- Phonetic inventory for SHA-speaking children complied via SWN task was almost completed for the shared phones before MSA phones because of the late exposure to MSA phones.
- Phonetic inventory for SHA-speaking children showed a superior acquisition rate via the SWN task compared to the stimulability task.
- The administration procedure could affect the results of the stimulability task.
- SHA-speaking children have an earlier age of acquisition and mastery for most Arabic phones compared to the age of acquisition reported in other Arabic studies, as the current study considered the phonetic feature of a phone without considering the correct use of the phone (phonemic production) that reported in other Arabic studies.
- The patterns of phones acquisition identified in SHA-speaking children matched the universal acquisition reported by Jakobson (1941/1968).
- The influence of language exposure on the development of a child’s speech was noticed in the speech of SHA-speaking children as they acquired some back fricatives /ħ, ʕ/ before other common fricatives /z, ʃ/.
- Consonantal co-occurrence analysed in this study could not be compared across Arabic dialects due to a lack of studies that conducted a similar analysis.
- The acquisition of the word-final CC and geminated consonants by SHA-speaking children was similar with the age of the emergence of word-final CCs (at 2;6) by children speaking other languages but contrasted with the age of acquisition (i.e. 4;0) compared to other languages (like English and Maltese which is almost at age 3;0).
- For heterosyllabic CCs age of acquisition, there is a general agreement between the current study and other studies on children speaking different languages that such a word structure is complex and is acquired later by children (after the age of 3;6).

6 Results and Discussion III – Phonological Variants in Saudi Hejazi Arabic-speaking children

In this chapter, the result of the phonological variants analysis is presented. These results will then be discussed to answer the third research question:

RQ3: Which phonetic and phonological error patterns are found in the speech of Saudi Hijazi Arabic monolingual speaking children aged between 2;6, and 5;11? Particularly in terms of:

3a) To what extent does the frequency of phonological variants (Tokens and pattern Types) alter over time as phonological development progresses?
3b) What is the frequency of infrequent variants across age groups?

3c) What phonological patterns can be observed as developmental at different ages using two distinct cut-off criteria?

6.1 Results III – Phonological variants

This section is divided into two subsections. First, the results of the phonological variants (PhonVar) analysis are presented, including the number of tokens, types of phonological patterns and infrequent variants (InfrVar). Second, a description of the types of developmental phonological patterns, as well as their age of occurrences, is provided.

6.1.1 Tokens, types, and infrequent variants

To address research questions 3a and 3b, the mean and standard deviation (SD) for Token, Type and InfrVar were also computed for each age group (see Table 6.1). It is important to note that this analysis was run twice to apply the two cut-off criteria (occurring at least four to six times; see Chapter 3, Section 3.6.3) for the classification of phonological variants as either patterns (Types) or InfrVar. As the number of Tokens covers all phonological variants independent of it being classified as patterns or InfrVar, it was calculated independent of the different cut-off criteria used in this study (Table 6.1).

Table 6.1 *Mean (M) and standard deviation (SD) of Tokens, Types of developmental phonological patterns and InfrVar in each age group according to the two cut-off criteria*

Age group	n	Cut-off criteria	Tokens M (SD)	Tokens Range	Type M (SD)	Types Range	InfrVar M (SD)	InfrVar Range
2;6–2;11	30	≥4	189.53 (71.73)	73–324	12.77 (3.93)	4–19	40.67 (9.42)	25–61
		≥6			9.83 (3.71)	4–17	53.57 (9.97)	32–75
3;0–3;5	27	≥4	163.74 (82.63)	51–320	10.85 (4.81)	2–20	34.93 (9.38)	16–52
		≥6			7.67 (4.10)	1–15	49.00 (14.13)	21–82
3;6–3;11	38	≥4	125.00 (75.22)	22–356	8.84 (4.58)	0–17	30.24 (7.89)	13–48
		≥6			6.11 (4.24)	0–16	42.37 (11.04)	21–72
4;0–4;5	39	≥4	75.46 (43.07)	15–188	5.82 (3.77)	1–14	24.69 (7.17)	11–37
		≥6			3.62 (2.87)	0–10	34.33 (11.81)	11–64
4;6–4;11	31	≥4	60.81 (33.77)	24–162	3.81 (2.86)	0–12	24.90 (5.09)	14–34
		≥6			2.32 (2.06)	0–8	31.42 (9.15)	17–53
5;0–5;5	37	≥4	42.38 (37.65)	9–181	2.73 (2.69)	0–10	18.11 (7.31)	6–36
		≥6			1.70 (1.87)	0–6	22.4 (10.29)	6–46
5;5–5;11	33	≥4	44.03 (32.01)	4–148	2.79 (2.48)	0–10	18.45 (8.09)	4–45
		≥6			1.39 (1.50)	0–5	24.45 (11.36)	4–58

Token values are not affected by the applied cut-off criterion

Overall, SHA-speaking children produced a comparably high number of phonological variants up to the age of 5;11, but the total number of Tokens decreases with increasing age. As can be noted from Table 6.1, there is a steady reduction in Tokens from the youngest age group (2;6–2;11), who produce the largest number of Token (*M* =189.53), to the oldest age group (5;6–5;11) where the mean for the number of Tokens reduced to 44.03. The SDs showed consistent scores from the age of 4;6–4;11, unlike the younger age groups where there was a steady reduction in SDs with increasing age. However, there is marked variation in the ranges across all age groups, which indicating large variability in the children's performances within each age group.

It is also clear from Table 6.1 that the youngest age group exhibited a larger number of phonetic/ phonological patterns Types on both criteria than the oldest children (i.e. at 2;6–2;11 *M* = 9.83 for ≥6 criterion, and for the age of 5;6–5;11 *M*= 2.79 for ≥4, and *M*=1.39 for ≥6). The standard deviation, which reflects the variability in the children's use of phonological pattern types, decreased from SD=3.71 for ≥6 in 2;6–2;11 age group, to be SD=1.50 for ≥6 in the 5;6–5;11 age group; the ranges also indicated a reduction in the variation of patterns used by SHA-speaking children across the age groups. It is also evident that the number of pattern Types calculated using the higher cut-off (i.e. ≥6 was higher than that found with the lower criterion (i.e. ≥4). Further, it is clear that InfrVar accounted for more than half of the Tokens of the first three age groups, while for children aged ≥4;0, the InfrVar accounted for almost half of their Tokens, indicating that most of the PhonVar of children older than 4;0 were InfrVar.

In general, the number of phonological pattern Types decreased considerably with age. Although there is a stability in the pattern Types from the age of 5;0–5;5, the phonological acquisition processes do not seem to have been fully completed.

Examining the InfrVar in Table 6.1, their number was also decreased with age. The results show that the youngest age group expressed a high number of InfrVar, which might reflect children's inconsistent performance across this age group. As expected, in contrast to the results for the pattern Types, when applying the higher criterion (≥6), the rate of InfrVar increased compared to the lower one (≥4). At the age of 2;6–2;11, the mean of InfrVar is almost a quarter the mean of Tokens of this age group. In contrast, at the age of 5;6–5;11, the mean of InfrVar equals almost half the mean of Tokens at this oldest age group. Although the number of InfrVar decreased with increasing age, this reduction is slow in rate and showed a high variability (see *SDs* in Table 6.1), which reflects some inconsistency in the productions of the older children as well.

6.1.2 Developmental phonological patterns and their age of suppression

To answer research question 3c, PhonVar occurring at least four (≥4) or six (≥6) times depending on the cut-off criterion were classified as phonological patterns and their percentage of occurrence in each age group was calculated; the results are reported in Table 6.2. Appendix Y presents the same table with further details about the number of children producing each type of pattern in each age group. The patterns have been defined; for example, fronting patterns are described in more detailed, so the labels reflect which sounds were affected (i.e. fronting of /ʃ ʒ/→ [s z θ ð], fronting of /s z/ → [θ ð], and fronting of /k g/→ [t d]**).**

Following Gruwell's (1981) classification, patterns were divided into systemic simplifications, structural simplifications, and phonetic distortions. Systemic simplifications relate to substitutions of single sounds of a syllable or word without altering the overall structure. In contrast, structural simplification applies to phonological changes in the syllable or word structure (Gruwell, 1981). Phonetic distortions patterns include distortion of phones that do not affect meaning.

Table 6.2 *Types, age, and percentages of occurrence of developmental phonological patterns presented in at least 10% of children in each age group (cut-off criteria ≥4 or ≥6)*

Pattern / Age group		2;6–2;11	3;0–3;5	3;6–3;11	4;0–4;5	4;6–4;11	5;0–5;5	5;6–5;11
		n= 30	n=27	n=38	n=39	n=31	n=37	n=33
Systemic simplifications								
Fronting of /ʃ ʒ/→ [s z θ ð]	≥4	60*	70	61	41	35	27	36
/ʃa:.riʕ/ → [sa:.riʕ] **(street)**	≥6	47	63	58	38	23	16	27
Fronting of /s z/ → [θ ð]	≥4	27	26	21	18	29	16	15
/ze:t/→ [ðe:t] **(oil)**	≥6	27	15	16	15	23	14	
Fronting of /k g/→ [t d]	≥4	17						
/kalb/→ [talb] **(dog)**	≥6	17						
Backing /x ɣ/→ [ħ ʕ]	≥4	30	26	26	13			
/xas/→ [ħas] **(lettuce)**	≥6	10	26	21	10			
Stopping of fricatives	≥4	57	48	61	28			
/sa.mak/→ [ta.mak] **(fish)**	≥6	40	37	34	15			
Lateralisation /r/→ [l]	≥4	80	56	42	41	19	19	
/ra:s/→ [la:s] **(head)**	≥6	70	52	37	26	16	14	
Glottal stop replacement to /ʔ/	≥4	77	48	45	36	10		
/nax.la/→ [ʔax.la] **(palm tree)**	≥6	50	37	37	15			
Glottal replacement to /h/	≥4	47	30	24				
/ħa.la:.wa/→ [ha.la:.wa] **(candy)**	≥6	40	22	13				
De–emphasisation	≥4	100	93	79	51	45	19	
/ tˤɑˈma:tˤim /→ [ta.ma:tim] **(tomato)**	≥6	100	81	63	38	35	16	
Denasalisation	≥4	13						
/mak.wa/→[bak.wa] **(iron)**	≥6							
Metathesis	≥4	77	52	50	21	10		
/ʔib.ti.da:.ʔi:/→[daʔibˈda:ʔi:] **(primary school)**	≥6	63	41	21	13			
Assimilation	≥4	87	81	76	49	13	19	27
/ʔasˤ.far/→ [ʔar.far] **(yellow)**	≥6	73	52	55	15		11	
Gemination of heterosyllabic CC	≥4	50	63	26	10			
/ʕan.ka.bu:t/→[ʕak.ka.bu:t] **(spider)**	≥6	33	30	16				
Devoicing	≥4	77	70	63	44	29	16	24
/ruz/→[rus] **(rice)**	≥6	70	63	39	28	23	11	15
Voicing	≥4	23	19	13				
/ta.mur/→[da.mur] **(date)**	≥6	10						
/dˤ/→ [ðˤ]	≥4		15	11	15	16	30	33
/be:dˤ/→[be:ðˤ] **(egg)**	≥6				10		22	
Vowel substitution (e.g. /a,i/→[ɛ], /o:/→[o]) /mi.gasˤ/→[mɛgasˤ] **(scissor)**	≥4	100	89	87	72	65	54	58
	≥6	97	89	79	49	42	41	33

Vowel deletion	≥4	13	11					
/ˈminʃafa/→[minʃaf] **(towel)**	≥6							
Structural simplifications								
Weak syllable deletion	≥4	87	52	47	23			
/mu.kaj.jif/→[kaj.jif] **(air conditioner)**	≥6	67	41	29	10			
Heterosyllabic CC reduction	≥4	70	52	32				
/kur.siː/ → [ku.siː] **(chair)**	≥6	53	37	16				
Word-final CC reduction	≥4	30	19	13				
/gird/→ [gid] **(monkey)**	≥6							
Word-final consonants deletion	≥4	53	52	32	31	10		
/ni.mir/→ [ni.mi] **(tiger)**	≥6	33	41	16	13			
Word-initial consonant deletion	≥4	33	33	11				
/ħi.sˤaːn/→ [isˤaːn**]** **(horse)**	≥6	20	15					
Reduplication	≥4	40	15					
/ħa.la:.wa /→[wawa] **(candy)**	≥6	13						
Phonetic distortions								
Distortion of sibilants (s z ʃ ʒ) **/s/**→[sˡ], [sʲ], /z/→[zˡ], [zʲ]	≥4	97	89	89	54	61	59	52
/ʃ/→ [ʃʲ] ,[ʃˡ] or [ç], /ʒ/→[ʒˡ] [ʒ] or [ʝ] e.g. /minʃafa/→ [minçafa] **(towel)**	≥6	90	78	82	28	28	49	42
Distortions of /r/→[ɹ] or [ɻ]	≥4	70	33	39	54	45	27	
/riʒil/→ [ɹiʒil] **(leg)**	≥6	63	22	21	41	35	24	
Distortions of /l/→[l̪] or [ɫ]	≥4	43	11					
/laiːmuːn/→ [l̪aiːmuːn] **(lemon)**	≥6	23						

Key: C: consonants, **WW**: within word, * all numbers are percentages.

The results revealed that, in general, 27 patterns were observed within SHA-speaking children aged 2;6–5;11: 18 patterns of systemic simplification, six structural simplifications, and three phonetic error patterns.

Out of the 27 patterns, three simplification phonological patterns occurred under both criteria across all age groups: fronting of /ʃ ʒ/→ [s z θ ð], devoicing, and vowel substitution. The number of occurrences steady decreased with increasing age, but the numbers in the oldest age–group indicate a later age of overcoming these patterns than was assessed in this study.

Fourteen patterns with individual age of acquisition across the different age groups were observed under both cut-off criteria (≥4 and ≥6): fronting of /k g/→ [t d], backing /x ɣ/→ [ħ ʕ], stopping of fricatives, lateralisation /r/→ [l], glottal stop replacement to /ʔ/, glottal replacement to /h/, de-emphasisation, metathesis, assimilation, gemination of heterosyllabic CC , weak syllable deletion, heterosyllabic CC reduction, word-final consonant deletion, word-initial consonant deletions.

Four other patterns only occurred under the ≥4 criterion: voicing, word-final CC reduction, reduplication, /dˤ/→ [ðˤ]. The occurrences of these patterns under the lower criterion may indicate that they are not developmental, considering the large number of words assessed. The occurrence of the substitution of /dˤ/→[ðˤ] was mainly observed between the ages of 4;0, and 5;5 years under the higher criterion (≥6). Under the lower criterion (≥4), the substituting of /dˤ/→[ðˤ] pattern showed an increasing trend in its appearance, with the highest occurrence at the age 5;6–5;11, which might be related to the acquisition of standard Arabic.

Another two patterns were observed under the lower criterion only: denasalisation and vowel deletion. Their occurrences may also be considered as non–developmental patterns because they occurred only under the lower criterion (≥4) rather than occurring under both criteria, and they were observed at low frequency.

Three phonetic patterns were observed: distortion of sibilants /s z ʃ ʒ/, distortion of /r/→[ɹ] or [ɻ], and distortion of /l/→[ɭ] or [ɫ] with one distortion pattern (i.e. distortion of /l/→[ɭ] or [ɫ]) only occurring at a low percentage under both criteria in the youngest age group.

6.1.3 Summary of results III

The phonological variations in SHA-speaking children analysis result could be summarized in the following points:

- SHA-speaking children showed a high number of PhonVar (Tokens) between the age of 2;6-5;11, which decreased with age.
- The number of InfrVar in each age group was higher than the number of phonological patterns (Type) identified with both criteria (≥4 and ≥6).
- In terms of phonological patterns (Types), there were 27 patterns identified in the speech of SHA-speaking children aged 2;6-5;11:
  - 18 systemic simplification patterns.
  - Six structural simplification patterns.
  - Three phonetic distortion patterns.
- The influence of cut-off criteria on the identification of patterns was apparent in this data, as there are several patterns occurred when both criteria were applied (i.e. a total of 17 patterns), and others occurred when the lower cut-off was applied, either at a high frequency (i.e. voicing, word-final CC reduction, reduplication, /dˤ/→ [ðˤ]) or with low frequency (i.e. denasalization and vowel deletion).

6.2 Discussion III – Phonological variants

One aim of this study was to investigate the typical phonetic and phonological development in 235 SHA-speaking children aged 2;6–5;11. In this section, the results for the number of the phonological variants (Tokens, Types and InfrVar) found in the speech of SHA-speaking children are discussed first in terms of quantitative measures (i.e. Token and InfrVar), followed by a discussion of the qualitative measure (Types) of phonological variants, including a description of the developmental patterns identified in the speech of these children, as well as identifying the language-universal, language-specific, and SHA-specific patterns.

6.2.1 Quantitative measures of the phonological variants in SHA-speaking children

As reported in Section 6.1.1, the phonological variants per age group were investigated utilising the total number of variants per child (Token), and the number of infrequent variants (InfrVar) as quantitative measurements (see Chapter 3, Section 3.6.3). The results revealed that SHA-speaking children expressed many Tokens and InfrVar. The mean number of Tokens and InfrVar showed a steady decrease with age, indicating a stabilisation of the phonological system over time. However, the standard deviations (SDs) of both measures (Tokens and InfrVar) were high and displayed fluctuation in values across all age groups, showing a large variability in phonological development among SHA-speaking children.

Looking at the number of Tokens for each age group, it is evident that SHA-speaking children showed a large number of phonological variants between the age of 2;6 and 5;11. However, it is important to remember that the number of Token includes all the occurrences of phonologically varied speech in children's productions, allowing for an analysis of the degree of inaccuracy in children’s speech based on the whole word, as opposed to the purely segmental perspective used in some other Arabic studies (cf. Saleh et al., 2007). Therefore, it was not possible to compare the Token number with those found in other Arabic studies, as they usually reported the total number of phonological patterns identified (e.g. Abou–Elsaad et al., 2019), or the total number of occurrences of each pattern occur (e.g. Alqattan, 2015). Abou–Elsaad et al. (2019) found that the 120 Egyptian Arabic-speaking children demonstrated a total number of 225 phonological patterns between the age of 2;0 and 5;0. Alqattan (2015) found that the number of errors of stopping pattern, for example, among her sample was 2255 errors out of 17,649 words; as she collected a spontaneous free speech from 80 Kuwaiti Arabic-speaking children aged between 1;4 and 3;7. Most of the previous Arabic studies only reported the number of occurrences of a particular set of phonological patterns without considering the total number of Token produced by their sample. However, some of these studies used the occurrence of one token as their cut-off criterion (e.g. Alqattan, 2015), meaning that they did not differentiate between the phonological patterns and misarticulations that occur by chance in children's speech (McReynolds &Elbert, 1981). Thus, their reported number of total numbers of phonological patterns could be inaccurate as it could include a number of phonological variants that did not reflect the presence of a pattern, but which were instead articulatory errors (McReynolds & Elbert, 1981).

Moreover, the test used in this study, SHAPA, included many words with multisyllabic words, which could lead to more mismatches in children's speech as reported by James et al. (2016). They found that most developmental mismatches were in words with multisyllabic words, suggesting that the residual development between the ages of 5;0 and 7;0 is mainly associated with syntagmatic aspect of speech as multisyllabic words are usually contained of one or more grammatical components of a language (as in UHA adding the plural suffix caused in increasing the number of syllables in that word, see Chapter 2, Section 2.1.3.4). This indicates that children are still learning about the complicated interaction of segmental and prosodic elements like syllable sonority and stress, rather than the pragmatic properties of particular sounds, which is the phonemic integrity. Such complex features of test items may result in identifying more phonological variants such as metathesis and gemination of heterosyllabic words, which could increase the total number of Tokens, Types and InfrVar for SHA-speaking children in this study.

Compared to other languages, SHA-speaking children produced a higher number of Token compared to those found in children speaking other languages (e.g. Italian). Tokens were calculated by Fox-Boyer et al. (2021) from 183 typically developing Italian–speaking children aged 3;0–4;11. They found that, for example, the age group 3;0–3;5 had a total of 55.21 Tokens, compared to 163.74 Tokens found for the same age group in the current study. Similarly, Albrecht (2017) found that the bilingual German-Turkish-speaking children aged 3;0-3;5 had 98.08 Tokens in German and 87.75 Tokens in Turkish. Considering these two studies, it is important to acknowledge the number of items elicited from their samples, which were in general much lower than the number of words elicited from children in the current study (i.e. 151 words). In the Italian study, their test included only 77 black and white pictures (Fox-Boyer et al., 2021), and in Albrecht's study there were 70 items in the Turkish test and 96 items in the German test (Albrecht, 2017). Such a methodological difference may illustrate the variants between the result of these studies results and those found for SHA-speaking children in the current study, where more items were named which may have led to more errors in these children's speech.

Another possible explanation can be provided for the high number of Tokens elicited in this study. The influence of phonotactics on the structures of Arabic words and the impact of the morphological acquisition, which reflects the grammar complexity of the Arabic language. Arabic words are like those from Maltese, in that they can be composed of a series of morphemes (Grech,2006), for example, the possession (/kalb/ (CVCC) 'dog'→ /kalbi/ (CVC.CV) 'my dog'), the feminine (/ʃaʔar/ (CV.CVC) 'hair'→ /ʃaʔraha/ (CVC.CV.CV) 'her hair'), and some plural and singular forms, which may be expressed through suffixes. Consequently, the syllabic complexity and the number of syllables increase as more morphemes are included in a word. The structure of such words may increase the phonotactic complexity of a language. Therefore, the progress in lexical and morphological development adds demands to the phonological development of children, which influences their ability to achieve phonological mastery of their language (Grech, 2006). The findings of the current study include some examples related to this issue, where the children made errors that could be explained by phonotactic structures, such as adding the singular feminine suffix /a/ to nouns or verbs that did not allow for such a suffix (e.g. /tˤaħ/ 'he fell'→ [tˤaħa], /ʃaːriʕ/ 'street'→ [ʃaːriʕa]). Such errors did not reach any of the cut-off criteria applied in this study, and all were counted as InfrVar, which increased the Tokens number consequently.

The results of the InfrVar calculation in this study also revealed high numbers under both cut-off criteria, considering that InfrVar were calculated with considering the cut-off criteria, unlike the Tokens. InfrVar accounted for more than half of the Tokens of the first three age groups (2;6–2;11, 3;0–3;5, and 3;6–3;11), indicating that the phonological patterns demonstrated by children in these three age groups were much more frequent than InfrVar. Furthermore, for children aged ≥4;0, the InfrVar were almost half the number of their Tokens, showing that most of the Tokens of children older than 4;0 was InfrVar. This highlighted the vital role of InfrVar in phonological development as they could signify the stability of children's phonological system (Fox-Boyer et al., 2021). The reduction of InfrVar across the age groups indicated that the consistency of SHA-speaking children increased with age (Dodd, 2005). As reported earlier, previous Arabic studies used to ignore any phonological pattern that occurred at a lower frequency than their cut-off criterion. Thus, they never calculated the number of Tokens or InfrVar among their samples. However, McReynolds and Elbert (1981) asserted that such variants should not be ignored and that they should be distinguished from other actual phonological patterns. Failure to differentiate between the errors occurring by chance and the errors that represent a pattern in children's speech could lead to misleading information about the children's articulation profile and the labelling of some errors as 'developmental patterns', while they are less common if a quantitative cut-off criterion is applied (McReynolds & Elbert, 1981).

Such a high number of InfrVar seems to reflect the variants within SHA-speaking children's speech, which cloud have occurred as a result of unfamiliar words included in SHAPA that caused children to mispronounce these unfamiliar items (Stackhouse & Wells, 1997). This explanation could lead to another reason for the increasing number of Tokens and InfrVar among children in this study. As reported earlier (see Chapter 4, Section 4.4.2.1), questions were raised regarding the familiarity of the

SHAPA test items, and the results of the PhonVar analysis reflected the effect of this low familiarity by showing a high number of Tokens and InfrVar. Given the different naming word-forms for SHAPA items (see Chapter 3, Section 3.3.1.1) due to the exposure to MSA and other dialects, some items form selected in SHAPA may not familiar to some SHA-speaking children, which make it unclear how stable the phonological representations and motor programs the items selected for SHAPA (Macrae, Tyler, & Lewis, 2014). It is possible that children had to rely more on their skills to create new motor programmes for unfamiliar words than activating their existing motor programs for known words (Stackhouse & Wells, 1997). The results of the SHAPA evaluation were in line with the results of the InfrVar analysis where the spontaneous naming rate was lower than that reported in the literature. This could have an effect, along with the previously mentioned reasons, on increasing InfrVar in SHA-speaking children's speech.

In addition, children need to master other aspects of language (e.g. prosodic features, syntactic structure) in order to be intelligible and communicate effectively (Kent, Miolo, & Bloedel, 1994). During phonological development, children rapidly acquire other dimensions of language, adding vocabulary to their language system and building longer sentences to increase their intelligibility (Davis & Bedore, 2013). Considering the linguistic complexity of their ambient language, Davis and Bedore (2013) explained that as children learn the intelligible production and perception of salient language forms, this process of learning may be accompanied by a decline in related behaviours as children shift their effort towards a more complex phonological target within more complex linguistic contexts. Therefore, the process of phonological acquisition could be affected either negatively or positively during the acquisition of these other aspects of language. Such a condition may lead some children to produce more phonological patterns in their speech while acquiring other aspects of their language (e.g. morphological acquisition), whereas other children may develop an advanced phonological system during the acquisition of other language features (e.g. lexical development). This hypothesis is in agreement with Slobin's (1973, as cited in Clausen & Fox–Boyer, 2017) assumptions about the influence of morphological acquisition and the complexity of a language domain on phonological development. Furthermore, it supports Stoel-Gammon (2011), who found that phonological development affects lexical acquisition to a greater degree than lexical factors affect phonological development; however, she asserted that the influence is bidirectional. Such a hypothesis could explain the high number of Tokens and InfrVar among SHA-speaking children as they are exposed to a complex linguistic system of Arabic including complicated phonological and morphological features. There is minimal research investigating the relationship between the acquisition of phonology and other aspects of language and how the acquisition of one aspect could enhance or inhibit the acquisition of another. Therefore, more research is needed to investigate such a relationship and to understand how the phonology develops in comparison to other aspects of language (morphology, syntax, semantics, and pragmatics).

In short, the high number of Tokens and InfrVar found in the speech of SHA-speaking children could highlight the importance of conducting further studies, including evaluating other linguistic skills in SHA-speaking children, editing the SHAPA test and increasing its items familiarity scores, and examining the effect of lexical development and other aspects of language on the phonological development.

6.2.2 Qualitative measures of phonological variants in SHA-speaking children

Another aim of this study was to investigate the Types of phonetic and phonological patterns in 235 SHA-speaking children aged 2;6–5;11 and their age of occurrence. The investigation included counting the number and kind of phonological patterns (Types) as reoccurring variants of the same kind of speech error within the same child for at least four (lower cut-off criterion) or six times (higher cut-off criterion). In terms of the phonological patterns, the results revealed 27 phonetic–phonological patterns in the speech of SHA-speaking children, which can be separated into systemic simplifications (n=18) and structural simplifications patterns (n=6), as well as three examples of phonetic distortions (see Table 6.2). Some of these patterns are considered as developmental for SHA-speaking children, while others are not, as discussed below. However, it is important to note that the total number of Types identified in this study is quite large. Using a precise description of each pattern, such as specifying the exact sound affected by each pattern, could lead to identifying more patterns within the children's speech in the study sample. In the current study, some common patterns (e.g. fronting and glottal replacement) were split in to more than one pattern (for fronting: fronting of /ʃ ʒ/, fronting of /s z/, and fronting of /k g/, and for glottal replacement: glottal stop replacement to /ʔ/ and glottal replacement to /h/); therefore, more patterns were identified, which could have increased the total number of phonological pattern Types.

In terms of the phonological patterns, three were found independent of the cut-off criteria across all age groups at a high frequency that showed a decreasing percentage of occurrence with age (i.e. fronting of /ʃ, ʒ/→ [s z θ ð], devoicing, and vowel substitution). Similarly, fronting of /s, z/□[θ, ð] is another pattern that occurred independent of the cut-off criteria but was overcome by age 5;5. This indicates that these patterns are developmental for SHA-speaking children between the age of 2;6 and 5;11, because they appeared at a high frequency under both cut-off criteria. Thus, their occurrences under both cut-off criteria supports the empirical evidence in favour of them being considered as patterns (McReynolds & Elbert, 1981). For the devoicing patterns, the percentage of occurrence fell to a low level (15% for the higher cut-off criterion) in the oldest age group. This could suggest that this pattern would not appear as typical development pattern in children older than 5;11 because they could stay below the threshold of >10% of children (e.g. Dodd et al., 2003). Data on older children are needed to confirm the age of disappearance for all three patterns. Further discussion of these patterns is presented below (see Section 6.2.2.1).

Fourteen patterns were observed with an individual age of acquisition for both criteria (≥4 and ≥6); these are: fronting /k, ɡ/→[t, d], backing /x ɣ/→ [ħ ʕ], stopping of fricatives, lateralisation of /r/→[l], glottal stop replacement to /ʔ/, glottal replacement to /h/, de–emphasisation, metathesis, assimilation, gemination of heterosyllabic CC, weak syllable deletion, heterosyllabic CC reduction, word-final consonant deletion, and word-initial consonant deletion. The occurrences of these patterns at different ages but under both cut-off criteria indicates that they are developmental patterns for SHA-speaking children between the age of 2;6 and 5;11. For the patterns observed in only one or two of the youngest age groups (e.g. fronting /k ɡ/, word-initial consonant deletion), their early age of suppression could indicate that these patterns are more frequent in younger children, which is in line with what was found for Jordanian Arabic-speaking children, where these two patterns were common in children younger than 2;6 (Dyson & Amayreh, 2000; Mashaqba et al., 2019). This finding agrees with Fox-Boyer et al. (2021), who noted the importance of including the age at which children start to be consistent in their speech production and show identifiable patterns (i.e. 2;0–2;6) when studying and analysing phonological development.

Another four patterns occurred under the lower criterion only (≥4) (i.e. voicing, the replacement of /dˤ/□[ðˤ], word-final CC reduction, reduplication). Furthermore, two other patterns were also occurred under the lower cut-off criterion with low frequency (i.e. denasalisation and vowel deletion). These two patterns appeared only in the speech of the youngest age group/groups (i.e. 2;6–2;11 [dentalisation and vowel deletion] and 3;0–3;5 [vowel deletion]). Due to the large number of words in the SHAPA test, the occurrences of these patterns under the lower criterion only could indicate that they cannot be considered as developmental patterns and that their appearance in the speech of SHA-speaking children was by chance due to the application of the low cut-off criterion. In line with Kirk and Vigeland's (2015) suggestion that a cut-off of at least four occurrences might be too low for some types of error, these patterns might have occurred by chance due to the application of an unsuitable cut-off criterion for these specific patterns, especially with the long SHAPA test where there are multiple opportunities for these patterns to appear in the speech of SHA-speaking children.

Finally, three phonetic patterns occurred among different age groups under both cut-off criteria. These phonetic distortions were distortion of sibilants /s, z, sˤ, ʃ/ and /ʒ/; distortion of r/→[ɹ] or [ɻ]; and distortion of /l/→[ɭ] or [ɫ]. These findings are in partial agreement with Arabic studies conducted by Dyson and Amayreh (2000) and Owaida (2015), who also reported that the distortion of sibilants, /s/ and /z/ in particular, was common among almost all age groups in their samples. Deviant of /r/ is a pattern that reported by many other Arabic studies (Al-Bader, 2009; Ammar & Morsi, 2006; Ayyad, 2011), where its description included realisation of the trill /r/ as the retroflex /ɻ/ or alveolar /ɹ/, and it continues to occur until age 5;0 for children speaking Egyptian and Saudi Najdi Arabic, which is similar to the description of /r/ distortion in the current study that disappeared by age 5;5. Beyond the Arabic language, the distortion of sibilants has also been found in other languages such as English (Smit, 1993), Maltese (Grech, 2006), Danish (Clausen & Fox-Boyer, 2017) and German (Fox, 2006). Interestingly, distortion of /l/→[ɭ] or [ɫ] has never been reported for other Arabic dialects, which might mean they have either not been accounted for or have not occurred. All these phonetic patterns are considered as developmental for SHA-speaking children aged 2;6–5;11.

In general, out of the 27 phonetic and phonological patterns identified in the speech of SHA-speaking children in this study, there are 18 developmental phonological patterns (i.e. fronting of /ʃ, ʒ/□ [s z θ ð], devoicing, vowel substitution, fronting of /s, z/□[θ, ð], fronting /k, g/→[t, d], backing/x ɣ/→ [ħ ʕ], stopping of fricatives, lateralisation of /r/→[l], glottal stop replacement to /ʔ/, glottal replacement to /h/, de–emphasisation, metathesis, assimilation, gemination of heterosyllabic CC, weak syllable deletion, heterosyllabic CC reduction, word-final consonant deletion, and word-initial consonant deletion), three developmental phonetic patterns (i.e. distortion of sibilants /s, z, sˤ, ʃ/ and /ʒ/, distortion of /r/→[ɹ] or [ɻ] and distortion of /l/→[ɭ] or [ɫ]). Six of the identified patterns in this study cannot be considered as developmental due to their low occurrences as they occurred under the lower criterion ≥4, these patterns are: voicing, replacement of /dˤ/→[ðˤ], word-final CC reduction, reduplication, dentalisation and vowel deletion. Considering the two cut-off criteria in this study, such a result highlights the influence of applying these two criteria as an innovative method to distinguish real developmental patterns from patterns that could occur by chance findings and should be considered as InfrVar.

In the following subsection, the different types of developmental and non–developmental patterns identified in the speech of SHA-speaking children are discussed and described according to language-universal, language-specific, and SHA-specific patterns, as well as being separated into systemic simplification and structural simplification patterns.

6.2.2.1 Types of phonetic and phonological patterns

The phonological patterns that classified as 'developmental' in the above section, these patterns can be separated into systemic simplification (n=14) and as structural simplification (n=4). Eight types of systemic simplifications and two types of structural simplifications have been reported as language universal. Five systemic patterns and two structural patterns can be classified as language specific since they reparteed to be typical for children speaking other Arabic dialects. Finally, one developmental pattern that can be descried as SHA-language specific is the systemic simplification pattern 'glottal replacement to /h/'. Figure 6.1 below summarized all the patterns identified in the speech of SHA-speaking children and their classification according to developmental and non-developmental. Furthermore, following subsections discuss these patterns according to their classifications.

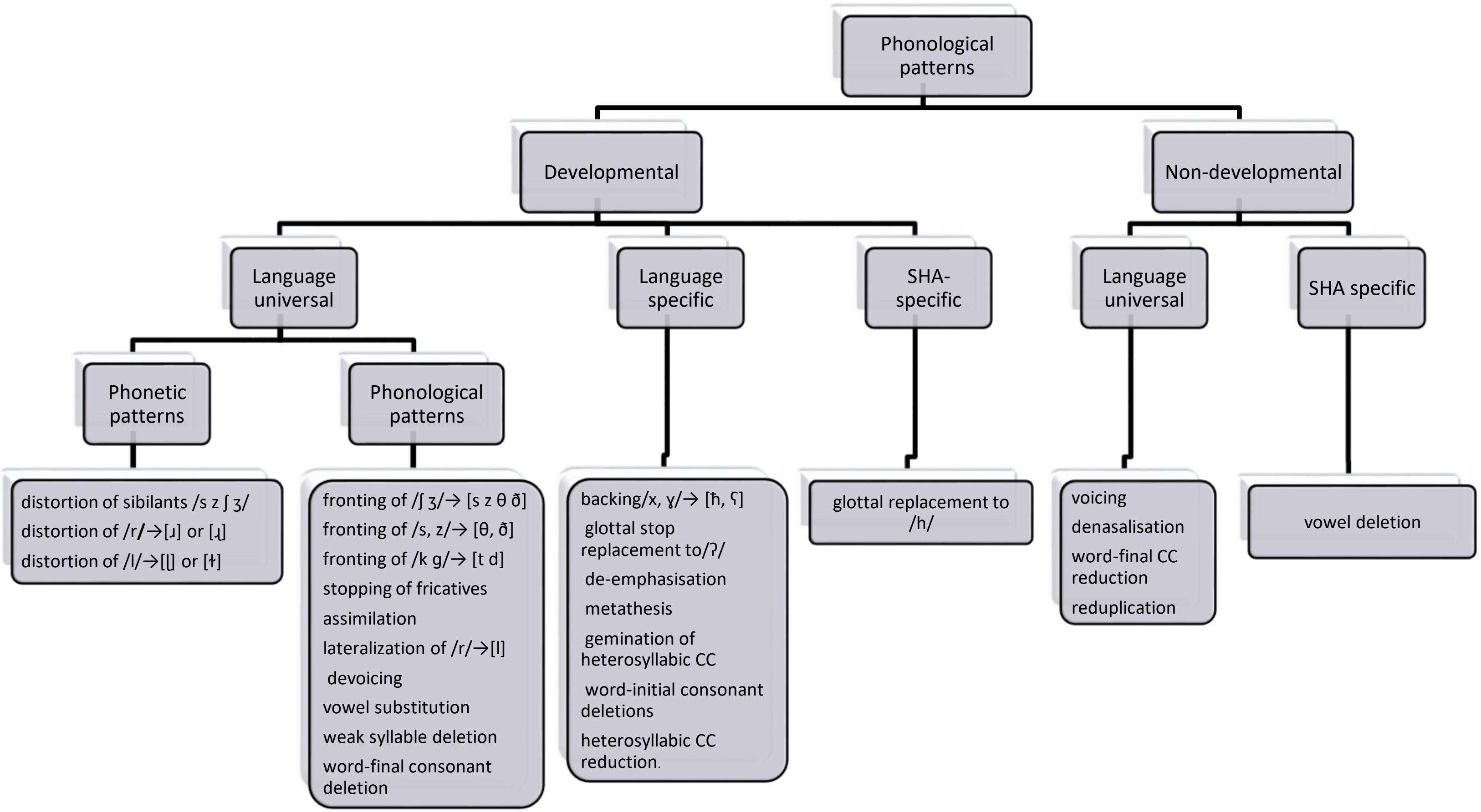


Figure 6.1 *A summary of phonological patterns identified in the speech of SHA-speaking children between 2;6-5;11*

6.2.2.1.1 Developmental language universal patterns

Ten of the phonological patterns identified in this study have been reported as *language universal* (Dodd & Lacano, 1989; Fox-Boyer et al., 2021; Gruwell, 1981) (i.e. fronting of /ʃ ʒ/→ [s z θ ð], fronting of /s, z/→ [θ, ð], fronting of /k g/→ [t d], stopping of fricatives, assimilation, lateralization of /r/→[l], devoicing, vowel substitution, weak syllable deletion and word-final consonant deletion). Further, three phonetic patterns (i.e. distortion of sibilants /s z ʃ ʒ/, distortion of /r/→[ɹ] or [ɻ], and distortion of /l/→[ɭ] or [ɫ]) have frequently been reported to occur in children's speech cross linguistically (McLeod, 2007). Thus, the occurrence of these phonological patterns in the speech of SHA-speaking children may indicate that children across languages apply the same simplification techniques to deal with complex phonological systems.

Fronting of /s, z/→ [θ ð] is one of the language universal patterns that occurs across Arabic dialects; however, it had been reported under the 'dentalisation' pattern, and it considered as a phonetic error or distortion (Dyson & Amayreh, 2000; Owaida, 2015). According to Dyson and Amayreh (2000), the sibilants /s, z/ are usually distorted because the usual place of articulation for these sibilants in Arabic is somewhat dental in many adult Jordanian speakers. For SHA-speaking children, this is one of the patterns where it is questionable whether it should be considered as a phonological or phonetic pattern, similar to that reported for Maltese-speaking children by Fox (2006). Although the dental sounds /θ ð/ were not part of the phonetic system of Maltese, she explained that the exposure of the Maltese-speaking children to English, even if they are not truly bilingual speakers, results in these sounds becoming part of the children's phonetic and phonemic inventory, which could lead to phonological confusion. In the case of SHA-speaking children, the dental sounds [θ and ð] are not UHA phones but are MSA phones, as well as are occurring in other Saudi dialects (e.g. Saudi Najdi Arabic). Because SHA-speaking children are continuously exposed to other Saudi dialects, this could explain why they tend to substitute /s, z/ with [θ, ð]. Interestingly, replacing /s, z/ with [θ ð] could lead to forming two different words, one of which is valid in all UHA, other dialects and MSA and the other which is valid only in MSA (e.g. /samar/ 'stay up at night'→[θamar] 'fruit'). However, the way in which children in this study used the two phones /θ, ð/ would not be correct or acceptable in any forms (e.g. /zeːt/→[ðeːt] 'oil'). Thus, such a substitution could be considered as phonetic in nature (i.e. a lisp, which is a phonetic error); or it could also be a product of phonological unsureness or confusion. Such a confusion may lead children to front /s z/ to /θ ð/, and this pattern continues to occur until the age of 5;5 when the higher criterion is applied. Therefore, it could be considered as a developmental pattern for SHA-speaking children. This is unlike the case for German-speaking children, where these two dental sounds are not part of the German phonetic system and the fronting of /s z/→ [θ ð] among German-speaking children does not affect the phonemic contrasts. Therefore, Fox (2006) considered it as articulatory substitution and not as phonological pattern. In contrast, Smit et al. (1990) reported that English-speaking children showed a dentalisation error (i.e. /s z/→ [θ ð]) until the age of 6;0 and such an error is considered a phonological pattern as both phones are in exitance in English, and such a substitution may affect the meaning. Similarly, James (2001b) considered dentalisation as phonological pattern, where its occurrence was common among Australian English–speaking children, increasing with age and peaking at age 4;0, and then decreasing and not used after the age of 6;0.

Lateralisation of /r/ is not a common pattern among English-speaking children, for whom gliding (i.e. replacing /r/ to [w] or [j]) is the common pattern for alveolar /ɹ/ deviation. This lateralisation pattern is commonly reported as a typical developmental pattern for many languages, such as Spanish (Goldstein, 2005), Maltese (Grech, 2006) and German (Fox, 2000), as well as for Arabic (e.g. Jordanian, Egyptian and Syrian dialects). Dyson and Amayreh (2000) noticed that in these languages, where the lateralisation of /r/ is common, include tap and trill /r/. These variants share the alveolar place of articulation with /l/, while English /ɹ/ is an approximant that involves either retroflexion or bunching of the back of the tongue with lip rounding without alveolar contact. This difference in /r/ articulation between English and Arabic explains the different realisations.

Vowel substitution is another language-universal pattern found in the speech of SHA-speaking children, with a high frequency of occurrence across all age groups under both criteria. However, Fox (2006) argued that it could not be considered universal as vowel errors are usually a mark of phonological disorders in monolingual-speaking children because of the universal acceptance that vowel development is completed before the age of 3;0 (Smit et al., 1990). However, for bilingual children, it has been proven that vowel changes or patterns of variants are developmental patterns due to the influence of the two different vowel systems in the languages they speak (Fox–Boyer, Fricke, & Albrecht, 2020; Pascoe, Mahura, & Le Roux, 2018). However, this explanation cannot be applied to the vowel changes found in the current study which considered only errors in vowel changes, with all dialectal changes counted as correct (see Chapter 3, Section 3.6.2). Examples of errors identified among the current study sample include reduced contrasts between long and short vowels (e.g. /doːlaːb/ ‘closet’ produced as [dolaːb]), using fewer front vowels (e.g. /migasˤ/→ [mɛgasˤ]), or vowel harmony (e.g. /ʒizam/→[ʒizim] ‘shoes’). However, the researcher did not analyse the vowel substitution pattern into different types; thus, the exact types of vowel substitution demonstrated by SHA-speaking children are unknown. Further, it cannot be assumed that vowel changes have been influenced by dialectal variants of Arabic (UHA and MSA) (cf. Fox-Boyer et al., 2021; Levy & Hanulíková, 2019). The types of error observed among children in this study did not reflect the influence of exposure to MSA, whose vowel system differs from UHA only in the usage of diphthongs and their replacement by /eː/, /oː/ (cf. Pascoe, Mahura, & Le Roux, 2018). However, the results of this study are in agreement with James (2001b), who found that vowel changes pattern was common among all age groups of her sample of monolingual Australian English-speaking children and persisted to the age of 7;0. Although she did not systematically analyse the types of vowel changes, James concluded that this vowel changes might be dialectal, but may also reflect developmental phenomena, where children are still learning how to control vowels productions in unstressed syllables.

Other Arabic studies never reported vowel change patterns as they rarely investigated vowels development during their phonological pattern analysis. This is a universal behaviour as it is expected that children over 2;0 make few vowel errors, and it is also believed that vowels vary according to regional accent (Reynolds, 1990, as cited in Dodd, 1995). However, one study conducted by Mashaqba et al. (2019) reported vowel shortening and lengthening as a pattern that occurs in the speech of young Jordanian Arabic-speaking children aged 1;0–1;6, where they used to either lengthening or shortening the vowel to compensate for the deletion of a consonant. Mashaqba and his colleagues did not notice the occurrence of this pattern in children older than 1;6, which contrasts with the results of the current study, where errors in vowels continued to occur until the age of 5;11. However, and as reported earlier, the types of errors in the current study were not analysed in detail. Therefore, it is recommended for future research to systematically analyse the types of errors that occurred in the vowels in the present study to understand the nature of this pattern and to examine the influence of MSA and other Arabic dialects on vowel acquisition among SHA-speaking children.

Weak syllable deletion was described in some Arabic studies without specifying the type of the syllable deleted (e.g. Abou-Elsaad et al., 2019; Dyson & Amayreh, 2000); these studies described a ‘syllable deletion pattern’ without determining whether the deleted syllable was a weak or unstressed syllable, which limits the comparisons that can be drawn between this data and the findings of the current study. Similarly, word-final consonant deletion was reported in some Arabic studies (e.g. Alqattan, 2015) as coda deletion. Such a label could refer to either deletion of the consonant in the syllable-final position within word, or in syllable-final-word-final position, which again limits the scope of comparisons with the current study.

In general, these patterns are considered as language-universal because they are commonly reported as developmental patterns in the speech of children speaking different languages (e.g. English, German, Italian, Maltese, as well as Arabic).

The three distortion phonetic patterns identified in this study are considered as language universal because these three types of distortion have been frequently reported in other languages such as English (Gruber, 1999; Shriberg, 1993) and Brazilian-Portuguese (Wertzner et al., 2005). Regarding distortions found in the other Arabic dialects, distortion of /s, z, sˤ/ was reported to be the most common distortion in the speech of Jordanian Arabic-speaking children (Dyson & Amayreh, 2000). The results of the current study confirm Dyson and Amayreh’s findings; however, the distortion of /ʃ/ and /ʒ/ is added as another frequently occurring type. Further, distortion of /r/, which was labelled as deviation of /r/ in many previous studies and was also frequently reported as a developmental phonological pattern (e.g. Al-Bader, 2009; Ammar & Morsi, 2006; Ayyad, 2011), was found to be a developmental phonetic pattern for SHA-speaking children aged 2;6–5;5. In some Arabic studies /r/ deviation, or distortion of /r/, was classified as a phonological pattern. However, the other form of /r/ used by children in this pattern did not change the meaning of the words and included substitution of trill/flap /r/ with retroflex /ɹ/, which is not an acceptable allophonic alteration for trill/flap /r/. Thus, in this study such an error was classified as a phonetic distortion. Distortion of /l/ has never been reported in any Arabic studies; however, it has been reported as a common distortion type in English and Brazilian-Portuguese-speaking children aged 3;0 to 5;0 (Gruber, 1999; Shriberg, 1993; Wertzner et al., 2005).

6.2.2.1.2 Developmental language-specific patterns

Seven other developmental patterns cannot be considered as language universal, but rather as *language specific* patterns because they have been reported as typical for other Arabic dialects but never or rarely reported as typical phonological patterns in other languages. These patterns are backing/x, ɣ/→ [ħ, ʕ], glottal stop replacement to /ʔ/, de–emphasisation, metathesis, gemination of heterosyllabic CC, word-initial consonant deletions, and heterosyllabic CC reduction. The influence of the ambient language phonology (e.g. motor difficulties, syllable structures, and phonological salience) could explain the occurrence of these patterns, and language universals cannot explain these types of patterns produced by SHA-speaking children (Dyson & Amayreh, 2000).

De-emphasisation is the only pattern that occurs across all Arabic dialects, and it has never been reported as occurring in any other language because emphatic consonants are specific to Arabic and its all forms. De-emphasisation has been found to occur at a high frequency in SHA-speaking and all other Arabic dialects; in the current study, it declined with age and disappeared at the age of 5;5 in line with the findings of other Arabic studies.

However, some of these Arabic-specific patterns have been reported as being idiosyncratic in some other languages. For example, backing and metathesis were reported to be unusual pattern for English- speaking children (Dodd et al., 2003; Roberts, Burchinal, & Footo, 1990), while glottal stop replacement was reported to as an acceptable variant for English-speaking children if it replaced for /t/ (Smit, 1993). This contrasts with the results of the current study, where glottal stop replacement is considered as a developmental pattern regardless of the sound replaced by the glottal stop /ʔ/. Further forms of glottal replacement were considered as non-developmental for English as well as for German- speaking children (Dodd, 1995; Fox, 2006).

On the other hand, some Arabic-specific patterns (e.g. heterosyllabic CCs reduction) have also been reported as developmental patterns occurring in the children's speech of languages that share the hetero syllabic CCs structure with the Arabic language, such as Maltese and Italian. Grech (2006) noted that Maltese-speaking children reduced the medial consonant sequence (i.e. heterosyllabic CCs) by age 3;5, which is in agreement with the findings of the current study. Fox-Boyer et al. (2021) reported that heterosyllabic CCs reduction was found in the speech of Italian-speaking children only with a lower cut-off criteria (i.e. >4), which may indicate that this pattern is more common for children younger than age 3;0. This is dissimilar to the results of the current study, where a heterosyllabic CCs reduction pattern occurred under both criteria until age 3;11.

Although all Arabic dialects shared the heterosyllabic CC structure and the gemination feature, only one study, conducted by Mashaqba et al. (2019) on Jordanian Arabic-speaking children, observed the pattern of geminating the heterosyllabic cluster, as also reported in the current study for SHA-speaking children. This pattern has been identified as a typical pattern among children up to the age of 3;11 for SHA-speaking children and up to the age of 3;0 for Jordanians, which was the oldest age group in Mashaqba et al.'s study. No other Arabic study has reported the occurrence of this pattern.

Further, the researcher in the current study investigated CCs reduction patterns in more depth by differentiating between word-final and heterosyllabic CCs reduction because the initial cluster is prohibited in SHA. Interestingly, this study found that children treat the two types of CCs differently. Heterosyllabic CC reduction was observed under both cut-off criteria. The syllable coda within a word is the consonant that was usually deleted by children who expressed this pattern, which is the first consonant in the cluster, and it is suppressed by the age 3;11 under both criteria. Similarly, almost all children who produced word-final CCs reduction pattern tended to delete the first consonant in the cluster (e.g. /gird/→[gid], 'monkey').

Compared to heterosyllabic CCs reduction, word-final CCs were only reduced by a very small percentage of children in the first three age groups and only under the lower cut-off criterion, which indicates that this pattern is not an actual pattern as it did not appear in the speech of children when the higher criterion was applied. However, it is not possible to conclude that the word-final CCs reduction pattern is not typical for SHA-speaking children because the SHAPA test administered in this study to assesses the children's speech is only had nine words with final CCs, which might be not a good representation of this syllable structure in the children's speech. Additionally, considering the dialectal changes (e.g. epenthesis as in /bint/–[binit] 'girl' as in Syrian Arabic) during the scoring of the words with final CCs reduced the occurrences of word-final CC reduction among SHA-speaking children. Therefore, this finding needs to be interpreted with care.

Compared to Arabic studies, although Dyson and Amayreh (2000) reported a heterosyllabic CC reduction pattern using a different label (i.e. 'consonant sequences reduction'), its description matches the pattern identified in the current study. They reported that the abutting consonant sequence was often reduced to a singleton and remained difficult even for the older children in their follow-up study (i.e. >6;0–year–old); in contrast, in the current study heterosyllabic CC reduction disappeared by age 3;11. Similarly, cluster reduction reported by Al-Bader (2009) and Alqattan (2015) who did not specify the word position of the cluster (initial, medial, or final). However, these two studies investigated two Arabic dialects that allowed for all types of clusters: initial, medial, and final (i.e. Kuwaiti and Saudi Najdi), which also limited comparison with this study's findings.

6.2.2.1.3 Developmental SHA-specific patterns

Another developmental pattern found in the current study which could be considered as *SHA-specific* pattern is glottal replacement to /h/. This glottal replacement to /h/ pattern is another pattern that is questionable as it has never been reported by any Arabic study but was found to occur in the speech of the SHA-speaking children before the age of 3;11. All previous studies tended to investigate the glottal replacement pattern in general without specifying if it was a glottal stop /ʔ/ only or if it also included the glottal fricative /h/ as well. However, the replacement with the glottal stop is what is meant by this pattern in most studies (e.g. Abou-Elsaad et al., 2019; Ammar & Morsi, 2006; Owaida, 2015; Saleh et al., 2007). Therefore, it is also difficult to claim that the glottal replacement to /h/ pattern is dialect specific, as it is unknown if it exists in children's speech in other Arabic dialects.

6.2.2.1.4 Non-developmental patterns

Six non-developmental patterns were identified in this study: four types of systemic simplification (i.e. voicing, denasalization, vowel deletion and substitution of /dˤ/→ [ðˤ]) and two structural simplifications (i.e. word-final CCs reduction, and reduplication). Two of the types of systemic simplification (i.e. voicing and denasalisation) and two of the types of structural simplification (i.e. word-final CCs reduction and reduplication) are universal language patterns; however, they are considered as non-developmental in the speech of SHA-speaking children because they appear under the lower cut-off criterion only (≥4). The other two systemic patterns were found as SHA-specific patterns (i.e. vowel deletion and substitution of /dˤ/→ [ðˤ]) but are still non-developmental.

As reported earlier, the word-final CCs reduction pattern needs further examination, which could be conducted by using an updated version of SHAPA to finalise its classification as a developmental or non–developmental patterns. Replacement of /dˤ/→[ðˤ] is a pattern that occurred among the children in this study when the lower criterion was used, except at the age of 5;0–5;5, where it appeared under the higher criterion. Increased exposure to MSA, as well as exposure to other Saudi dialects, by SHA-speaking children by this age (5;5) could explain the occurrence of this pattern because the emphatic dental /ðˤ/ is not a phone in the UHA inventory, but it is part of the MSA inventory. Further, it is a phone in the Saudi Najdi dialect where it is the acceptable replacement of /dˤ/ and such an exchange is not considered an error. Thus, it could be considered that the dental fricative /ðˤ/ is also a part of the phonetic inventory of SHA-speaking children.

In general, comparing the current study's results to other Arabic studies investigating phonological patterns, the differences are not remarkable, and the methodological and analytical variants could explain the variability in their findings compared to the results of the current study. For example, there was a high level of variation among the studies regarding the types of phonological patterns identified in each study, as there is no definitive list of phonological patterns (James, 2001b). Therefore, they focused primarily on common phonological patterns that identified in the literature for English- speaking children, as reported by Ingram (1989). This led to missing many patterns that could be apparent in their data. Furthermore, the speech sampling method, for example, may contribute to the different types of phonological patterns identified across Arabic studies. Yeh and Liu (2021) found that the sampling method may stimulate the elicitation of different phonological patterns, including atypical patterns. They explained that the grammatical and word differences in spontaneous CS had been shown to affect pronunciation accuracy rates and speech simplification.

Table 6.3 demonstrate the language universal and language–specific patterns that reporeted to occur in the speech of children speaking different Arabic dialects as well as SHA-speaking children. However, some of these patterns are found to be non–developmental for SHA-speaking and they are highlighted in yallow in Table 6.3 below

6.2.2 Summary of discussion III

The discussion of the phonological variation analysis findings could be summarized in the following:

- The result of the quantitative analysis of phonological variants in the current study were in line with what was reported in other Arabic studies as well as cross-linguistic studies, where there is a steady decreasing in the number of Tokens, Types and InfrVar.
- The high number of Tokens and InfrVar found in the speech of SHA-speaking children could be related to methodological issues of this study, to the complexity of Arabic gramme, morphological structures, and other linguistic aspects.

- The 27 Types of phonological and phonetic patterns found in the speech of SHA-speaking children were included 18 developmental phonological patterns, three developmental phonetic patterns, and six non-developmental patterns.
- The 18 developmental phonological patterns were included:
    - 10 language universal patters: fronting of /ʃ ʒ/→ [s z θ ð], fronting of /s, z/□[θ, ð], fronting of /k g/→ [t d], stopping of fricatives, assimilation, lateralisation of /r/→[l], devoicing, vowel substitution, weak syllable deletion and word-final consonant deletion.
    - Seven language specific patterns backing /x, ɣ/→ [ħ, ʕ], glottal stop replacement to /ʔ/, de-emphasisation, metathesis, gemination of heterosyllabic CC, word-initial consonant deletions, and heterosyllabic CC reduction.
    - One SHA-specific patterns: glottal replacement to /h/.
- The six non-developmental included:
- Four language universal patterns: voicing, denasalization, word-final CCs reduction and reduplication.
- Two SHA-specific patterns: vowel deletion and substitution of /dˤ/→ [ðˤ].

Table 6.3 *Comparison of the presence of phonological patterns observed in this study with their presence on other children speaking different Arabic dialects*

Phonological Patterns[1]	Saudi Hejazi	Egyptian	Jordanian	Syrian	Kuwaiti	Saudi Najdi
Language–universal patterns						
1) Phonological patterns						
Stopping of fricatives	2;6–4;5	2;0–3;5	2;0–2;11	NA	1;4–5;2	NA
Fronting of /ʃ ʒ/	2;6–5;11	2;0–3;11	Velar: 2;0–2;11	Not specified 2;5–5;5	NA	NA
Fronting of /k g/	2;6–2;11	2;0-3;11	NA	NA	NA	NA
Fronting of /s z/	2;6–5;5	NA	Dentalization 3;0–3;11	Dentalization 2;5–4;5	Dentalization 4;0–4;11	NA
Lateralization of /r/	2;6–5;5	2;0–4;5	2;0–4;0	2;0–5;5	1;4–5;2	2;0–3;11*
De–voicing	2;6–5;11	2;0–5;0	2;0–2;5	2;0–4;5	2;4–2;7	NA
Assimilation	2;6–5;0	2;0–4;5	NA	NA	NA	NA
Voicing	2;6-2;11	NA	NA	NA	NA	NA
Vowel substitution 2;6-5;11	2;6-5;11	NA	NA	NA	NA	NA
Final consonant reduction	2;6–4;5	2;0–2;5	2;0–2;11	2;5–4;5	NA	NA
Word-final CC	2;6–3;5^	2;0–3;5	NA	NA	3;0–3;5	2;0–3;11
Weak syllable deletion	2;6–4;5	2;0–3;11(not specified)	1;0–2;11*	2;5–4;5	3;0–4;5 (weak)	NA
Reduplication	2;6–2;11^	1;0–2;6**	NA	NA	NA	NA
2) Phonetic patterns						
Distortion of /r/	2;6–5;5	/r/ deviation 3;0–5;0	NA	NA	/r/ deviation 2;0–4;11	/r/ deviation 2;0–3;11
Distortion of sibilants /s z ʃ ʒ/	2;6-5;11	NA	NA	NA	NA	NA
Distortion of /l/	2;6-2;11	NA	NA	NA	NA	NA
Language–specific patterns						
Glottal Replacement	2;6–4;5	2;0–3;5	NA	2;5–3;5	NA	NA
Backing	2;6–4;5	2;0–2;5	NA	2;5–3;5	NA	NA
De–emphasisation	2;6–5;5	3;0–5;0*	2;0–6;0	2;5–4;5	1;4–5;2	2;0–2;5
Denasalization	2;6–2;11^	NA	2;0–3;5	NA	NA	NA
Metathesis	2;6–4;5	NA	1;0–3;0*	NA	3;0–3;7	NA
Gemination	2;6–3;11	NA	1;0–3;0*	NA	NA	NA
Word-initial consonant deletion	2;6–3;11	NA	1;0–2;6*	NA	NA	NA
Heterosyllabic CC reduction	2;6–3;11	NA	2;0–6;0	NA	NA	NA
			SHA-specific patterns			
Glottal replacement to /h/	2;6-3;11	NA	NA	NA	NA	NA
Vowel deletion	2;6-3;5	NA	NA	NA	NA	NA
/dˤ/→ [ðˤ]	4;0-5;5	NA	NA	NA	NA	NA

[1]The patterns' named as reported in in the current study**,** structural simplification**,** unshaded: systemic simplification**, Egyptian Arabic** data come from Abou–Elsaad et al. (2019) and Ammar and Morsi (2006)* and Saleh et al. (2007)**; **Jordanian data come from** Dyson and Amayreh (2000) and Mashaqba et al. (2019)*; **Syrian Arabic** data from Owaida (2015); **Kuwaiti Arabic** data from Ayyad (2011) and Alqattan (2015); **Saudi Najdi** data from Al- bader (2009) and Alawwad (2009)*; **Saudi Hejazi** data based on the result of the current study considering the higher criterion ≥6, except for the two highlighted patterns, ≥4 criterion considered). ^

These are non–developmental phonological patterns for SHA-speaking children. **NA:** the pattern did not report in this study.

7 General Discussion and Conclusion

The purpose of this study was to describe the phonological development of 235 typically developing monolingual Saudi Hejazi Arabic (SHA)-speaking children aged between 2;6 and 5;11. This final chapter summarises the outcomes discussed in Chapters 4, 5, and 6, emphasising the main points and addressing the overarching research goals outlined in Section 2.5. The study's strengths and limitations, as well as the clinical implications, will be discussed before concluding with directions for future research.

7.1 Review of the overall study aims and findings

Despite an increasing body of research on monolingual Arabic phonological development over the last decade (Elrefaie et al., 2021), information remains limited on how development proceeds in typically developing SHA-speaking children, its characteristics, and how to best obtain normative data for the phonological development in such a population. Therefore, the present research aimed to:

1) Create a phonological assessment tool (single-word naming test) for the Saudi Hejazi Arabic dialect considering international construction criteria.
2) Describe the typical phonetic and phonological development of SHA–speaking monolingual children aged 2;6–5;11.

The following subsections summarise the extent to which these aims could be achieved with the data collected and analysed in this research.

7.1.1 Creating a phonological assessment tool for the Saudi Hejazi Arabic dialect considering international construction criteria (Aim 1)

The present study has shown that the SHAPA tool has some strengths and weaknesses in terms of being an appropriate clinical phonological assessment tool for SHA-speaking children. Chapter 4, Section 4.4, included a discussion about items that need to be removed because they do not add value to the test (e.g. items that are unfamiliar to all children). In this section, a general discussion will be provided of suggested revisions that could be made to SHAPA to improve its practical application.

Taking into consideration the weaknesses of SHAPA reported in Chapter 4, Section 4.4.4, the limited number of items that represent some UHA features is discussed first, followed by a discussion about increasing the spontaneous naming of SHAPA pictures. Before recommending the addition of more items or changing some other items to overcome such weaknesses, it is important to consider several aspects.

First of all, one of the major concerns with SHAPA is its limited number of items representing consonantal co-occurrences (see Chapter 5, Section 5.2.1.3). Thus, it is important to investigate an appropriate number of items that represent this feature in the speech of SHA-speaking children. Therefore, a question is raised about the number of words with a final CCs and geminations structure required in a phonological test for SHA-speaking children; this information is needed before it can be judged whether or not the ten items included in SHAPA test are sufficient, even though at this point it does seem to not be enough. Further, it is important to keep in mind that the results of this study showed that SHA-speaking children had a low PCCC score for word-final CCs, where they only approached an average score of 80% by age 5;0, while they achieved a score of >80% by age 4;0 for gemination PCCC. At the same time, the word-final CCs reduction pattern did not appear when the higher cut-off criterion (≥6) was applied, which means that word-final CCs reduction cannot be considered as a developmental pattern for SHA-speaking children. This result could indicate the influence of an unrepresentative sample of word-final CCs in SHAPA. However, for gemination, no geminated consonant was reduced among SHA-speaking children, and this finding is in line with what reported by Dyson and Amayreh (2000) and Davis and Ragheb (2014a) where the geminated feature was acquired early by Arabic-speaking children and even when the geminated consonants were substituted, the feature was still preserved. This result may indicate that the number of items with geminated consonants in SHAPA is sufficient and parallels similar findings to what reported in other Arabic literature.

According to Al-Ani, (1970), Arabic language includes several words with word-final CCs, as well as words with word-medial CCs (i.e. geminates); Al-Ani’s descriptions focused on the MSA form of Arabic. However, to date, there is a lack of Arabic studies in general, and Saudi Hejazi studies in particular, that investigate the frequency of occurrence and inventory of CCs in the dialectal Arabic-speaking children’s speech. McLeod et al., (2001b) reported that in English, a large number of CCs are permissible in adult daily speech, either in word-initial or word-final positions, as well as the occurrence of morphological endings which create even more complex phoneme sequences. If these morphological endings are excluded, 18% of English monosyllabic words ends with CCs. Therefore, it is important for SLTs to be knowledgeable about the normal development of CCs because such knowledge can help them to determine whether the speech development of English-speaking children is progressing normally and can assist in selecting targets for intervention. However, such information is still missing about the frequency of CCs in adult speech and the inventory of CCs in dialectal Arabic speech, making it challenging for SLTs working with Arabic-speaking children to determine the normal development of children’s speech, particularly the inclusion of CCs in their speech.

Even though a few studies reported the nature of the acquisition of CCs among details of all other aspects of phonological development (see Chapter 2, Section 2.2.1.2), these studies were limited to small sample sizes (between 2 and 13 children). Further, none of them has reported the frequency or the inventory of the CCs in adult or children's speech (Abdoh, 2010; Dyson & Amayreh, 2000; Ragheb & Davis, 2014). Therefore before adding more items to SHAPA to represent a sufficient sample of consonantal co-occurrences (word-final CCs and geminates), it is recommended to conduct a study similar to that of McLeod et al. (2001a) investigating the frequency of CCs in SHA-speaking children and the inventory of CCs in a free speech sample to determine if the ten items of CCs and geminates in SHAPA are representative or not. Such a step is needed to avoid adding more words to SHAPA to fulfil this missing criterion; it is already an extended test with 151 items, and one of the goals of creating SHAPA is to develop a clinically relevant and applicable phonological test. The results of such studies could guide the researchers/SLTs on how to assess the acquisition of these aspects (word-final CCs and gemination). One possible hypothesis from such a study is that these ten items may be sufficient for inclusion in the general word list of SHAPA. If some children struggle to name these items correctly or have many speech errors in producing the CCs, they could be tested using more items with word-final CCs and geminates by administering a subtest that includes more items with CCs to determine their level of difficulty. Thus, creating a supplementary list for items including these two features could be a possible solution.

Second, as reported in Chapter 4, Section 4.2.1, the rate of spontaneous naming of SHAPA items was lower than that reported in other literature (e.g. Ceron et al., 2020; James, 2001a; Nojavan-Pirehyousefan et al., 2021). Because one of the construction criteria followed in creating single-word phonological assessment tools for children is to ensure that the naming of test items is done spontaneously, the new version of SHAPA needs to include more items familiar to children in the target age range. However, as also reported in Chapter 4, Section 4.4.2.1, deleting items with a high difficulty score (i.e. the items named with a whole-word imitation cue) needs to be done with caution as it could increase the spontaneous naming rate. Another option could be to change the elicitation mode of the words by considering other prompting options (e.g. sentence completion, asking questions). Ceron et al. (2020) reported that using prompts and clear images should guide the child towards the target word. For example, they used sentence completion, such as "The boy will use the pencil to ….? (write)". Such a procedure resulted in high recognition rates for Ceron and colleagues' test items (i.e. >80%). Adding such a procedure in the elicitation of the SHAPA items may improve its spontaneous naming rate.

Age of acquisition is one of the criteria that is usually considered when developing a phonological test for children to ensure items familiarity for the target children and spontaneous naming (Stoel- Gammon & Williams, 2013). According to Section 4.4.2.1, such data were missing in the present study. Most of the SHAPA words were selected from JA-CDI, with some other words selected based on asking Saudi

Hejazi parents and teachers about the age of acquisition of the initial list of SHAPA words, and some other words were selected from other Arabic articulation tests (see Chapter 3, Section 3.3.1.1). These steps were followed by the researcher due to the absence of any database for early Arabic-speaking children's vocabularies. For this reason, it is strongly recommended that future research establish a database for typically Arabic-speaking children's vocabularies and their age of acquisition. Such data are needed to provide information about Arabic word frequency among children, their age of acquisition, and their familiarity with the words (Eisenberg & Hitchcock, 2010). It is also suggested that such a database be constructed with consideration of the variety of Arabic dialects to ensure the cultural appropriateness of words in the database.

Lastly, shortening the SHAPA test could ensure a more spontaneous naming rate because, as reported in Chapter 4, Section 4.4.1, the results of the SHAPA evaluation showed that the spontaneous naming rate was reduced among children in all age groups after they named one hundred items. Therefore, removing items that were shown to lack phonetic or phonological value should increase the applicability of SHAPA as a clinically appropriate assessment tool (see Chapter 4, Section 4.4.2).

In general, the SHAPA evaluation revealed that Aim 1 of this study was not fully achieved. Using the construction criteria, which are universally applied, to create a phonological evaluation tool for SHA-speaking children did not yield a clinically valuable instrument. The vast number of UHA and MSA phones may have impeded the application of the construction criteria, mainly due to the lack of data on the frequency of each UHA phone in each word position. Such information, if available, may lead to the elimination of particular consonants being represented in some rare word positions, leading to a shortened test. Another limitation of the construction principles was that all the essential criteria to develop appropriate phonological tools in specific languages were not explicitly addressed, as including the consonantal co-occurrences feature was not one of the standards. This led to a limited number of items representing this feature in the SHAPA test, which negatively affected the results obtained from these items. Therefore, the construction standards need to be used with consideration of whether the test will be used in a language other than English by considering phonological aspects of the target language; in addition, the goal of the created test should be considered in terms of whether it is for clinical application or research use. These considerations could help SLTs in finding a balance between applying as many criteria as possible to ensure the inclusion of the linguistic guidelines and preserve practicality in administering the test.

Further revisions of this test are essential to improve its applicability in a clinical setting. Once such revisions have been made, the updated version of SHAPA needs to be piloted again to examine the effectiveness of the latest version of SHAPA a practical clinical tool.

7.1.2 Describing the typical phonetic and phonological development in monolingual SHA-speaking children (Aim 2)

The present research has shown that the phonetic and phonological skills of 2;6– to 5;11–year–old monolingual SHA-speaking children are still developing. However, universal developmental trends are apparent in the phonological development of SHA-speaking children, for instance, in the development and sequence of phones acquisition (e.g. nasals and plosives were acquired before fricatives), and for the phonological patterns where SHA-speaking children produced some universal patterns (e.g. stopping of fricatives, fronting of /ʃ, ʒ/, and devoicing). In the following general discussion, an overview is presented of this study's quantitative and qualitative measures. Using these analyses could boost the general understanding of phonological development in this targeted population and evaluate the achievement of Aim 2 of this research. Two subsections follow; the first is to describes the quantitative measurements applied in this study (i.e. PCC, PCC-R, PVC, PhonVar including Tokens and InfrVar), and the second describes the qualitative measures (i.e. the phonetic inventory and Types of phonological patterns). The conclusion of this section will link all these measures and recommend the best clinical practice for these measurements.

7.1.2.1 Performance on quantitative measures

A general developmental trend in the phonological development of SHA-speaking children across the majority of the quantitative applied measures was evidenced in this study. The analysis of two of the quantitative tasks revealed that the average scores of PCC, PCC-R, and PVC increased with age due to the acquisition of more consonants and vowels. In contrast, the number of the other two quantitative measures, Tokens and InfrVar, decreased with increasing age as the number of errors in children's speech decreased. This emphasises the importance of chronological age and biological maturation in monolingual phonological acquisition.

In terms of the influence of chronological age, this study showed that SHA-speaking children's speech becomes more accurate as they get older. Their phonetic inventory of UHA phones using the single-word naming task was mastered by age 5;5 for consonants and by age 4;11 for vowels. It was also found that the average PCC and PCC-R scores reached 90% by age 5;5, showing that SHA-speaking children have typical articulatory competence when they get older (i.e. by age 5;5). The PVC score, on the other hand, reached >90% by the age of 3;0–3;5, and the maximum score was 96.48% by age 5;11. This result is in line with the phonological pattern (vowel substitution), which appeared in the speech of SHA-speaking children at high frequency until the age of 5;11.

In this study, the sounds prone to distortion were taken into consideration when calculating the PCC. A second calculation was conducted where the phonemic production (i.e. the distortions) of sounds was counted as correct. It was revealed that the differences between the PCC and PCC-R scores were minimal as the variation between the age groups scores showed that the distortion output in SHA- speaking children did not considerably reduce PCC scores. This might imply that the low PCC scores, if they are considered low at all, are not attributable to the distortion produced by those phones. As reported in Chapter 5, Section 5.2.1.1, the total average PCC scores for SHA-speaking children did not differ markedly from those reported for Syrian and Jordanian Arabic-speaking children, while they may be considered lower than those reported for the other languages. However, the length of the SHAPA test could justify these lower scores, as well as possibly leading to more errors in the children's productions due to tiredness and fatigue.

This finding raises the question of whether these distorted outputs can be considered errors and whether SLTs should address them as treatment goals. According to Shriberg (1993), distortion mistakes occur in a substantial percentage of young children (i.e. younger than 3–years–old) with regular development as well as children with SSDs. Increased number of distortions with high variability in distorted phonemes, on the other hand, have been connected with phonological problems (Wertzner et al., 2005). Gruber (1999) proposed that there are approximately two paths for children when acquiring typical speech production. He discovered that frequent clinical distortions increased in some children as deletions and substitutions diminished as they 'normalised' their speech by age 5;5; then, after a brief period of roughly five months, a dramatic surge in distortion corrections occurred (i.e. by age 6;0). Gruber found that some other children had a different path to normalisation in which the frequency of deletions, substitutions, and omissions reduced as proper speech output rose. According to Gruber (1999), these distortion errors could be a step toward normalising speech sounds acquisition. Children's distortion production indicated they could use those sounds correctly in words where the meaning did not change because of distorted phonemes (Wertzner, Sotelo, & Amaro, 2005). Such a path could be observed in the acquisition of phones prone to distortions by SHA- speaking children. For example, before the postalveolar fricative /ʃ/ is acquired phonetically by age 5;5, children expressed their ability to produce it with distortion at an earlier stage of their development (i.e. at age of 2;6). Therefore, the phonetic acquisition of consonants may take a longer time, and distortions should not be treated before a certain age. Shriberg, Gruber, and Kwiatkowski (1994) suggested that the distortions that remain in English-speaking children's speech after age 8;0 are considered 'residual speech-sound errors' (p. 1173). In this study, the distortion of sibilant patterns, for example, continue to appear until the age of 5;11 at high frequency, which indicated the need for examining older children to determine the age at this patterns are overcome, it could then be possible to label any distortions that appeared after that age as 'residual speech-errors'. Therefore, distortion in SHA-speaking children may require treatment if the age of overcoming the distortion production of each phoneme passed without this distortion error being corrected (e.g. the age of acquisition of /z/ (no distortion) is 4;5). However, more clinical research is required to support such a hypothesis.

Tokens and InfrVar were two other quantitative measures employed in this study. As mentioned in Chapter 6, Section 6.2.1, the findings of this study revealed that SHA-speaking children had a higher number of Tokens and InfrVar than children speaking other languages (for example, Italian) (Fox-Boyer et al., 2021). Yet, when the SDs and ranges of Tokens and InfrVar calculations were examined across age groups, it was shown that some children had a large number of these measurements (e.g. for age group 5;6–5;11, the maximum rage of InfrVar was 58). On the other hand, some children in this age range had a low number of InfrVar (i.e. 4). These findings were consistent with the mean PCC and PCC-R scores, which showed substantial variability among age groups. For instance, the PCC range for the age group 5;6–5;11 was 69.09–98.58. Indeed, the substantial variation in SD and range of PCC scores, Tokens, and InfrVar for specific children in this study sample might be an additional signal of abnormal phonological development. Macrae et al. (2014) stated that speech output variability for children might indicate speech-language development since it relates to developmental stage transitions. However, one cannot be sure that the variants shown by all children reflects developmental stages without evaluating their performance and comparing their total Token and InfrVar scores to the mean and SD of the age group.

Further research is needed to identify atypically developing children in this study by comparing their performance in the quantitative (i.e. PCC and PCC-R, Tokens and InfrVar) and the qualitative (Types of phonological patterns) measurements to identify typically developing children versus delayed and atypically developing ones. Such a comparison could shed light on the actual number of typically developing children in this study as their performance could be used as norms for SHA-speaking children without the influence of atypical or delayed development children.

Generally, the differences in the phonological acquisition rate of SHA-speaking children in terms of these quantitative measurements compared to other languages suggest the influence of the different factors discussed in Chapter 6, Section 6.2.1. Arabic language complexity, the possible effect of the other linguistic aspects (e.g. lexical development and morphological acquisition), as well as the items familiarity issue of SHAPA are among the factors that could affect the rate of the acquisition of SHA-speaking children's phonology.

7.1.2.2 Performance on qualitative measures

Regarding the qualitative measures, the two relevant aspects are the phonetic inventory and phonological patterns. First, the phonetic inventory of SHA-speaking children was described in this study considering UHA sounds and MSA sounds, as well as the distortion product of those phones prone to distortions similar to the PCC score calculations. Further, this inventory was also compiled using two different tasks: a single-word naming task and a stimulability task. The results showed that the UHA phonetic inventory via the single-word naming task was completed by age 5;5, while it was never completed via the stimulability task where /z, ʃ, ʒ/ were still missing for children aged 5;11 when all distortions were counted as errors. The MSA inventory was completed by age 5;11 via the single- word naming task, and again was never completed via the stimulability task where /z, ʃ, ʒ/ were still missing as these are shared phones between UHA and MSA. Plus, it lacked the trill /r/, with SHA- speaking children finding its imitation in isolation was a challenging task. The differences in the results of the complied phonetic inventories highlighted some essential aspects such as the importance of the stimulability task and the importance of considering the MSA phones, as discussed below.

First, few differences were found between the phonetic inventories complied via both tasks; this finding could demonstrate the importance of the stimulability task as a quick and easy–to–administer clinical screening tool (Kubaschk et al., 2015). The stimulability task showed that SHA-speaking children could imitate most UHA phones in isolation with at a mastery level, except for the trill /r/. In contrast, they could imitate all UHA and MSA phones with an acquisition level of accuracy by age 5;11. The results of the stimulability task could be valuable for SLTs working with SHA-speaking children with SSDs, as this result indicated that not all UHA and MSA phones can to be elicited in isolation. The articulatory complexity of some phones, such as the emphatic consonants and the trill /r/, prevent many children from imitating these sounds in isolation, especially young children aged 2;6–3;6.

On the other hand, this result adds value to such a tool and supports the evidence that its application could provide an overview of the phonetic ability of children aged between 2;6 and 511. One of the advantages of the stimulability task is that it is an easy task to administer and analyse and could be administered by a non–SLTs clinical practitioner. This suggests that this tool could be used as a quick clinical screening tool to identify children who need further evaluation by SLTs. Although the results showed that some phones could not be imitated in isolation, it is still possible to create such a tool by including only sounds that could be imitated in isolation based on the findings of this study or by including sounds that are difficult to imitate in isolation by targeting them in syllables instead. It would also be interesting to investigate in future research the connections between the result of the stimulability task and the phonological ability of children with SSDs, as this could examine the ability of the stimulability task to predict SSD. Miccio et al. (1999) found that children who failed to imitate some sounds during stimulability tasks were not able to acquire these sounds without treatment. Such a finding could support the importance of the stimulability task in identifying children with SSD. Further study is needed to support this hypothesis.

Second, investigating the MSA phone acquisition in SHA-speaking children sheds light on the importance of including those phones to be acquired in the investigation, even if the study mainly targeted SHA-speaking children. As described in Chapter 5, Section 5.2.1.2, the results indicated that the phonetic inventory of UHA phones was completed earlier than the inventory of MSA phones. This late mastery is related to the late exposure to MSA phones. Further, and as reported in Chapter 2, Section 2.1.2, Arabic-speaking children have to learn MSA to be able to read and write in Arabic. Thus, children's ability to develop their Arabic literacy skills would not be fully completed until all the MSA phones are mastered, which is an essential part of their phonological development. Abu– Rabia (2000) found that MSA exposure was substantially linked with the acquisition of reading comprehension in MSA. Saiegh-Haddad (2004) explained that oral exposure to MSA and its linguistic structures positively impacted the acquisition of basic bottom–level processes that underlie reading comprehension. Therefore, investigating the acquisition and mastery of MSA phones cannot be ignored in any phonological development study, even if a study targets the colloquial form of Arabic (cf. Ammar & Morsi, 2006; Owaida, 2015).

Another qualitative measure applied in this study was Types of phonological patterns. The study showed that the total number of Types reduced as age increased, which was true for both cut-off criteria (≥4 and ≥6). Interestingly, applying these two proposed cut-off criteria in this study helped in differentiating between developmental versus non–developmental patterns among SHA-speaking children, as well as showing the importance of clearly and precisely identifying the criteria used to identify the patterns. Applying such clearly defined criteria made the comparison between this study's results and the results of other studies impossible because other researchers did not specify their method in identifying phonological patterns in their sample (cf. Al-buainain, 2012; Ammar & Morsi, 2006; Ayyad, 2011; Mashaqba et al., 2019; Saleh et al., 2007).

This study also emphasised the problem of pattern labelling and how mislabelling a pattern might lead to inaccurate identification of the error and its age of suppression. For example, in this study, three different fronting patterns were identified, and the results revealed various ages of suppression for each type. As a result, labelling a pattern as just fronting without considering the type of error or distinguishing the sound or set of sounds being fronted may result in the misidentification of the delayed children. Smit (1993) found that many phonological patterns are restricted in the range of consonants or word positions they could nominally apply. She argued that some commonly described phonological patterns do not completely specify the range of errors children make (Smit, 1993). However, using accurate descriptions for the patterns with an emphasis on the error sounds increased the number of Types identified in this study. Further, it resulted in finding some patterns that have never been identified in previous Arabic studies, such as 'Glottal replacement to /h/', which was differentiated from 'glottal stop replacement'.

Providing information about the type of errors when identifying the phonological patterns, and even for distortion pattern descriptions, serves to develop criteria for evaluating whether a child's output is atypical and not likely to be developmental, so that appropriate intervention may be provided. (Smit, 1993).

7.1.2.3 Overall performance and clinical application of these measures

This study showed that the phonological development of SHA-speaking children is not completed by age 5;11. The quantitative measures revealed that even by age 5;11, some children still demonstrate relatively low scores for PCC and PCC-R, as well as a high amount of PhonVar (Token and InfrVar). The phonetic inventory for UHA phones completed (i.e. mastered) by age 5;5. However, the acquisition of MSA phones only completed by age 4;5. Some phonological patterns continue to appear in children's speech by age 5;11.

In general, the information in this study about the typical speech acquisition of SHA-speaking children provides a foundation for clinical decision–making. The quantitative measures (PCC and InfrVar) and qualitative measures (phonetic inventory via single-word naming task, stimulability task, and phonological pattern Types) form the proposed analysis for the speech data collected using SHAPA to diagnose SSD in SHA-speaking children. For clinical purposes, it would not be practical to apply all the measures used to establish the normative data as followed in this study. Therefore, it is proposed to use the measures that assist SLTs in identifying typical versus atypical performance, such as the quantitative measures of InfrVar and PCC scores (Holm et al., 2021).

Hua (2002) clarified that the quantitative measures applied in her study (i.e. PCE[16] and inconsistency) are very effective in screening delayed phonological development, but qualitative measures, phonological patterns in particular, as they are used to diagnose consistent and inconsistent disorders. Using more than one measure to differentiate typical from atypical phonological development is recommended. Because of the several underlying deficiencies of SSD, children’s speech processing abilities must be examined using numerous measures targeting distinct speech processing channels to correctly identify children’s strengths and weaknesses (Stackhouse & Wells, 1997). Moreover, the normative data provided in this study seems to be a valuable tool for clearly identifying atypical phonological performance, as all measures for SSD require the availability of normative data for comparison. For SHA-speaking children, the effectiveness of using these measures to accurately diagnoses SSD needs more evaluation as this study did not include children with atypical phonological development or SSD. However, it would be interesting for future research to compare the speech

[16]PCE= Percentage of consonants in errors, which is a measure applied by Hua (2002) on her sample

measurements used in this study (i.e. Token, Types, InfrVar, PCC, and PCC- R) by comparing the individual performance of each child to the performance of the whole age group in order to identify which measure could be an effective clinical marker for classifying children's speech into delay and disorders (e.g. SSD).

Considering all the data provided in this study, the researcher can assume that Aim 2 has been achieved. The normative data established from the analysis of data collected in this study seems to provide a comprehensive description of the typical phonological development of SHA-speaking children between the ages of 2;6 and 5;11.

7.2 Evaluation of the present study

This is the first research project to investigate phonetic and phonological acquisition in SHA-speaking children. Its design offered the opportunity to examine a comparably large number of children (n=235), which allowed for some more confident generalisations than those provided in previous Arabic studies. By assessing the phonological skills of 235 SHA-speaking children aged 2;6–5;11, a large database for their typical phonological acquisition in this population could be created, which allowed for a detailed description of the acquisition of phones (singleton and consonantal co– occurrence), phonological accuracy scores (PCC, PCC-R and PVC), as well as a description of PhonVar in adult–like speech (i.e. phonological patterns Types and InfrVar). The outcomes offered valuable insights into the specific characteristics of typical phonological acquisition in this language.

7.2.1 Strengths and limitations of the study

The present study included the design of a newly created phonology assessment tool, SHAPA, used for examining Saudi Hejazi Arabic phonological skills in SHA-speaking children. A design of a new tool was necessary due to the sparse availability of phonological assessment tools for SHA-speaking children and the lack of some important international test construction criteria for phonology assessments (e.g. linguistic criteria, psychometric criteria; see Chapter 2, Section 2.3.4). The SHAPA was geared to international test construction criteria and was preliminary examined regarding its validity and reliability within this research. Thus, the first aspect this research adds is a practical one, which is the creation of a new and thoroughly designed assessment tool for Saudi Hejazi Arabic phonology skills.

Another strength of this research is the use of a stimulability task to compile the phonetic inventory in addition to a single-word naming task. It has been claimed that the stimulability task could be used as a developmental measure that has a predictive value for typical phonetic acquisition in children (Kubaschk et al., 2015). Comparing the phonetic inventory compiled from these two tasks could influence the age of acquisition of phones (see Chapter 5, Section 5.1.2.3). Such type of comparison has been conducted for the first time in the Arabic literature, and the result of this comparison showed that the phonetic inventory via stimulability task is almost similar to that complied ~~vis~~ single-word naming task. Thus, because the stimulability task is an easy task to administer and analyse, it could be used as a quick screening tool in the Pediatric clinics for the speech sounds acquisition.

Finally, this research included analysis and evaluation not only of the phonological patterns produced, but also the number of InfrVar that occurred in children's speech provided quantitative and qualitative insights into children's phonological skills and qualitative insights into children's phonological skills. This measure has been known for its sensitivity towards the stability of a child's phonological system (Albrecht, 2017), and it provided a potential additional indicator of atypical phonological development (Fox-Boyer et al., 2021).

Besides these strengths, some limitations of the current study must be acknowledged. The first limitation in this study related to creating the SHAPA test. The SHAPA word list was lengthy (n=151 items), which might have influenced the performance of the participating children on both tasks (single-word naming task and stimulability task). The SHAPA item analyses revealed that only 64.24% (n=97) of its items were named spontaneously by more than 50% of participant, which is less than 80% of test items (cf. Ceron et al., 2020, see Chapter 4, Section 4.2.1). Thus, a large number of items were not named spontaneously by all children, which could predict limited lexical development in some children. Such difficulty led to excessive use of imitation cues, which may have produced more accurate speech (Maphalala et al., 2014). In addition to the length of the SHAPA test, the word selection did not provide a perfect sample to examine and analyse the acquisition of consonantal co-occurrence, word-final CCs and gemination in particular. As reported in Chapter 5, Section 5.2.1.3, the word list included only nine words with word-final CCs and only 14 words with middle geminated consonants, which is not a good representative sample to understand the acquisition of such phonological features.

Second, considering the psychometric measures applied to SHAPA, the validity measures were limited and did not provide sufficient validation of this newly designed assessment tool. Concerning the reliability check, only the inter- and intra-rater reliability for the phonetic transcription and phonological patterns identification were reported in this study. This is a limitation in the reliability and validity of the SHAPA as there is a need to conduct more psychometric testing (e.g. test-retest reliability and diagnostic validity) to improve its clinical application.

Third, some children refused to participate in the stimulability task, or they refuse to imitate some sounds (see Chapter 5, Section 5.1.2.2). Such a condition could be explained by children's willingness and motivation to do the stimulability task after the long single-word naming task. Administering this task as a second task during the testing session, could have influenced the total results obtained in this study.

Thus, the design of the testing session may have had an impact on the number of children who fully participated in this task, which adds to the limitations of this study. Moreover, the procedure followed in this task (see Chapter 3, Section 3.3.2.1) may also have affected the results of this stimulability task, as the researcher did not provide a full cuing hierarchy that included some tactile cues (Tyler & Tolbert, 2002).

Finally, the mode of speech sample elicitation used in this study may have limited the obtained results, as only the single-word naming task and phone imitation were applied. Therefore, the generalisations to the connected speech may be limited.

7.2.2 Theoretical implications of the study findings

This research resulted in some theoretical implications for phonological development in monolingual SHA-speaking children under the scope of the emergentist theory. As discussed in Chapter 1, Section 1.1.3, the emergence approach, which considered children's speech acquisition by combining both general principles (e.g. formalist approaches) and individual capacity (e.g. functionalist approaches), may provide a better theoretical framework for SHA-speaking children regarding the acquisition of phonology (Davis & Bedore, 2013; McLeod & Crowe, 2018).

The following sections address the theoretical implications of this study's findings and the general observations of the results in line with aspects reported in the literature review (Chapters 1 and 2). These aspects include tasks and measures, test construction and the psychometric properties, Arabic phonological acquisition, and assessment tools.

7.2.3.1 Tasks and measures

As reported in Chapter 1, Section 1.3.2, the SWN and stimulability tasks were chosen for this study as they are suitable for the phonetic and phonological analysis based on the emergence approach. Using SHAPA, including the SWN and stimulability tasks, could be linked to the emergence approach in some ways.

First, using the SWN task in this study has resulted in collecting appropriate speech samples from the SHA-speaking children that were analysed phonetically and phonologically to explore the phonological development in those children. The collected samples encompassed many of the phonological features of the SHA dialect that are commonly found in adult speakers' speech. For instance, the naming task included a variety of words with UHA phone /g/, bi- and multisyllabic word length, and consonant co-occurrences structure such as word-final CC, heterosyllabic CC, and medial geminated consonants. To ensure the children's cooperation, the test words were presented using child-friendly, colourful cartoon pictures (refer to Chapter 3, Section 3.3.1.1). Including such words and pictures in SHAPA could reflect the children's perception-cognitive-production system capacities, which are essential components of the emergence of phonological knowledge (Davis & Bedore, 2013). Second, based on the emergence approach, data collected via the stimulability task (i.e. the phone imitation task) suggests that the peripheral speech structure capacities for SHA-speaking children were fully developed for all children by age 5;5. For example, the result of the stimulability task analysis showed that all Arabic phones including the emphatic sounds (i.e. /tˤ, sˤ, ðˤ/) were acquired by SHA-speaking children by age 5;5, (see Chapter 5, Section 5.1.2.2). This could be attributed to the effect of language exposure on phonological development. The emergence approach states that children develop precise phonological knowledge structures based on their language requirements through internal connections and social interaction skills, which are improved by external stimuli (Davis & Bedor, 2013).

7.2.3.2 Test construction and psychometric properties

In Chapter 4, the analysis of the SHAPA test construction indicated that it provided a suitable speech sample for phonetic and phonological analysis. As an example, the test items' familiarity results demonstrated that more than half of the children who participated could name the test words without any additional cues (see Chapter 4, Section 4.2.1). Under the emergence theory, imitated behaviours resulting from sensory-motor-based bodily movements may not accurately reflect the phonological knowledge stored in children's memory (Davis & Bedore, 2013). Therefore, such words may not provide a reliable sample for examining the permanently stored phonological structures in children's memory. The emergence approach has emphasized the important connection between phonological working memory and the development of vocabulary and skills in young children. This highlights the crucial role of this mechanism in the complex phonological system (Davis & Bedore, 2013). A thorough analysis of the SHAPA test items revealed a reasonable construction of the test regarding word selection and word structures (as reported in the previous section). This study also considered children's cognitive skills in the normative sample by selecting the age range of children expected to name the SHAPA items as intended (see Chapter 4, Section 4.2.2).

7.2.3.3 Arabic phonological acquisition

In terms of Arabic phonological acquisition, the results of this study may have theoretical implications for phonological development in SHA-speaking children in the following ways. First, as the results reported in Chapter 5, Section 5.1.2 have shown, the trajectory of SHA-speaking children's phonetic development was in line with the universal acquisition trend. Although such findings support the formalist approach of phonological theory, the study also reveals variations that cannot be accounted for by that theory. For example, the early acquisition of the back fricatives such as /ħ, ʕ/, which are not common in many languages, before other common fricatives such as /z, ʃ/ supports the notion of the influence of ambient language phonology on the development of a child's speech. Within the emergence approach, the effect of ambient language is one source of external capacities available to children, linked to the internal capabilities (e.g. biological) and to the process of physical maturation to constitute the complex phonological system (Davis & Bedore, 2013). Therefore, understanding the differences in phonetic development among SHA-speaking children compared to other languages could be enhanced under the emergence theoretical approach

Based on the phonetic development of SHA-speaking children, it was discovered that their trajectory aligned with those of children who spoke other Arabic dialects. Previous Arabic studies have implemented different methodologies, but the general findings of this study did not reveal any significant differences in phonetic acquisition across the compared dialects. It is important to note that these dialects have varying phonetic features, including the use of dental phones (/θ, ð, ðˤ/) which can affect the age at which Arabic phones are acquired by children. In certain cultures, children who are exposed to more words with dental sounds tend to acquire these sounds at an earlier age compared to children exposed to an Arabic dialect that lacks dental sounds. This was observed in Kuwaiti-speaking children who acquired dental sounds (/θ, ð, ðˤ/) earlier than Jordanian-speaking children. However, SHA-speaking children were found to acquire some of these sounds (/ð, ðˤ/) at an early age (by 3;5), despite their absence in the dialect's inventory. In fact, these sounds are part of the MSA inventory, and SHA-speaking children exposed to this form of Arabic at an early age (≤3;0). In contrast, Jordanian-speaking children acquired these sounds by age 6;4, as they did not expose to these dental fricatives before the age of 6;0. Again, such findings support the emergence approach which considered the sociocultural environment as a significant role in shaping the overall phonological system (Davis & Bedore, 2013).

In Chapter 6, Section 6.1.2, the phonological analysis showed that there was a consensus regarding the phonological patterns found in the speech of SHA-speaking children and those speaking other Arabic dialects. The few discrepancies reported between the patterns observed in SHA-speaking children and those speaking different dialects, such as Saudi Najdi and Kuwaiti, were associated with differences in phonological characteristics. Additionally, the presence of language-specific patterns in SHA-speaking children indicated the influence of their surroundings, as described in Chapter 6, Section 6.1.2. These findings align with emergentist theory's stance on ambient language and its impact on interactional styles in the child's environment. These cultural interaction patterns are strong enough to support the emergence of phonology in various cultural contexts across these phonological systems (Davis & Bedore, 2013). While the occurrence of the universal phonological patterns in the speech of SHA-speaking children may also support the formalist theory (e.g. structuralist and Natural phonology), there is no consensus on the age at which these patterns disappear. For example, while such patterns occur universally in children who speak SHA, they disappear at different ages than in children who speak other languages (see Chapter 6, Section 6.1.2).

Other theoretical implications found in the result of this study and maybe also related to the Arabic phonological acquisition in general, are the influence of individual differences and oral-motor system maturation on the phonological system (Davis & Bedore, 2013). First, the wide range and large standard deviations found in the score of Types and InfrVar of the phonological patterns for some age groups among SHA-speaking children provide support for individual variability (see Chapter 6, Section 6.2.1). The high means of the InfrVar across all age groups show evidence of variability and individual differences in the SHA-speaking children's phonology. Such an influence is one of the perspectives of the biological model of the functionalist theory (Vihman, 2014), and it also highlighted in the emergence approach view (Divas & Bedore, 2013). Second, the distortion productions that occurred in the speech of SHA-speaking children and the low percentage of PCC could prove the effect of oral-motor skills. Reducing the distortion productions of speech sounds prone to distortion shows how structural maturation affects phonetic development and phonetic production accuracy (see Chapter 5, Section 5.2.1). The production of these sounds requires finely grained articulator placements to be superimposed on the rhythmic jaw cycle during the closure phase (Davis & Bedore, 2013).

7.2.3.4 Assessment tools

The criteria reported in Chapter 3, Section 3.3.1.1, that were used to develop the assessment tool for this study (SHAPA test), were primarily based on the perspective of the emergence approach. Unlike other assessment tools reviewed in Chapter 2, Section 2.3, the SHAPA word list was developed considering the variations in children's speech, which may occur by chance. Therefore, the cut-off criteria used to identify phonological patterns set to prevent such assumption. Plus, using of the phonetic transcription of the whole word of the SHAPA word list during the analysis is an essential component of the emergence approach (Davis & Bedore, 2013). As reported in Section 7.2.3.2, the results of the SHAPA analysis results were also another theoretical implication for this study by the emergence approach regarding the spontaneous naming rate.

To summarize, the emergence approach seems to be a suitable phonological theory that explains the impact of external factors, such as adults' speech patterns and cultural background, on the phonological development of children who speak SHA. However, additional research is necessary to fully explore this approach and its application to Arabic phonology overall. Such research should investigate other areas that were not considered in this study, such as children's production, perception, cognitive abilities, and interaction skills. Additionally, the study should examine the influence of factors like phonetic complexity, functional load, phonetic frequency, and phonotactic probability and compare data from SHA-speaking children to that of children who speak other Arabic dialects or multiple languages (McLeod & Crowe, 2018).

7.2.3 Clinical implications of the study's findings

The outcomes of the present study provide some implications for the clinical practice, especially the assessment process of developmental SSD in SHA-speaking children. These are addressed in the following discussion.

First, approximately 67% of SLTs' caseload in Saudi Arabic are children with suspected SSD (Alanazi, 2017), and they are required to assess and decide whether a child's speech sound skills developed normally or if he/she needs intervention. The normative data reported in the current study should enable SLTs to make their clinical decision, because this data contribute to the evidence base of practice and provide some practical material for clinicians to guide their evaluation of SHA-speaking children. This study's norms are based on a large representative sample of the dialect under investigation. It was designed to include different age groups of different children acquiring SHA so that the norms would be sensitive to sociolinguistic variation. SLTs can use this information to assess speech sound acquisition (phonetic inventory), accuracy (linked to intelligibility), and whether the path of speech sounds development is typical or not (phonological patterns).

Second, as this is the first study on the acquisition of phonology in Saudi Hejazi Arabic, it provides SLTs with a clinical tool (SHAPA) to assess the development of phonological skills in SHA-speaking children. The information obtained is most likely to be considered as a prerequisite to establishing a standardised phonological assessment tool after further examination of its reliability and validity properties.

Finally, the qualitative differences regarding the types of phonological patterns identified in this study and in children acquiring other languages may have led to an over–identification of SSD, since some phonological patterns that appeared in SHA-speaking children may be considered as atypical for children acquiring other languages (e.g. English). For example, it was found that initial consonant deletion is typical for SHA-speaking children up to the age of 3;5, but it has been labelled as one of the atypical patterns for children acquiring English (Dodd et al., 2003).

7.3 Future directions

In the present research, all the research questions could be answered. However, this project also motivated some further questions and suggestions which require future work to be conducted. These suggestions are summarised below.

First, it has been hypothesised throughout this thesis that children's low performance in the item's familiarity score may have biased their phonological outcome. Therefore, it is essential to revise the word list of SHAPA further and pilot the new version of the test. These changes should be done by considering the presentation of each phone in each possible position at least twice, following the international requirements (e.g. Eisenberg & Hitchcock, 2010; Stokes et al., 2005). Therefore, the following steps should be carried out in order to revise the SHAPA:

1) All items named spontaneously by fewer than 50% of children across all age groups should be reviewed carefully in terms of their phonetic–phonological features. They should be removed if their phonetic features are presented in another item, or if after their deletion there will still be two occurrences of their phones in the targeted position.
2) All items named by imitation by more than 50% of children should be removed if their phonetic–phonological features are presented in other items. If not, then their mode of elicitation needs to be changed, or they should be exchanged with other items, if possible.
3) All items named with other prompts also need to be reviewed for their phonetic–phonological features. If they are replaceable, then these items need to be deleted. Otherwise, changes need to be made to their mode of elicitation.

Such changes may improve the practicality of SHAPA to be used as a clinical tool to assess phonological development in SHA-speaking children. Therefore, the new revision of the SHAPA should be piloted in future investigation to evaluate its clinical practicality (see Section 7.1.1).

Second, after revising the SHAPA word list, this tool must still undergo additional psychometric testing to determine its validity and reliability (e.g. Fabiano-Smith, 2019). Further, SHAPA needs to be administered on children with SSD to compare their performance on SHAPA with the typically developing children as such a comparison could increase the clinical importance of this test for diagnosing children with SSD (Ahmadi et al., 2018). Dodd et al. (2006) stated that one of the characteristics of the normative sample is to include children with a history of speech or language problems to avoid over-identification of typically developing children that fall at the lower end of a normal distribution curve. However, according to Hua (2006), the more children a study examines, the less the effect of inclusion of children with difficulties and the more representative the description of phonological acquisition. Because there were 235 typically developing children in the normative sample of this study, it was hypothesised that including children with SSD could not affect the description of the phonological development of SHA-speaking children. However, this is still a point to consider in a future study to exclude the possibility of over-identification of normal development. A further step is needed to improve the clinical application of SHAPA to include children with SSD in the normative sample.

Third, it is recommended for future research to investigate the effectiveness of the stimulability task to be used as clinical screening tool, as reported in Section 7.1.2.2.

Finally, one of the possible consequences of the present study is to encourage more research on Arabic phonological development in general. Further investigation of phonological development in more children speaking different Arabic dialects is needed to develop normative data applicable to all or most Arabic dialects considering all similarities and variation between Arabic varieties. It would also be beneficial to investigate bilingual Saudi Arabic-English-speaking children to compare their phonological development to the results of the monolingual SHA-speaking children found in this research. The number of children speaking English is increasing every day around the world, and in SA, it has become the norm for children to grow up speaking Arabic and English. Therefore, investigating the trajectory of Arabic phonological development in these bilingual children would add to the general understanding of how these children acquired the Arabic phonological system compared to monolingual children based on the result of the current study.

7.4 Summary and conclusion

This study investigated the phonological development of SHA-speaking children using a newly created phonological assessment tool (SHAPA) that was linguistically controlled and designed following international construction criteria. Although it has advantages, SHAPA requires some revisions before it can be used as a clinical tool to evaluate children's speech. Therefore, the first aim of the study was not fully achieved. However, the data collected via SHAPA was analysed in terms of quantitative and qualitative measures to achieve the second aim of this study.

The data of the present research has provided insight into important aspects of the phonological acquisition of SHA-speaking children aged 2;6 to 5;11. In order to complete the picture and further our understanding of children's entire phonological acquisition process (e.g. identifying the number of phonological variants in children's speech), it would be necessary to extend the analysis to phonetic aspects (e.g. the age of acquisition of consonants and vowels) of both UHA and MSA. Furthermore, some of the 2;6–year–old children had already mastered some of the phones (UHA and MSA), and some of the 5;11–year–old children were still in the process of acquiring aspects of the phonological system of SHA (i.e. not mastering all MSA phones and still using some of the systemic phonological patterns and phonetic distortion patterns, as well as InfrVar that were still appearing in their speech). Thus, it would be revealing to assess younger and older SHA-speaking children to find out the age at which UHA phones emerge and the age of completion of the phonological development. It can otherwise be assumed that the second aim of the study has been achieved and that the phonological development of SHA-speaking children has been described with consideration of the phonological and phonetic aspects. In general, it can be concluded that the phonological analysis conducted in this study will help SLTs working with SHA-speaking children by providing them with normative data regarding their typical phonological development, which could help with clinical decision–making during the evaluation of SSDs.

As a next step to build on this study, it is recommended to create a second version of the SHAPA test, and then pilot it to ensure an increase in its items' familiarity score and spontaneous naming rate. The psychometric evidence of this test needs to be examined and established, to ensure that SHAPA is a reliable clinical tool for SLTs. Moreover, the normative data presented in this study could provide the basis required for a clear specification of clinical markers for SSDs in SHA-speaking children.

8 References

Aalto, E., Saaristo-Helin, K., & Stolt, S. (2020). Phonological development of Finnish speaking children at 3;6 and associations to previous and simultaneous lexical ability. *Clinical Linguistics and Phonetics*, *34*(7), 617–633. https://doi.org/10.1080/02699206.2019.1681517

Abdoh, E. (2010). *A Study of the phonological Structure and Representation of First Words in Arabic*. [Unpublished doctoral thesis]. University of Leicester.

Abou-Elsaad, T., Afsah, O., & Rabea, M. (2019). Identification of phonological processes in Arabic–speaking Egyptian children by single-word test. *Journal of Communication Disorders*, *77*(October 2018), 80–93. https://doi.org/10.1016/j.jcomdis.2018.10.003

Abou-Elsaad, T., Baz, H., & El-Banna, M. (2009). Developing an articulation test for arabic-speaking school-age children. *Folia Phoniatrica et Logopaedica*, *61*(5), 275–282. https://doi.org/10.1159/000235650

Abu-Abbas, K. H., M. Zuraiq, W., & Abdel-Ghafer, O. A. (2011). Geminates and Long Consonants in Jordanian Arabic. *International Journal of Linguistics*, *3*(1). https://doi.org/10.5296/ijl.v3i1.735

Abu-Rabia, S. (2000). effects of exposure to literary Arabic on reading comprehension in a diglossic situation. *Reading & Writing: An Interdisciplinary Journal*, *13*(1–2), 147–157. Retrieved from http://link.springer.com/article/10.1023/A:1008133701024#

Abu Guba, M. N. (2018). Stress assignment in polysyllabic words in Levantine Arabic: An Optimality-Theoretic analysis. *Lingua Posnaniensis*, *60*(2), 7–24. https://doi.org/10.2478/linpo- 2018-0009

Ahmadi, A., Ebadi, A., Kamali, M., Zarifian, T., Dastjerdi Kazemi, M., & Mohamadi, R. (2018). Single word test for the assessment of speech sound production in Persian speaking children: Development, validity and reliability. *International Journal of Pediatric Otorhinolaryngology*, *114*(May), 61–66. https://doi.org/10.1016/j.ijporl.2018.08.018

Al-Ani, S. (1970a). *Arabic phonology: An acoustical and physiological investigation.* Paris: The Hague Mouton. Retrieved from https://doi.org/10.1515/9783110878769

Al-Ani, S. (1970b). Arabic Phonology, 13–22. https://doi.org/10.1515/9783110878769

Al-Bader, E. (2009). *Phonological Processes in Saudi Najdi Arabic-Speaking Children: Aged two to three years*. [Unpublished master thesis]. Imam Muhammad bin Saud Islamic University.

Al-buainain, H. (2012). Baseline Data for Arabic Acquisition with Clinical Applications: Some Phonological Processes in Qatari Children's Speech. *International Journal of Business and Social Research*, *2*(6), 18–33. https://doi.org/10.18533/ijbsr.v2i6.131

Al-huri, I. (2015). English Literature and Language Review Arabic Language: Historic and Sociolinguistic Characteristics, (August 2016), 28–36. https://doi.org/10.13140/RG.2.2.16163.66089/1

Al-Mohanna, F. (1998). *Syllabification and metrification in urban Hijazi Arabic: Between rules and* *constraints* [Unpublished doctoral thesis]. University of Essex, Colchester, United Kingdom.

Al-Muhanna, F. (2009). Positional Syllable Minimality: Syllabification in Hejazi. *King Saud University*.

Al-Sabi, Y. N. (2017). The JISH Speech, Language, and Hearing School Readiness Screening in Jeddah, Saudi Arabia. *Journal of Otolaryngology-ENT Research, 7*(5). https://doi.org/10.15406/joentr.2017.07.00221

Al-Sabi, Y. N., & Naqawah, A. (2013). JISH Articulation Test (1st ed.). Jeddah Institue for Speech and Hearing.

Al Amro, M. (2019). Word Stress in Saudi Najdi Arabic. *American International Journal of Social Science*, *8*(3), 30–34. https://doi.org/10.30845/aijss.v8n3p5

Al Nassir, A. (1985). *Sibawayh The Phonologist, A critical study of the phonetic and phonological theory of Sibawayh as presented in his treatise Al Kitab.* (Order No. U352187). Available from ProQuest Dissertations & Theses A&I. (301397992). Retrieved from .

Al Qenaie, S. (2011). Kuwaiti Arabic: A Socio-Phonological Perspective. [Unpublished doctoral thesis]. University of Durham.

Alanazi, A. A. (2017). Audiology and speech-language pathology practice in Saudi Arabia. *International Journal of Health Sciences*, *11*(5), 43–55. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/29114194%0A

Alawwad, G. (2009). *Phonological Development of 3- to 4-year-old Najdi Children*. [Unpublished master thesis]. Imam Muhammad bin Saud Islamic University.

Albrecht, K. M. (2017). *The Identification of Typical and Atypical Phonological Acquisition in Turkish-German Bilingual Children* [Doctoral thesis]. University of Sheffield. White Rose eTheses Online. http://etheses.whiterose.ac.uk/18545/

Alfaifi, A. A. (2019). *Syllabification of Coda Consonant Clusters in Najdi and Hejazi Arabic* [Unpublished doctoral thesis]. George Mason University .Retrieved from http://search.proquest.com/openview/17b448613eeb1a7db48b66bfd75ba71f/1?pqorigsite=gscholar&cbl=18750&diss=y

Alghamdi, M. (2015). *Arabic phonetics and phonology* (1st ed.). Riyadh: Tawbah Publisher.

Alghamdi, M. M. (1998). A Spectrographic Analysis of Arabic Vowels: A Cross-dialect Study. *Journal of King Saud University*.

Alhazmi, L. (2018). *Perceptions of Hejazi Arabic Dialects: An attitudinal approach* [Doctoral thesis]. University of Sheffield. White Rose eTheses Online.

Almbark, R. A., & Hellmuth, S. J. (2015). Acoustic Analysis of the Syrian Arabic Vowel System. Proceedings of the 18th International Congress of Phonetic Sciences (ICPhS 2015).

Almurashi, W., Al-Tamimi, J., & Khattab, G. (2020). Static and dynamic cues in vowel production in Hijazi Arabic. *The Journal of the Acoustical Society of America*, *147*(4), 2917–2927. https://doi.org/10.1121/10.0001004

Alqahtani, M. S. M. (2014). *Syllable structure and related processes in optimality theory: an examination of Najdi Arabic* [Doctoral thesis]. Newcastle University. Retrieved from http://hdl.handle.net/10443/2757

Alqattan, S. (2015). *Early Phonological Acquisition by Kuwaiti Arabic Children* [Doctoral thesis]. Newcastle University. Retrieved from http://theses.ncl.ac.uk/jspui/handle/10443/2787

Alzaidi, M. S. A. (2014). *Information Structure and Intonation in Hijazi Arabic* [Unpublished doctoral thesis]. University of Essex.

Amayerh, M. M. (1994). *The acquisition of Arabic consonants* [Unpublished doctoral thesis]. University of Florida.

Amayreh, M. M. (2003). Completion of the consonant inventory of Arabic. *Journal of Speech, Language, and Hearing Research*, *46*(3), 517–529. https://doi.org/10.1044/1092-4388(2003/042)

Amayreh, M. M., & Dyson, A. T. (1998). The Acquisition of Arabic Consonants. *Journal of Speech and Hearing Research*, *41*(3), 642–653. Retrieved from 10.1044/jslhr.4103.642

Amayreh, M. M., & Dyson, A. T. (2000). Phonetic inventories of young Arabic-speaking children. *Clinical Linguistics and Phonetics*, *14*(3), 193–215. https://doi.org/10.1080/026992000298823

Ammar, W., & Morsi, R. (2006). Phonological Development and Disorders: Colloquial Egyptian Arabic. In Zhu Hua & B. Dodd (Eds.), *Phonological Development and Disorders in Children* (pp. 204–232). Multilingual Matters Limited. https://doi.org/10.21832/9781853598906-011

Arabian business. (2016). arabianbusiness.com. Retrieved July 16, 2022, from https://www.arabianbusiness.com/politics-economics/expats-outnumber-saudi-nationals-injeddah-jubail-635579

Ayyad, H. (2011). *Phonological Development of Typically developing Kuwaiti Arabic-Speaking Preschoolers* [Unpublished doctoral thesis]. The University of British Columbia.

Ayyad, H., & Bernhardt, B. M. (2009). Phonological development of Kuwaiti Arabic: Preliminary data phonological development of Kuwaiti Arabic. *Clinical Linguistics and Phonetics*, *23*(11), 794–807. https://doi.org/10.3109/02699200903236493

Ayyad, H. S., Bernhardt, B. M., & Stemberger, J. P. (2016). Kuwaiti Arabic: acquisition of singleton consonants. *International Journal of Language and Communication Disorders*, *51*(5), 531–545. https://doi.org/10.1111/1460-6984.12229

Baker, E. (2004). Phonological analysis summary and management plan. *Acquiring Knowledge in Speech, Language and Hearing*, *6*(1), 14–21.

Barlow, J. A., & Gierut, J. A. (1999). Optimality Theory in Phonological Acquisition, (December), 1482–1498.

Basalamah, M. (1990). *A Study of Certain Aspects of the Suprasegmental of Arabic and their Influence on the English Pronunciation of Arab Teachers of English* [Unpublished doctoral thesis].

University of London.

Belur, J., Tompson, L., Thornton, A., & Simon, M. (2021). Interrater Reliability in Systematic Review Methodology: Exploring Variation in Coder Decision-Making. *Sociological Methods and Research*, *50*(2), 837–865. https://doi.org/10.1177/0049124118799372

Bernhardt, B. H., & Holdgrafer, G. (2001a). Beyond the Basics I. *Language, Speech, and Hearing Services in Schools*, *32*(1), 18–27. https://doi.org/10.1044/0161-1461(2001/002)

Bernhardt, B. H., & Holdgrafer, G. (2001b). Beyond the basics II: Supplemental sampling for in-depth phonological analysis. *Language, Speech, and Hearing Services in Schools*, *32*(1), 28–37. https://doi.org/10.1044/0161-1461(2001/003)

Bernhardt, B., & Stemberger, J. (1998). Handbook of phonological development: From the perspective of constraint-based nonlinear phonology: Academic Press San Diego, CA.

Bin-Muqbil, M. S. (2006). *Phonetic and phonological aspects of Arabic emphatics and gutturals* [Unpublished doctoral thesis]. University of Wisconsin-Madison.

Birnstiel, D. (2019). Classical Arabic. In J. Huehnergard & N. Pat-El. In *The Semitic Languages* (2ed ed., pp. 367–402). New York: Routledge.

Bland-Stewart, L. M. (2003). Phonetic Inventories and Phonological Patterns of African American, *24*(03), 109–120. https://doi.org/10.1177/15257401030240030201

Bokhari, H. (2020). *A Comprehensive Analysis of Coda Clusters in Hijazi Arabic: An Optimality Theoretic perspective*. [Unpublished doctoral thesis]. Indiana University.

Brame, M. (1970). Arabic Phonology: Implications for Phonological Theory and Historical Semitic. The Massachusetts Institute of Technology.

Brosseau-Lapré, F., & Rvachew, S. (2018). *Introduction to speech sound disorders*. (F. BrosseauLapré & S. Rvachew, Eds.). Plural Publishing.

Bybee, J. (2001). Language Use as Part of Linguistic Theory. In *Phonology and Language Use* (pp. 1–18). https://doi.org/10.1017/cbo9780511612886.001

Campbell, T. F., Dollaghan, C., Janosky, J. E., & Adelson, P. D. (2007). A Performance Curve for Assessing Change in Percentage of Consonants Correct–Revised (PCC-R). *Journal of Speech, Language, and Hearing Research*, *50*(4), 1110–1119. https://doi.org/10.1044/1092-4388(2007/077)

Carmines, E. G., & Zeller, R. A. (2011). Introduction in Reliability and Validity Assessment. *Sage Publications, Inc.*, 8.

Ceron, M. I., Gubiani, M. B., de Oliveira, C. R., & Keske-Soares, M. (2017). Factors influencing consonant acquisition in Brazilian Portuguese–Speaking children. *Journal of Speech, Language, and Hearing Research*, *60*(4), 759–771. https://doi.org/10.1044/2016_JSLHR-S-15- 0208

Ceron, M. I., Gubiani, M. B., de Oliveira, C. R., & Keske-Soares, M. (2020). Phonological assessment instrument (INFONO): A pilot study. *Codas*, *32*(4), 1–13. https://doi.org/10.1590/2317-1782/20202019105

Chen, R. K., Bernhardt, B. M., & Stemberger, J. P. (2016). Phonological assessment and analysis tools for Tagalog: Preliminary development. *Clinical Linguistics and Phonetics*, *30*(8), 599–627. https://doi.org/10.3109/02699206.2016.1157208

Chomsky, N., & Halle, M. (1968). *The Sound Pattern of English. Harper & Row*. New York. https://doi.org/10.1037/031431

Clausen, M. C., & Fox-Boyer, A. V. (2017). Phonological development of Danish-speaking children: A normative cross-sectional study. *Clinical Linguistics and Phonetics*, *31*(6), 440–458. https://doi.org/10.1080/02699206.2017.1308014

Clements, N. (1990). The role of the sonority cycle in core syllabification. In J. Kingston & M. Beckman (Eds.), Papers in Laboratory Phonology I: Between the Grammar and Physics of Speech. (pp. 283–333). Cambridge: Cambridge University Press.

Cohen, W., & Anderson, C. (2011). Identification of phonological processes in preschool children's single-word productions. *International Journal of Language and Communication Disorders*, *46*(4), 481–488. https://doi.org/10.1111/j.1460-6984.2011.00011.x

Cole, P. A., & Taylor, O. L. (1990). Performance of Working Class African-American Children on Three Tests of Articulation. *Language, Speech, and Hearing Services in Schools*, *21*(3), 171–176. https://doi.org/10.1044/0161-1461.2103.171

Cowan, W. (1970). The Vowels of Egyptian Arabic. WORD, *26*(1), 94–100. https://doi.org/10.1080/00437956.1970.11435584

Crais, E. R. (2011). Testing and Beyond: Strategies and Tools for Evaluating and Assessing Infants and Toddlers. *Language Speech and Hearing Services in Schools*, *42*(3), 341. https://doi.org/10.1044/0161-1461(2010/09-0061)

Crowe, K., & McLeod, S. (2020). Children's English consonant acquisition in the united states: A review. *American Journal of Speech-Language Pathology*, *29*(4), 2155–2165. https://doi.org/10.1044/2020_AJSLP-19-00168

Crystal, D. (2008). *A Dictionary of Linguistics and Phonetics*. ProQuest Ebook Central. https://doi.org/10.1002/9781444302776

Dashash, N., & Safi, S. (2014). *JISH Arabic Communicative Development Inventory: Saudi population JACDI: User's guide and technical manual.* Jeddah: Jeddah Institute for Speech and Hearing

Davis, B. L., & Bedore, L. M. (2013). *An Emergence Approach to Speech Acquisition*. East Sussex: Psychology Press. https://doi.org/10.4324/9780203375303-9

Davis, S., & Ragheb, M. (2014). Geminate representation in Arabic, 3–20. Retrieved from https://www.jbe-platform.com/content/books/9789027270665-sal.1.04dav

De Bel-Air, F. (2018). Demography, Migration, and Labour Market in Saudi Arabia. *Gulf Research*

Center, (5),19. Retrieved from http://cadmus.eui.eu/bitstream/handle/1814/32431/GLMM_ExpNote_08-2014.pdf?sequence=1

de Castro, M. M., & Wertzner, H. F. (2012). Stimulability: Auxiliary measure in the identification of difficulty in speech sounds production. *Jornal Da Sociedade Brasileira de Fonoaudiologia*, *24*(1), 49–56. https://doi.org/10.1590/S2179-64912012000100009

Dinnsen, D. A., Chin, S. B., Elbert, M., & Powell, T. W. (1990). Some constraints on functionally disordered phonologies: Phonetic inventories and phonotactics. *Journal of Speech and Hearing Research*, *33*(1), 28–37. https://doi.org/10.1044/jshr.3301.28

Dodd, B. (2005). *Differential diagnosis and treatment of children with speech disorder*. (Ed.) (2ed ed.). London: Whurr.

Dodd, B. (2014). Differential Diagnosis of Pediatric Speech Sound Disorder. *Current Developmental Disorders Reports*, *1*(3), 189–196. https://doi.org/10.1007/s40474-014-0017-3

Dodd, B., Holm, A., Crosbie, S., & Hua, Z. (2005). Children's acquisition of phonology. In B Dodd (Ed.), *Differential Diagnosis and Treatment of Children with Speech Sound Disorders* (2nd ed., pp. 24–43). Whurr Publishers Ltd.

Dodd, B., Hua, Z., Crosbie, S., & Broomfield, J. (2006). English phonology: Acquisition and disorder. In *Phonological Development and Disorders in Children: A Multilingual Perspective* (pp. 25–55). Tonawanda, NY: Multilingual Matters.

Dodd, B, Holm, A., Hua, Z., & Crosbie, S. (2003). Phonological development: A normative study of British English-speaking children. *Clinical Linguistics and Phonetics*, *17*(8), 617–643. https://doi.org/10.1080/0269920031000111348

Dodd, B & Lacano, T. (1989). Phonological disorders in children: Changes in phonological process use during treatment. *International Journal of Language & Communication Disorders*, *24*(3), 333–352. https://doi.org/10.3109/13682828909019894

Dollaghan, C., Biber, M., & Campbell, T. (1993). Constituent syllable effects in a nonsense-word repetition task. *Journal of Speech and Hearing Research*, *36*(5), 1051–1054. https://doi.org/10.1044/jshr.3605.1051

Dyson, A., & Amayreh, M. M. (2000). Phonological errors and sound changes in Arabic-speaking children. *Clinical Linguistics & Phonetics*, *14*(2), 79–109. https://doi.org/10.1080/026992000298850

Dyson, A. T. (1988). Phonetic inventories of 2- and 3-year-old children. *Journal of Speech and Hearing Disorders*, *53*(1), 89–93.

Dyson, A. T., & Amayreh, M. M. (2000). Phonological errors and sound changes in Arabic-speaking children. *Clinical Linguistics and Phonetics*, *14*(2), 79–109. https://doi.org/10.1080/026992000298850

Eagan, B., Brohinsky, J., Wang, J., & Shaffer, D. W. (2020). Testing the reliability of inter-rater reliability. *ACM International Conference Proceeding Series*, (March), 454–461. https://doi.org/10.1145/3375462.3375508

Edwards, J., & Beckman, M. E. (2008). Methodological questions in studying consonant acquisition. Clinical Linguistics and Phonetics, 22(12), 937–956. https://doi.org/10.1080/02699200802330223

Eisenberg, S. L., & Hitchcock, E. R. (2010). Using Standardized Tests to Inventory. *Language, Speech and Hearing Services in Schools*, *41*(October), 488–503.

El-badarin, M., & Bani-Yasin, R. (1993). Consonant Clusters in a Northern Jordanian Arabic Dialect. *Language Research*, 29(2), 201–242.

Elizabeth D. Pena, T. J. S. E. P. (2006). The Composition of Normative Groups and Diagnostic Decision Making: Shooting Outselves in the Foot, *15*(August 2006), 247–254.

Elrefaie, D. A., Hegazi, M. A. E.-F., El-Mahallawi, M. M., & Khodeir, M. S. (2021). Descriptive analysis of the development of the Arabic speech sounds among typically developing colloquial Egyptian Arabic-speaking children. *The Egyptian Journal of Otolaryngology*, *37*(1). https://doi.org/10.1186/s43163-021-00094-w

Fabiano-Smith, L. (2019). Standardized Tests and the Diagnosis of Speech Sound Disorders. Perspectives of the ASHA Special Interest Groups, 4(1), 58–66. https://doi.org/10.1044/2018_pers-sig1-2018-0018

Fee, E. J. (1995). The phonological system of a specifically language-impaired population. *Clinical Linguistics & Phonetics*, *9*(3), 189–209. https://doi.org/10.3109/02699209508985332

Ferguson, C. A., & Farwell, C. B. (1975). Words and sounds in early language acquisition. *Language*, 419-439.

Ferguson, W. A. (1959). Diglossia. In *Language and Social Context* (pp. 232–251). Oxford: Oxford University Press.

Flipsen Jr, P., & Ogiela, D. A. (2015). Psychometric Characteristics of Single-Word Tests of Children's Speech Sound Production. *Language, Speech and Hearing Services in Schools*, *46*(2), 166–178. https://doi.org/10.1044/2015

Foulkes, P. (2006). Phonological Variation: A Global Perspective. In A. Bas & A. McMahon (Eds.), *The Handbook of English Linguistics* (pp. 625–669). John Wiley & Sons, Incorporated. https://doi.org/10.1002/9780470753002.ch27

Fox-Boyer, Annette, Fricke, S., & Albrecht, K. M. (2020). Phonological Patterns in German in Turkish-German Bilingual Children. *Sprache Stimme Gehor*, *44*(4), 216–223. https://doi.org/10.1055/a-1127-5908

Fox-Boyer, Annette, Lavaggi, S., & Fricke, S. (2021). Phonological variations in typically-developing Italian-speaking children aged 3;0-4;11. *Clinical Linguistics and Phonetics*, 0–19. https://doi.org/10.1080/02699206.2021.1945686

Fox-Boyer, A. V. (2016). *Childish pronunciation disorders* (7 ed). Idstein: Schulz-Kirchner.

Fox, A. V. (2000). The acquisition of phonology and the classification of speech disorders in German-speaking children.[Doctoral thesis]. University of Newcastle Upon Tyne.

Fox, A. V. (2007). German speech acquisition. In S. McLeod (Ed.), *The international guide to speech acquisition* (pp. 386–397). New York: New York: Thomson Delmar Learning.

Fox, A. V. (2006). Evidence from German-Speaking Children. In Z. Hua & B. Dodd (Eds.), *Phonological Development and Disorders in Children: A multilingual perspective* (pp. 56–80). Channel View Publications.

Gangji, N., Pascoe, M., & Smouse, M. (2015). Swahili speech development: Preliminary normative data from typically developing pre-school children in Tanzania. *International Journal of Language and Communication Disorders*, *50*(2), 151–164. https://doi.org/10.1111/1460-6984.12118

Gierut, J. A., Simmerman, C. L., & Neumann, H. J. (1994). Phonemic structures of delayed phonological systems. *Journal of Child Language*, *21*(2), 291–316. https://doi.org/10.1017/S0305000900009284

Glaspey, A., & Stoel-Gammon, C. (2007). A dynamic approach to phonological assessment. *Advances in Speech Language Pathology*, *9*(4), 286–296. https://doi.org/10.1080/14417040701435418

Goldstein, B. A. (2005). Substitution patterns in the phonology of Spanish-speaking children. *Journal of Multilingual Communication Disorders*, *3*(3), 153–168. https://doi.org/10.1080/14769670500096724

Goldstein, B., Fabiano, L., & Iglesias, A. (2004). Spontaneous and Imitated Productions in Spanish-Speaking Children with Phonological Disorders. *Language, Speech, and Hearing Services in Schools*, *35*(1), 5–15. https://doi.org/10.1044/0161-1461(2004/002)

Grech, H. (1998). *Phonological Development of Normal Maltese-Speaking Children*. (Publication No. 10870949) [Doctoral thesis]. University of Manchester]. ProQuest Dissertations Publishing

Grech, H. (2006). Phonological Development of Maltese-Speaking Children. In Zhu Hua & B. Dodd (Eds.), Phonological Development and Disorders in Children: A Multilingual Perspective (1st ed., pp. 135–178). Multilingual Matters Limited.

Gruber, F. A. (1999). Probability estimates and paths to consonant normalization in children with speech delay. *Journal of Speech, Language, and Hearing Research*, *42*(2), 448–459. https://doi.org/10.1044/jslhr.4202.448

Gruwell, P. (1981). The development of phonology: a descriptive profile. In *First Language* (2nd ed., pp. 161–191).

Hall, J. W. (2016). Effective and efficient pre-school hearing screening: Essential for successful early hearing detection and intervention. *Journal of Early Hearing Detection and Intervention*, *1*(1), 2–12.

Hamdan, J. M., & Amayreh, M. M. (2007). Consonant profile of Arabic-speaking school-age children in Jordan. *Folia Phoniatrica et Logopaedica*, *59*(2), 55–64. https://doi.org/10.1159/000098338

Hedlund, G., & Rose, Y. (2019). Phon 3.1.

Hellmuth, S. (2013). Phonology. In O. Jonathan (Ed.), *The Oxford Handbook of Arabic Linguistics*. (pp. 45–70). Oxford University Press.

Hodson, B. W., & Paden, E. P. (1981). Phonological Processes which Characterize Unintelligible and Intelligible Speech in Early Childhood. *Journal of Speech and Hearing Disorders*, *46*(November 1981), 369–373.

Holes, C. (2004). *Modern Arabic*. Washington, D.C.: Georgetown University Press.

Holes, C. (2007). Kuwaiti Arabic. In K. Versteegh, M. Eid, A. Elgibali, M. Woidich, & A. Zaborski (Eds.), *Encyclopaedia of Arabic Language and Linguistics* (pp. 608–620). Netherlands: Brill Academic Publishers.

Holm, A., Sanchez, K., Crosbie, S., Morgan, A., & Dodd, B. (2021). Is children's speech development changing? Preliminary evidence from Australian English-speaking 3-year-olds International Journal of Speech-Language Pathology, 0(0), 1–10. https://doi.org/10.1080/17549507.2021.1991474

Hua, Z., & Dodd, B. (2006). *Phonological Development and Disorders in Children*. Multilingual Matters.

Hua, Z, & Zhu, H. (2002a). Language Universals and Cross- linguistic Studies of Phonological Acquisition. In *Phonological Development in Specific Contexts: Studies of Chinese-speaking Children* (pp. 1–29). Multilingual Matters.

Hua, Z, & Zhu, H. (2002b). Phonological Development in Specific Contexts: Studies of Chinese-*Speaking Children*. Multilingual Matters. Retrieved from https://ebookcentral.proquest.com/lib/sheffield/detail.action?docID=3007711.

Hua, Zhu. (2006). The Need for Comparable Criteria in Multilingual Studies. In Zhu Hua & B. Dodd (Eds.), Phonological development and disorders in children: A multilingual perspective (pp. 15–22). Multilingual Matters.

Ingham, B. (1971). Some characteristics of meccan speech. *Bulletin of the School of Oriental and African Studies*, *34*(2), 273–297. https://doi.org/10.1017/S0041977X00129544

Ingram, D. (1976). *Phonological disability in children* (1st ed.). London: Whurr.

Ingram, D. (1989). First Language Acquisition: Method, description and explanation. New York: Cambridge University Press.

Ingram, D. (1991). An historical observation on "Why 'Mama' and 'Papa'?" *Journal of Child Language*, *18*(3), 711–713. https://doi.org/10.1017/S0305000900011326

IstiZada. (2020). Arabic Speaking Countries List - 2020 Update | IstiZada.

James, D. G. H. (2001a). An item analysis of Australian English words for an articulation and phonological test for children aged 2 to 7 years. *Clinical Linguistics and Phonetics*, *15*(6), 457–485. https://doi.org/10.1080/02699200110045344

James, D. G. H. (2001b). Use of phonological processes in Australian children ages 2 to 7;11 years.

International Journal of Speech-Language Pathology, *3*(2), 109–127. https://doi.org/10.3109/14417040109003718

James, D. G. H., Ferguson, W. A., & Butcher, A. (2016). Assessing children's speech using picture naming: The influence of differing phonological variables on some speech outcomes. *International Journal of Speech-Language Pathology*, *18*(4), 364–377. https://doi.org/10.3109/17549507.2015.1101159

Jarrah, M. A. S. (1993). *The phonology of Madina Hijazi Arabic: a non-linear analysis*. [Doctoral thesis]. University of Essex. Retrieved from http://search.mandumah.com/Record/599918%250

Kehoe, M., Niederberger, N., & Bouchut, A. L. (2021). The development of a speech sound screening test for European French-speaking monolingual and bilingual children. *International Journal of Speech-Language Pathology*, *23*(2), 135–144. https://doi.org/10.1080/17549507.2020.1750699

Kent, R. D., Miolo, G., & Bloedel, S. (1994). The Intelligibility of Children's Speech: A Review of Evaluation Procedures. *American Journal of Speech-Language Pathology*, *3*(2), 81–95. Retrieved from http://ajslp.asha.org/cgi/content/abstract/3/2/81

Khattab, G., & Al-tamimi, J. (2015). the Acquisition of Gemination in Lebanese-Arabic Children. Proceedings of the 18th International Congress of Phonetic Sciences (ICPhS 2015).

Khattab, G., & Al-Tamimi, J. (2012). Influence of geminate structure on early Arabic templatic patterns. The Emergence of Phonology: Whole-Word Approaches and Cross-Linguistic Evidence. https://doi.org/10.1017/CBO9780511980503.018

Khoja, M. A. (2019). A survey of formal and informal assessment procedures used by speechlanguage pathologists in Saudi Arabia. *Speech, Language and Hearing*, *22*(2), 91–99. https://doi.org/10.1080/2050571X.2017.1407620

Kirk, C., & Demuth, K. (2002). Onset / Coda Asymmetries in the Acquisition of Clusters In *Proceedings of the 27th annual Boston University conference on language development* (Vol. 1, pp. 437-448). Cascadilla Press.

Kirk, C., & Vigeland, L. (2014). A psychometric review of norm-referenced tests used to assess phonological error patterns. *Language, Speech, and Hearing Services in Schools*. American Speech-Language-Hearing Association. https://doi.org/10.1044/2014_LSHSS-13-0053

Kirk, C., & Vigeland, L. (2015). Content Coverage of Single-Word Tests Used to Assess Common Phonological Error Patterns. *Language, Speech, and Hearing Services in Schools*, *46*(4), 14–29. https://doi.org/10.1044/2014

Klein, H. B. (1984). Procedure for Maximizing Phonological Information from Single-Word Responses. *Language, Speech, and Hearing Services in Schools*, *15*(4), 267–274. https://doi.org/10.1044/0161-1461.1504.267

Kubaschk, L., Fox-Boyer, A., & Klann, J. (2015). Die Stimulierbarkeit isolierter Phone im Deutschen bei 2;5 – 4;0-Jährigen. *Sprache · Stimme · Gehör*, *39*(S 01), S5–S6. https://doi.org/10.1055/s-

0034-1387612

Languages and dialects in Saudi Arabia. (2020). Retrieved from https://saudiarabiaevisa.org/languages-and-dialects-in-saudi-arabia/

Levy, H., & Hanulíková, A. (2019). Variation in children's vowel production: Effects of language exposure and lexical frequency. *Laboratory Phonology*, *10*(1), 1–26. https://doi.org/10.5334/labphon.131

Leonard, L. (1995). Phonological impairment in children. In P. Fletcher & B. MacWhinney, (Eds.), *Handbook of child language*. Oxford, England: Basil Blackwell.

Limbrick, N., Mc Cormack, J., & Mcleod, S. (2013). Designs and decisions: The creation of informal measures for assessing speech production in children. *International Journal of Speech-Language Pathology*, *15*(3), 296–311. https://doi.org/10.3109/17549507.2013.770552

Lipinski, E. (1997). *Semitic languages: Outline of a Comparative Grammar*. Leuven: Peeters. Low, A., Koh, G., Ee-li Young, S., & Chandler-Yeo, H. (2019). Effect of dialect on phonological analysis. *Clinical Linguistics and Phonetics*, *33*(5), 457–478. https://doi.org/10.1080/02699206.2018.1550812

Macrae, T., Tyler, A. A., & Lewis, K. E. (2014). Lexical and phonological variability in preschool children with speech sound disorder. *American Journal of Speech-Language Pathology*, *23*(1), 27–35. https://doi.org/10.1044/1058-0360(2013/12-0037)

Maphalala, Z., Pascoe, M., & Smouse, M. R. (2014). Phonological development of first language isiXhosa-speaking children aged 3;0-6;0 years: A descriptive cross-sectional study. *Clinical Linguistics and Phonetics*, *28*(3), 176–194. https://doi.org/10.3109/02699206.2013.840860

Marklund, U., Lacerda, F., Persson, A., & Lohmander, A. (2018). The development of a vocabulary for PEEPS-SE—profiles of early expressive phonological skills for Swedish. *Clinical Linguistics and Phonetics*, *32*(9), 844–859. https://doi.org/10.1080/02699206.2018.1445780

Mashaqba, B. M., Al-Shdifat, K. G., Al Huneety, A. I., & Alhala, M. A. (2019). Acquisition of syllable structure in Jordanian Arabic. *Communication Sciences and Disorders*, *24*(4), 953–967. https://doi.org/10.12963/csd.19657

Masterson, J. J., Bernhardt, B. H., & Hofheinz, M. K. (2005). A comparison of single words and conversational speech in phonological evaluation. *American Journal of Speech-Language Pathology*, *14*(3), 229–241. https://doi.org/10.1044/1058-0360(2005/023)

McAllister-Byun, T., & Rose, Y. (2016). Analyzing clinical phonological data using Phon. *Seminars in Speech and Language*, *37*(2), 85–105. https://doi.org/10.1055/s-0036-1580741.Analyzing

McCarthy, J. J. (1994). The Phonetic and phonology of Semitic pharyngeals. In P. Keating (Ed.), *Phonological Structure and Phonetic Form* (pp. 191–233). Cambridge: Cambridge Press.

McCarthy, J. J., & Prince, A. S. (1990). Foot and Word in Prosodic Morphology: The Arabic Broken Plural Author (s): John J. McCarthy and Alan S. Prince Published by : Springer Stable URL : http://www.jstor.org/stable/4047699 Accessed : 25-04-2016 18 : 23 UTC Your use of the

JSTOR archive in, *8*(2), 209–283.

McCauley, R. J., & Strand, E. A. (2008). A Review of Standardized Tests of Nonverbal Oral and Speech Motor Performance in Children. *American Journal of Speech-Language Pathology*, *17*(1), 81–91. https://doi.org/10.1044/1058-0360(2008/007)

McCauley, R. J., & Swisher, L. (1984). Review of Language and Articulation Tests for Preschool Children, *49*(February), 34–42.

McCormack, J., McLeod, S., McAllister, L., & Harrison, L. J. (2009). A systematic review of the association between childhood speech impairment and participation across the lifespan. *International Journal of Speech-Language Pathology*, *11*(2), 155–170. https://doi.org/10.1080/17549500802676859

McHugh, M. L. (2012). Interrater reliability: the kappa statistic. *Biochemia Medica*, *3*(22), 276–282. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3900052/

McIntosh, B., & Dodd, B. (2008). Two-year-olds' phonological acquisition: Normative data. *International Journal of Speech-Language Pathology*, *10*(6), 460–469. https://doi.org/10.1080/17549500802149683

McLeod, S. (2007). *The international guide to speech acquisition*. Clifton Park, NY: Thomson Delmar Learning.

McLeod, S. (2012). 14 Translation to Practice: Creating Sampling Tools to Assess Multilingual Children's Speech, In S. McLeod & B. A. Goldstein (Eds.), *Multilingual aspects of speech sound disorders in children* (pp. 144-153). Multilingual Matters.

McLeod, S., & Baker, E. (2014). Speech-language pathologists' practices regarding assessment, analysis, target selection, intervention, and service delivery for children with speech sound disorders. *Clinical Linguistics and Phonetics*, *28*(7–8), 508–531. https://doi.org/10.3109/02699206.2014.926994

McLeod, S., & Crowe, K. (2018). Children's Consonant Acquisition in 27 Languages: A CrossLinguistic Review. *American Journal of Speech-Language Pathology*. https://doi.org/10.1044/2018_AJSLP-17-0100

McLeod, S., Van Doorn, J., & Reed, V. A. (2001a). Consonant cluster development in two-year-olds: General trends and individual difference. *Journal of Speech, Language, and Hearing Research*, *44*(5), 1144–1171. https://doi.org/10.1044/1092-4388(2001/090)

McLeod, S., Van Doorn, J., & Reed, V. A. (2001b). Normal Acquisition of Consonant Clusters. *American Journal of Speech-Language Pathology*, *10*(2), 99–110. https://doi.org/10.1044/1058-0360(2001/011)

McLeod, S., & Verdon, S. (2014). A Review of 30 Speech Assessments in 19 Languages Other Than English. *AJSLP*, *23*, 708–723. https://doi.org/10.1044/2014_AJSLP-13-0066

McLeod, S., Verdon, S., Bowen, C., Bacsfalvi, P., Crowe, K., Davis, B., … Yavas, M. (2013).

International aspirations for speech-language pathologists' practice with multilingual children with speech sound disorders: Development of a position paper. *Journal of Communication Disorders*, *46*(4), 375–387. https://doi.org/10.1016/j.jcomdis.2013.04.003

McReynolds, L. V., & Elbert, M. (1981). Criteria for Phonological Process Analysis. *Journal Of Speech and Hearing Disorders*, 197–205.

Menn, L. (1981). Theories of Phonological Development. *Annals of the New York Academy of Sciences*, *379*(1), 130–137. https://doi.org/10.1111/j.1749-6632.1981.tb42003.x

Menn, L., Schmidt, E., & Nicholas, B. (2013). Challenges to theories, charges to a model: The Linked-Attractor model of phonological development. In *The Emergence of Phonology: Whole-Word Approaches and Cross-Linguistic Evidence* (pp. 460–502). https://doi.org/10.1017/CBO9780511980503.022

Messick, S. (1980). Test validity and the ethics of assessment. *American Psychologist*, *35*(11), 1012–1027. https://doi.org/10.1037/0003-066X.35.11.1012

Miccio, A. W., Elbert, M., & Forrest, K. (1999). The Relationship between Stimulability and Phonological Acquisition in Children with Normally Developing and Disordered Phonologies. *American Journal of Speech-Language Pathology*, *8*(4), 347–363. https://doi.org/10.1044/1058-0360.0804.347

Morris, S. R., Wilcox, K. A., & Schooling, T. L. (1995). The Preschool Speech Intelligibility Measure. *American Journal of Speech-Language Pathology*, *4*(4), 22–28. https://doi.org/10.1044/1058-0360.0404.22

Morrison, J. A., & Shriberg, L. D. (1992). Articulation testing versus conversational speech sampling. *Journal of Speech and Hearing Research*, *35*(2), 259–273. https://doi.org/10.1044/jshr.3502.259

Nojavan-Pirehyousefan, H., Zarifian, T., Ahmadi, A., Jahan, A., Teymouri, R., & Hosseinzadeh, S. (2021). Development of a speech production test in Azeri Turkish speaking children. *Applied Neuropsychology: Child*, *0*(0), 1–10. https://doi.org/10.1080/21622965.2021.1935255

Nojavan-Pirehyousefan, H., Zarifian, T., Ahmadi, A., & Pascoe, M. (2022). Speech sound acquisition in Azeri Turkish-speaking children in Iran Speech sound acquisition in Azeri Turkish-speaking children in Iran. *Journal of Communication Disorders*, *99*(July), 106244. https://doi.org/10.1016/j.jcomdis.2022.106244

Omar, M. (1973). *The Acquisition of Egyptian Arabic as a Native Language* (1st ed.). Washington, D.C.: Georgetown University Press.

Omar, M. (1975). *Saudi Arabic Urban Hejazi Dialect Basic Course*. Washington DC: Foreign Service Institute.

Owaida, H. (2015). *Speech sound acquisition and phonological error patterns in child speakers of Syrian Arabic: a normative study*. [Unpublished doctoral thesis]. City University London.

Pascoe, M., Mahura, O., & Le Roux, J. (2018). South African English Speech Development:

Preliminary Data from Typically Developing Preschool Children in Cape Town. *Clinical Linguistics* *and Phonetics*, *32*(12), 1145–1161. https://doi.org/10.1080/02699206.2018.1510985

Pearson, B. Z., Velleman, S. L., Bryant, T. J., & Charko, T. (2009). Phonological Milestones for African, *40*(July), 229–245.

Peña, E. D., Spaulding, T. J., & Plante, E. (2006). The composition of normative groups and diagnostic decision making: Shooting ourselves in the foot. *American Journal of Speech-Language Pathology*, *15*(3), 247–254. https://doi.org/10.1044/1058-0360(2006/023)

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. https://doi.org/10.3109/02699206.2015.1095244

Phoon, H. S., Abdullah, A. C., & MacLagan, M. (2012). The effect of dialect on the phonological analysis of Chinese-influenced Malaysian English-speaking children. *International Journal of Speech-Language Pathology*, *14*(6), 487–498. https://doi.org/10.3109/17549507.2012.719549

Plante, E., & Vance, R. (1994). Selection of Preschool Language Tests. *Language, Speech, and Hearing Services in Schools*, *25*(1), 15–24. https://doi.org/10.1044/0161-1461.2501.15

Polit, D. F. (2014). Getting serious about test-retest reliability: A critique of retest research and some recommendations. *Quality of Life Research*, *23*(6), 1713–1720. https://doi.org/10.1007/s11136-014-0632-9

Powell, T. W., & Miccio, A. W. (1996). Stimulability: A useful clinical tool, *29*, 237–253.

Preston, J. L., Hull, M., & Edwards, M. L. (2013). Preschool speech error patterns predict articulation and phonological awareness outcomes in children with histories of speech sound disorders. *American Journal of Speech-Language Pathology*, *22*(2), 173–184. https://doi.org/10.1044/1058-0360(2012/12-0022)

Prezas, R. F., Hodson, B. W., & Schommer-Aikins, M. (2014). Phonological Assessment and Analysis of Bilingual Preschoolers' Spanish and English Word Productions. *Raúl*, *23*(May), 176–185. https://doi.org/10.1044/2013

Priester, G. H., Post, W. J., & Goorhuis-Brouwer, S. M. (2011). Phonetic and phonemic acquisition: Normative data in English and Dutch speech sound development. *International Journal of Pediatric Otorhinolaryngology*, *75*(4), 592–596. https://doi.org/10.1016/j.ijporl.2011.01.027

Ragheb, M., & Davis, S. (2014). On the L1 development of final consonant clusters in Cairene Arabic. In R. Khamis-Dakwar & K. Froud (Eds.), Perspectives on Arabic Linguistics XXVI: Papers from the annual symposium on Arabic Linguistics (pp. 263–282). New York: John Benjamin's Publishing Company.

Raykov, T., & Marcoulides, G. A. (2010). *Introduction to Psychometric Theory*. Retrieved from https://ebookcentral.proquest.com/lib/sheffield/reader.action?docID=668336&ppg=198

Roberts, J. E., Burchinal, M., & Footo, M. M. (1990). Phonological process decline from 2 1 2 to 8 years. *Journal of Communication Disorders*, *23*(3), 205–217. https://doi.org/10.1016/0021-

9924(90)90023-R

Rose, Y., & MacWhinney, B. (2014). The Phonbank Project: Data and software-assisted methods for the study of phonology and phonological development. *The Oxford Handbook of Corpus Phonology*, 380–401.

Rvachew, S., & Andrews, E. (2002). The influence of syllable position on children's production of consonants. *Clinical Linguistics and Phonetics*, *16*(3), 183–198. https://doi.org/10.1080/02699200110112222

Ryding, K. C. (2005). Arabic word structure: an overview. In *A Reference Grammar of Modern Standard Arabic* (pp. 44–56). Cambridge University Press. https://doi.org/https://doi.org/10.1017/CBO9780511486975.004

Ryding, K. C. (2010). Arabic linguistics: overview and history, *1*(3), 1–12.

Saidat, A. M., & Khlifat, J. A. (2019). Phonetics and Phonology Paradox in Levantine Arabic: An Analytical Evaluation of Arabic Geminates' Hypocrisy. *Theory and Practice in Language Studies*, *9*(7), 854. https://doi.org/10.17507/tpls.0907.16

Saiegh-Haddad, E. (2004). The impact of phonemic and lexical distance on the phonological analysis of words and pseudowords in a diglossic context. *Applied Psycholinguistics*, *25*(4), 495–512. https://doi.org/10.1017/s0142716404001249

Saiegh-Haddad, E., & Haj, L. (2018). Does phonological distance impact quality of phonological representations? Evidence from Arabic diglossia. *Journal of Child Language*, *45*(6), 1377–1399. https://doi.org/10.1017/S0305000918000302

Saiegh-Haddad, E., Shahbari-Kassem, A., & Schiff, R. (2020). Phonological awareness in Arabic: the role of phonological distance, phonological-unit size, and SES. *Reading and Writing*, *33*(6), 1649–1674. https://doi.org/10.1007/s11145-020-10019-3

Saleh, M., Shoeib, R., Hegazi, M., & Ali, P. (2007). Early phonological development in Arabic Egyptian children: 12-30 Months. *Folia Phoniatrica et Logopaedica*, *59*(5), 234–240. https://doi.org/10.1159/000104461

Sander, E. K. (1972). Are Speech Sounds Learned? *Journal Of Speech and Hearing Disorders*, (1957), 55–64.

Savinainen-Makkonen, T. (2007). Geminate template: A model for first Finnish words. *FirstLanguage*, *27*(4), 347–359. https://doi.org/10.1177/0142723707081728

Schaefer, B., Fricke, S., Szczerbinski, M., Fox-Boyer, A. V., Stackhouse, J., & Wells, B. (2009). Development of a test battery for assessing phonological awareness in German-speaking children. *Clinical Linguistics and Phonetics*, *23*(6), 404–430 https://doi.org/10.1080/02699200902770187

Schwartz, R. G. (1992). Clinical Applications of Recent Advances in Phonological Theory. *Language, Speech, and Hearing Services in Schools*, *23*(3), 269–276. https://doi.org/10.1044/0161-1461.2303.269

Shamsan, M. A.-H. A., & Attayib, A. (2015). Inflectional Morphology in Arabic and English: A Contrastive Study. *International Journal of English Linguistics*, *5*(2). https://doi.org/10.5539/ijel.v5n2p139

Shriberg, L. D. (1993). Four new speech and Prosody-Voice measures for genetics research and other studies in developmental phonological disorders. *Journal of Speech and Hearing Research*, *36*(1), 105–140. https://doi.org/10.1044/jshr.3601.105

Shriberg, L. D., Gruber, F. A., & Kwiatkowski, J. (1994). Developmental phonological disorders III: Long-term speech-sound normalization. *Journal of Speech and Hearing Research*, *37*(5), 1151–1177.

Shriberg, L. D., & Kwiatkowski, J. (1982). Phonological disorders. III: A procedure for assessing severity of involvement. *Journal of Speech and Hearing Disorders*, *47*(3), 256–270. https://doi.org/10.1044/jshd.4703.256

Shriberg, Lawrence D., Austin, D., Lewis, B. A., McSweeny, J. L., & Wilson, D. L. (1997a). The percentage of consonants correct (PCC)metric: Extensions and reliability data. *Journal of Speech, Language, and Hearing Research*, *40*(4), 708–722. https://doi.org/10.1044/jslhr.4004.708

Shriberg, Lawrence D., Austin, D., Lewis, B. A., McSweeny, J. L., & Wilson, D. L. (1997b). The speech disorders classification system (SDCS): Extensions and lifespan reference data. *Journal of Speech, Language, and Hearing Research*, *40*(4), 723–740. https://doi.org/10.1044/jslhr.4004.723

Shweta, Bajpai, R. C., & Chaturvedi, H. K. (2015). Evaluation of inter-rater agreement and inter-rater reliability for observational data: An overview of concepts and methods. *Journal of the Indian Academy of Applied Psychology*, *41*(Special Issue 3), 20–27.

Skahan, S. M., Watson, M., & Lof, G. L. (2007). Speech-language pathologists' assessment practices for children with suspected speech sound disorders: Results of a national survey. *American Journal of Speech-Language Pathology*, *16*(3), 246–259. https://doi.org/10.1044/1058-0360(2007/029)

Smit, A. B., Hand, L., Freilinger, J. J., Bernthal, J. E., & Bird, A. (1990). The Iowa Articulation Norms Project and its Nebraska replication. *Journal of Speech and Hearing Disorders*, *55*(4), 779–798. https://doi.org/10.1044/jshd.5504.779

Smit, Ann B. (1987). Ages of Speech Sound Acquisition: Comparisons and Critiques of Several Normative Studies. *Language, Speech, and Hearing Services in Schools*, *18*(3), 288–288. https://doi.org/10.1044/0161-1461.1803.288

Smit, Ann B. (1993). Phonologic error distributions in the Iowa Nebraska articulation norms project: consonant singletons. *Journal of Speech and Hearing Research*, *36*, 533–547.

Smith, B. L., & Anderson, R. (1987). Phonological development of two-Year-Old monolingual Puerto Rican Spanish-Speaking children*. *Journal of Child Language*, *14*(1), 57–78. https://doi.org/10.1017/S0305000900012733

Smith, B. L., Mcgregor, K. K., & Demille, D. (2006). Phonological development in lexically precocious 2-year-olds. *Applied Psycholinguistics*, *27*(3), 355–375. https://doi.org/10.1017/s0142716406060310

Smith, F. (1973). *Psycholinguistics and reading*. Holt, Rinehart & Winston.

So, L. K., & Dodd, B. J. (1994). Phonologically disordered Cantonese-speaking children. *Clinical Linguistics & Phonetics*, *8*(3), 235-255.

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. https://doi.org/10.1044/1092-4388(2011/10-0113)

Stackhouse, J, & Wells, B. (1997). Children's Speech and Literacy Difficulties: A Psycholinguistic Framework. London: Whurr.

Stampe, D. (1969). The acquisition of phonetic representation. *Papers from Vth Regional Meeting*, 443–454.

Stockman, I. J., & Stephenson, L. W. (1981). Children's articulation of medial consonant clusters: Implications for syllabification. *Language and Speech*, *24*(2), 185–204.

Stoel-Gammon, C. (1985). Phonetic inventories, 15-12 months: A longitudinal study. *Journal of Speech and Hearing Research*, *28*(4), 505–512. https://doi.org/10.1044/jshr.2804.505

Stoel-Gammon, Carol. (1987). Phonological Skills of 2-Year-Olds. *Language, Speech, and Hearing Services in Schools*, *18*(4), 323–329. https://doi.org/10.1044/0161-1461.1804.323

Stoel-Gammon, Carol. (2011). Relationships between lexical and phonological development in young children. *Journal of Child Language*, *38*(1), 1–34. https://doi.org/10.1017/S0305000910000425

Stoel-Gammon, Carol, & Dunn, C. (1985). *Normal and disordered phonology in children.* Baltimore: University Park Press.

Stoel-Gammon, Carol, & Sosa, A. V. (2007). Phonological Development. In E. Hoff & M. Shatz (Eds.), *Blackwell Handbook of Language Development* (pp. 238–254). https://doi.org/https://doi.org/10.1002/9780470757833.ch12

Stoel-Gammon, Carol, & Williams, A. L. (2013). Early phonological development: Creating an assessment test. *Clinical Linguistics and Phonetics*, *27*(4), 278–286. https://doi.org/10.3109/02699206.2013.766764

Stokes, S. F., Klee, T., Carson, C. P., & Carson, D. (2005). A phonemic implicational feature hierarchy of phonological contrasts for English-speaking children. *Journal of Speech, Language, and Hearing Research*, *48*(4), 817–833. https://doi.org/10.1044/1092-4388(2005/057)

Storkel, H. L. (2019). Using Developmental Norms for Speech Sounds as a Means of Determining Treatment Eligibility in Schools. *Perspectives of the ASHA Special Interest Groups*, *4*(1), 67–

75. https://doi.org/10.1044/2018_pers-sig1-2018-0014

Topbas, S. (1997). Phonological acquisition of Turkish children: Implications for phonological disorders. *International Journal of Language and Communication Disorders*, *32*(4), 377–396. https://doi.org/10.3109/13682829709082255

Tranter, N. (2000). *The phonology of English loan-words in Korean*. *Word*. University of Sussex. https://doi.org/10.1080/00437956.2000.11432504

Tyler, A. A., & Macrae, T. (2010). Stimulability: Relationships to other characteristics of children's phonological systems. *Clinical Linguistics and Phonetics*, *24*(4–5), 300–310. https://doi.org/10.3109/02699200903532516

Tyler, A. A., & Tolbert, L. C. (2002). Speech-language assessment in the clinical setting. *American Journal of Speech-Language Pathology*, *11*(3), 215–220. https://doi.org/10.1044/1058-0360(2002/022)

Van Haaften, L., Diepeveen, S., van den Engel-Hoek, L., de Swart, B., & Maassen, B. (2020). Speech sound development in typically developing 2–7-year-old Dutch-speaking children: A normative cross-sectional study. *International Journal of Language and Communication Disorders*, *55*(6), 971–987. https://doi.org/10.1111/1460-6984.12575

Velleman, S. L., & Pearson, B. Z. (2010). Differentiating speech sound disorders from phonological dialect differences. *Topics in Language Disorders*, *30*(3), 176–188. https://doi.org/10.1097/TLD.0b013e3181efc378

Velleman, S. L., & Vihman, M. M. (2002). Whole-Word Phonology and Templates: Trap, Bootstrap, or Some of Each?, *33*(January), 9–23.

Versteegh, K. (2006). Encyclopaedia of Arabic language and linguistics. In *Encyclopaedia of Arabic language and linguistics* (1st ed.). Leiden, The Netherlands: Brill.

Vihman, M., & Croft, W. (2007). Phonological development: toward a "radical" templatic phonology. *Linguistics*, 683–725.

Vihman, M. M., & Greenlee, M. E. L. (1987). Individual Differences Development: Ages One. *Journal of Speech and Hearing Research*, *30*(December), 503–521.

Vihman, M. (2007). Phonological development: toward a "radical" templatic phonology. *Linguistics*, 683–725.

Vihman, M. M. (2014). *Phonological development: the first two years* (Second Ed.). Wiley-Blackwell.

Vogel Sosa, A., & Bybee, J. (2011). A Cognitive Approach to Clinical Phonology. In *The Handbook of Clinical Linguistics* (pp. 480–490). Oxford: Wiley-Blackwell.

Waring, R., & Knight, R. (2013). How should children with speech sound disorders be classified? A review and critical evaluation of current classification systems. *International Journal of Language and Communication Disorders*, *48*(1), 25–40. https://doi.org/10.1111/j.1460-6984.2012.00195.x

Waterson, N. (1971). Child phonology: a prosodic view. *Linguistics*, *7*(11), 179–221.

Watson, J. C. (2002). Phonology and Morphology of Arabic, The. The Phonology of the World's Languages (1st ed). Oxford: Oxford University Press.

Watson, J. C. E. (2011). Word Stress in Arabic. *The Blackwell Companion to Phonology*. Retrieved from http://usir.salford.ac.uk/17633/1/135_watson.pdf

Watson, M. M., & Scukanec, G. P. (1997). Profiling the phonological abilities of 2-year-olds: A longitudinal investigation. *Child Language Teaching and Therapy*, *13*(1), 3–14. Retrieved from https://doi.org/10.1177/026565909701300102

Werker, J. F., Hall, D. G., & Fais, L. (2004). Reconstruing U-shaped functions. *Journal of Cognition and Development*, *5*(1), 146-151.

Wertzner, H. F., Sotelo, M. B., & Amaro, L. (2005). Analysis of distortions in children with and without phonological disorders. *Clinics (São Paulo, Brazil)*, *60*(2), 93–102. https://doi.org/10.1590/S1807-59322005000200004

Wolk, L., & Meisler, A. W. (1998). Phonological assessment: A systematic comparison of conversation and picture naming. *Journal of Communication Disorders*, *31*(4), 291–313. https://doi.org/10.1016/S0021-9924(97)00092-0

Yeh, L. L., & Liu, C. C. (2021). Comparing the informativeness of single-word samples and connected speech samples in assessing speech sound disorders. *Journal of Speech, Language, and Hearing Research*, *64*(11), 4071-4084. http://pubs.asha.org/doi/10.1044/2021_JSLHR-20-00172

9 Appendices

Appendix A

A Narrative description of Arabic consonants

Bilabial. In UHA there are two bilabials /b/ and /m/, plus one bilabial approximant /w/ (Basalamah, 1990). This is similar to MSA, except for /w/, which was described by Al-Ani (1970) as a bilabial sonorant with a vowel–like formant, while Watson (2002) treated it as a labio–velar consonant. Other Arabic dialects also match UHA in these two bilabials /b/ and /m/; however, in Kuwaiti and Egyptian Arabic, /w/ is described as a velar approximant, while in Najdi and Jordanian Arabic it is a bilabial glide (Al Qahtani, 2014; Al Qenaie, 2011; Amayreh & Dyson, 1998; Ammar & Morsi, 2006).

Labiodental. The voiceless labiodental fricative /f/ is maintained in all Arabic varieties, as confirmed by all Arabic grammarians.

Dental. This place of articulation is missing in UHA. MSA has a full set of dental fricative consonants: voiceless /θ/, voiced /ð/, and pharyngeal (emphatic) /ðˤ/. Similar to UHA, Egyptian and Syrian Arabic also do not have this dental place of articulation in their phonetic system. However, Holes (2004) noted that the Bedouin Arabic dialects preserve the interdental fricatives, while at the same time the voiced emphatic alveolar plosive /dˤ/ is realised as a voiced emphatic dental fricative /ðˤ/. For example, in the Bedouin Jordanian dialect /dˤ/ does not exist and merging with /ðˤ/, similar to Najdi Arabic.

Alveolar. This classification includes most of the Arabic consonants incorporated in UHA, MSA, and the other dialects (/t, d, tˤ, dˤ, n, r, s, sˤ, z, l/). However, UHA includes one more emphatic alveolar fricative, /zˤ/, which is an allophonic variant of the MSA voiced dental fricative /ðˤ/, similar to Syrian and Egyptian Arabic. Another consonant that can be emphasised in UHA, MSA, and the other dialects is /l/, which is found as an emphatic alveolar lateral in only one word, /ʔaɫɫaːh/ (‘God’), and in some of its variants (e.g. /waɫɫaːh/ [I swear to God]) (Al-Ani, 1970).

Postalveolar. In UHA, there is one consonant that holds this place of articulation, the voiceless postalveolar fricative /ʃ/, similar to MSA and other Arabic dialects. In addition, the phoneme /ʒ/ or /dʒ/, is considered a unique consonant in the phonemic system of Arabic. For UHA, the phoneme /ʒ/ had also different description among Arabic linguists. Basalamah (1990) considers such phoneme a voiced palate-alveolar affricative, but Al Oufi (2016), who studied the phonology of UHA in loanwords, considers the classification of /ʒ/ as a voiced palate-alveolar fricative. Ingham (1971) provided a detailed description of the manner of allophonic variation of this phoneme as it occurs in UHA according to word position and the adjacent consonants:

- Affricate /dʒ/: in syllable initial position, in the word-final position, and when preceding the voiceless fricatives /h, ħ, s, f/; • Fricative /ʒ/: before the plosive /b, d, t/.
- Plosive /ɟ/: before alveolar /l, z, n/.

In MSA, many Arabic grammarians have described this phoneme as a voiced palatal plosive /ɟ/, with the allophonic variation of the voiced alveopalatal affricate /dʒ/ (Al-Ani, 1970; Alghamdi, 2015; Brame, 1970; Ingham, 1971). Al Nassir (1985) described this consonant as a sound 'that shows a great deal of variation' (p.78). In the Egyptian dialect, as it is clear from Table 2.3, this phoneme has a different representation, where it is replaced mainly by the velar plosive [g]. But in some loanwords, it is realised as the palatal fricative [j].

Palatal. The palatal consonants in UHA, MSA, and the other dialects are similar. They have only one voiced palatal approximate: /j/. Like /w/, the consonant /j/ has a vowel–like formant structures (Al- Ani, 1970).

Velar and uvular. The velar consonants are the consonants that differ the most markedly between MSA and UHA. In UHA, there is a voiceless plosive velar /k/ and a voiced plosive velar /g/, with an absence of the uvular place of articulation. As illustrated in Table 2.3, the /g/ in UHA is an allophonic replacement for the MSA voiced uvular plosive /q/. There are also the voiced and voiceless velar fricatives /x, ɣ/ (Basalamah, 1990; Tranter, 2000). In MSA, there is only the velar /k/, while /q/ is the uvular plosive phone (Al Nassir, 1985) and there are two uvular fricatives /χ, ʁ/ (Alghamdi, 2015; Watson, 2002), which are the allophonic variants of the UHA /x, ɣ/ (see Table 2.3). Velar and uvular consonants in other dialects are similar to UHA, except for Egyptian Arabic which has the voiced plosive velar /g/ but as a different phoneme (i.e. a replacement for the MSA voiced palatal affricate /dʒ/) (see Table 2.3). The uvular plosive /q/ is also absent from the phonetic system of the other dialects; it is replaced with the glottal plosive /ʔ/ in Egyptian, Jordanian and Syrian Arabic (Al Nassir, 1985), while, similar to UHA, it is replaced by /g/ in Kuwaiti and Najdi Arabic.

Pharyngeal and glottal. These two classifications are similar for UHA, MSA, and the other dialects, where all varieties have the voiceless and voiced pharyngeal fricatives /ħ, ʕ/, as well as the glottal plosive /ʔ/ and the glottal fricative /h/. However, there are some differences between UHA, MSA, and the other dialects in terms of the occurrence of the glottal plosive /ʔ/. In UHA and the other dialects, the medial glottal plosive /ʔ/ used to be assimilated with the vowel that proceeds it. For example, the MSA words

/raʔs/ (‘head’) and /biʔr/ (‘well’) are pronounced as /raːs/ (‘head’) and /biːr/ (‘well’) in UHA and the other dialects. This glottal plosive /ʔ/ is deleted in the final position in UHA and the other dialects, as in /da.waʔ/ (‘medicine’), with some exceptions in a few words, such as the word /hu.duːʔ/ (‘quiet’), where it is apparently pronounced in UHA (Ingham, 1971).

Appendix B

Table 1 *Allophonic variation of short vowels (as reported by Al-Ani (1970) for MSA and Ingham (1971) for UHA)*

Vowel	MSA	Environment	UHA	Environment
High front /i/	/ɨ/	Next to emphatic	/ɪ/	Next to emphatic
	/ɪ/	Near the pharyngeals	/ɛ/	Near the pharyngeals
	/i/	Elsewhere	/i/	Elsewhere
High back rounded /u/	/ʊ/	Next to emphatic	/u/	Next to emphatic
	/u/	Elsewhere	/ɔ/	Near the pharyngeals
Low central /a/	/ə/	Word-final and next to a non–emphatic.	/a/	Next to palatal /j, ʒ/
	/ɑ/	Next to emphatic	/ɑ/	Next to emphatic
	/ʌ/	Near the pharyngeals	/æ/	Near the pharyngeals
	/a/	Elsewhere	/ʌ/	Elsewhere

Table 2 *Allophonic variation of long vowels (as reported in Alani (1970) for MSA and in Ingham (1971) for UHA)*

Vowel	MSA	Phonetic context	UHA	Phonetic context
High front /iː/	/ɨː/	Next to emphatic	/iː/	Next to emphatic
	/ɪː/	Near the pharyngeals		Near the pharyngeals
	/iː/	Elsewhere		Elsewhere
High back rounded /uː/	/ʊː/	Next to emphatic	/uː/	Next to emphatic
	/uː/	Elsewhere		Near the pharyngeals
Low central /aː/	/ʌː/	Near the pharyngeals	/æː/	Near the pharyngeals
	/ɑː/	Next to emphatic + /q/ & /r/	/ɑː/	Next to emphatic
	/aː/	Elsewhere	/ʌː/	Near palatal and elsewhere
Diphthongs	/au/	Elsewhere	/oː/	Elsewhere
	/ai/	Elsewhere	/eː/	Elsewhere

Table 3 *Vocalic variation in Arabic dialects*

	MSA	Egyptian	Jordanian	Syrian	Kuwaiti	Saudi Najdi	UHA
Monophthongs	/a, aː, i, iː, u, uː/	/a, aː, i, iː, u, uː/	/a, aː, i, iː, u, uː/	/a, aː, i, iː, u, uː/	/a, aː, i, iː, u, uː/	/a, aː, i, iː, u, uː/	/a, aː, i, iː, u, uː/
Diphthongs	/au, ai/	/eː, oː/	/eː, oː/	/eː, oː/	/eː, oː/	/eː, oː/	/eː, oː/
Allophonic variants		/eː/ → /e/ /oː/ → /o/ /aː/ → /æː/ /a/ → /æ/	/a/ → /ɑ/	/eː/ → /e/ /oː/ → /o/ /a/ → /ɑ/	/a/ → /ɑ/ /a/ → /ɒ/ /i/ → /e/ /ɛː/→ /ə/	/aː/ → /ɑː/ /a/ → /ɑ/	/i/ → /ɪ/ /i/ → /ɛ/ /a/ → /æ/ /a/ → /ʌ/ /a/ → /ɑ/ /aː/ → /æː/ /aː/ → /ʌː/

Appendix C

Comparison of the psychometric criteria of the available phonology assessment tools for Arabic-speaking children (including Jordanian, Egyptian, and Saudi Arabic) (adopted from Albrecht, 2017)

Criteria			Assessments		
			AAT Jordanian	**MAT** Egyptian	**JAT** Saudi
Linguistic criteria	**Phonemes**	**All included?**	Yes (all MSA)	No; /q, ðˤ θ ð dʒ/ are missing	Yes (all MSA)
		All possible word positions?	Not for all consonants (5/28)	Yes	Yes
		≥ 4 – 5 occurrences per included consonant?	No; /ʔ ð ʁ w j/ x2 All the other x3	No; All consonants x3	No; /k, dˤ/ x2 All the other x3
	Syllables	**Test items' number of syllables?**	1 – 3	1 – 4	1 – 3
		Consonant clusters?	No	No	No
	Phonological patterns	**≥ 5 occurrences for all typical patterns?**	NA	NA	NA
		≥ 5 occurrences for all atypical patterns?	NA	NA	NA
	Vocabulary	**Appropriate for target population?**	Evidence N/A	Evidence N/A	Evidence N/A
	Item difficulty	**Item familiarity analyses?**	NA	Yes	NA
		Different syllable complexities?	Yes	Yes	Yes
Psychometric criteria	**Validity**	**What type of analyses are conductible?**	Phonemic inventory	Phonetic inventory	phonemic inventory
		Cultural–specific words	Yes; only 2	Yes; 6 words	No
		Cultural–sensitive words	No	No	No
		Cultural–specific pictures	No	Yes; only 2	No
		Cultural–sensitive pictures	No	No	No
	Reliability	**Test–retest reliability**	Yes; N= 18, Agreement= 83% for children scoring.	Yes	NA
		Inter–rater reliability **Transcriptions** **Classification**	NA	NA	NA
		Intra–rater reliability **Transcriptions** **Classification**	NA	NA	NA
		Internal (item) consistency	NA	NA	Cronbach's Alpha: >0.96
	Norms	**Normative investigation?**	Yes (N=180)	Yes (*N* = 100)	Yes (*N* = 506), with a subgroup of children with SSD (N=352) but analysis procedures non–transparent

Appendix D

The University Of Sheffield.

Downloaded: 08/07/2018
Approved: 06/07/2018

Deema Turki
Registration number: 170117258
Human Communication Sciences
Programme: Human Communication Science

Dear Deema

PROJECT TITLE: Phonological Development in Monolingual Saudi Hijazi-speaking Children.
APPLICATION: Reference Number 020204

On behalf of the University ethics reviewers who reviewed your project, I am pleased to inform you that on 06/07/2018 the above-named project was **approved** on ethics grounds, on the basis that you will adhere to the following documentation that you submitted for ethics review:

- University research ethics application form 020204 (dated 29/06/2018).
- Participant information sheet 1047955 version 1 (29/06/2018).
- Participant information sheet 1044585 version 2 (27/05/2018).
- Participant consent form 1047956 version 1 (29/06/2018).
- Participant consent form 1044586 version 2 (27/05/2018).

If during the course of the project you need to deviate significantly from the above-approved documentation please inform me since written approval will be required.

Yours sincerely

Traci Walker
Ethics Administrator
Human Communication Sciences

Appendix E

Parents Questionnaire

Dear Parent,

Thank you for signing the consent form and approving to take part in this study. In this questionnaire, there are some brief questions about your child's communication skills and some background information that is needed by the researcher to check the participants are representative. The time required to fill this survey is only 10 to 20 min. Kindly try to fill all the area with the appropriate answers. Feel free for not answering any of the survey questions, and/or withdraw from participating in the project at any time. Your cooperation is highly appreciated.

A) Language/ dialect exposure:

<table>
<tr><td>• Is Hijazi the primary dialect you use in your home?</td><td>□Yes □ No</td></tr>
<tr><td>• Is your child exposed to standard modern Arabic (MSA)?</td><td>□Yes □ No</td></tr>
<tr><td colspan="2">• Could you please put a cross on one of the vertical line to answer the following questions?
1) How much your child is exposed to MSA and Hijazi dialect from you (his/her parent and family members at home)?

Always Hijazi dialect |--------------------|------------------|-------------------|--------------------| Always MSA
Never MSA Almost Hijazi/rarely MSA 50% Almost MSA/rarely Hijazi Never Hijazi</td></tr>
<tr><td colspan="2">2) How much your child is exposed to MSA and Hijazi dialect from media (TV, YouTube, Videos, laptop…etc.)?

Always Hijazi dialect |--------------------|------------------|-------------------|--------------------| Always MSA
Never MSA Almost Hijazi/rarely MSA 50% Almost MSA/rarely Hijazi Never Hijazi</td></tr>
<tr><td colspan="2">3) How much your child is exposed to MSA and Hijazi dialect in school?

Always Hijazi dialect |--------------------|------------------|-------------------|--------------------| Always MSA
Never MSA Almost Hijazi/rarely MSA 50% Almost MSA/rarely Hijazi Never Hijazi</td></tr>
<tr><td colspan="2">4) How much your child is exposed to MSA and Hijazi dialect in other daily activities (e.g. sport club, after school club…etc.)?

Always Hijazi dialect |--------------------|------------------|-------------------|--------------------| Always MSA
Never MSA Almost Hijazi/rarely MSA 50% Almost MSA/rarely Hijazi Never Hijazi</td></tr>
</table>

B) Language development:

<table>
<tr><td>• Do you believe that your child's understanding language skills are typically developing? In another word, do you feel your child can understand and follow your instructions?</td><td>□Yes □No</td></tr>
</table>

If no, would you state your concern about your child's understanding skills briefly in one sentence:	
• Do you believe that your child knows enough vocabularies for his/her age?	□Yes □No
If no, would you state your concern about your child's vocabulary development briefly in one sentence:	
• Do you believe that your child produces appropriate sentences or express his needs appropriately according to his/her age?	□Yes □ No
If no, would you state your concern about your child's sentences development briefly in one sentence:	
• Do you believe that your child's pronunciation skills are typically developing? For example, can he/she produce variety of speech sounds that is appropriate to his/her age?	□Yes □ No
If no, could you state your concern about your child's pronunciation skill:	

C) Oral motor skills:

• Does your child have any problems in his /her oral motor structures? For example, does he/she have any difficulties moving these/or one of these structures (lips. tongue, or jaw)?	□Yes □ No
If yes, could you tick which one of the oral structure that has a problem:	□Lips □Teeth □Tongue □Jaw
• Does this problem affect your child's speech production skills?	□Yes □ No
If yes, to what extent do you believe this problem affect your child's intelligibility? a) severely affected b) moderately affected c) slightly affected d) Not affected at all	

D) Hearing skill:

Does your child suffer from frequent ear infections?	□Yes □No
If yes, do you feel this problem affects his hearing skills?	□Yes □No
Does your child have a diagnosed hearing problem?	□Yes □No
If yes, could you briefly describe his/her problem and how you manage it:	

E) Demographic information:

Child Name:	Date of Birth: ___ day, ____month, ________year____
Age:	Grade:
Place of Born:	
Mother Name:	Father Name:
Number of Siblings:	Siblings' Age:

Mother/female carer level of education: □ Some high school, no diploma □ high school □ diploma or the equivalent □ Some college credit, no degree □ bachelor's degree □ Postgraduate education	Mother/female carer's age: □ 18-24 years old □ 25-34 years old □ 35-44 years old □ 45-54 years old □ 55 or older	Occupation: (if any) Job title:
Father/male carer level of education: □ Some high school, no diploma □ high school □ diploma or the equivalent □ Some college credit, no degree □ Bachelor's degree □ Postgraduate education	Father/male carer's age: □ 18-24 years old □ 25-34 years old □ 35-44 years old □ 45-54 years old □ 55 or older	Occupation: (if any) Job title:

Home address: House type: □Flat □ Villa District: Nearby:	Email address: Telephone:

Who does the child live with: □ both parents □ mother □ father □ other (specify) ___________________________

Thank you for your participation in this project. It would be very helpful if you could return this questionnaire to your child's kindergarten within one week or two. If you have any question, please do not hesitate to contact the researcher or her supervisors. All the contact details are below:

Deema Turki Human Communication Sciences University of Sheffield S10 2TA Telephone: +966538366222 DFTurki1@sheffield.ac.uk	Dr Silke Fricke Human Communication Sciences University of Sheffield S10 2TA s.fricke@sheffield.ac.uk	Dr Traci Walker Human Communication Sciences University of Sheffield S10 2TA traci.walker@sheffield.ac.uk

Appendix F

The exclusion reasons for children with signed consents forms

Reasons for exclusion	Sources	N
1. Saturation of age ranges by testing day	On testing day	18
2. Children were not responding during the test.	During testing session	12
3. The child was absent on testing day.	On testing day	6
4. Testing incomplete	During testing session	23
5. Language, hearing, or oral structure problems were reported.	Parent questionnaire	10
6. The children were bilingual (Arabic/English).	Parent questionnaire	7
7. Hejazi dialect was not the dialect of the parents.	Parent questionnaire	19
Total		**95**

Appendix G

Example of the word Excel matrix including all the initial phonological features targeted for SHAPA

UHA Consonant	Frequency	Arabic Word	Hejazi IPA	MSA IPA	English meaning	SIWI	Phonetic shape	SIWW	Phonetic shape	SFWW	Phonetic shape	SFWF	Phonetic shape
/n/	2	نخلة	nax.la	nax.lah	palm tree	/n/	**C**VC						
	2	نجمة	naʒ.ma	naʒ.mah	star	/n/	**C**VC						
	1	نمر	ni.mir	namir	tiger	/n/	CV						
	2	بطانية	batˤ.tˤa:.ni.ja	batˤ.tˤa:.ni.jah	blanket			/n/	CV				
	2	منشفة	min.ʃa.fa	min.ʃa.fah	towel					/n/	CVC		
	3	منديل	man.di:l	man.di:l	tissue					/n/	CVC		
	2	ألوان	ʔal.wa:n	ʔal.wa:n	coloring pencils							/n/	CVV**C**
	3	صحن	sˤa.ħin	sˤaħn	plate							/n/	CV**C**
	3	صابون	sˤa:.bu:n	sˤa:.bu:n	soap							/n/	CVV**C**
/m/	2	موز	mu:z	mawz	bananas	/m/	CVVC						
	1	ملح	milħ	milħ	Salt	/m/	CVCC						
	2	منديل	man.di:l	man.di:l	tissue	/m/	CVC						
	2	مدرسة	mad.ra.sa	mad.ra.sah	school	/m/	CVC						
	3	سمك	sa.mak	sa.mak	fish			/m/	CVC				
	3	ليمون	laj.mu:n	laj.mu:n	lemon			/m/	CVVC				
	2	شمعة	ʃam.ʕa	ʃam.ʕah	candle					/m/	CVC		
	2	لمبة	lam.ba	lam.bah	lamp					/m/	CVC		
	2	طماطم	tˤa.ma:.tˤim	tˤa.ma:.tˤim	tomato			/m/				/m/	CVC
	2	هرم	ha.ram	ha.ram	pyramid							/m/	CVC
	2	خشم	xu.ʃum	ʔanf	nose							/m/	CVC

Appendix H

shutterstock

Invoice/Receipt

Order ID: **SSTK-07A3D-E6F9**

Billed from:

Shutterstock Netherlands, B.V.
Hoogte Kadijk 39
1018 BE Amsterdam
The Netherlands
Registered with Dutch Chamber of Commerce under Trade Register number 61173851

EIN: 80-0812659

Billed to:

DEEMAH F TURKI
C23, ATLANTIC ONE
16 ST GEORGE CLOSE
SHEFFIELD, SHEFFIELD S3 7AN
United Kingdom

User ID: 195572016

Purchase Date
6 Nov 2018

Payment Method
MC ending in 9292

Payment Status
Paid

Qty	Description	EUR Approx.	Amount
1	1-Month Subscription, Standard License with 350 Downloads per Month	€104.13	£119.00
	TOTAL:	**€104.13**	**£119.00**

Please note: Charges were made in GBP. Prices in other currencies are approximate.
Line item amounts may not add up to the total charged due to rounding approximations.

Sequence Number: **NLNL-07D081353-1**

Thank you for your business!

Appendix I

The acceptable variants of the words in the single naming task

The word (English)	The word (Arabic)	Hijazi IPA	A	B	C	D	E
Sheep	خروف	xa.ru:f	xa:.ru:f				
Frog	ضفدع	dˤif.daʕ			dˤuf.dˤaʕ		
Horse	حصان	ħu.sˤa:n					ħi.sˤa:n
Parrot	بغبغان	baɣ.ba.ɣa:n		bab.ba.ɣaʔ			
Spider	عنكبو ت	ʕan.ka.bu:t		ʕan.ka.bu:.ta			
Circle	دائرة	ha.ram	do.wai:.ra				
Leg	رجل	ri.ʒil	ri. ʒu:l				
Head/	راس	ra:s /waʒh					raʔs
Face	وجه		waʃ				
Washing	يغسل	ji.ɣas.sil	ɣa.si:l				
cloth	مالب س	ma.la:.bis					
Diving	يغو ص	ji.ɣu:sˤ	ɣau.wa:sˤ				
Watermelon	بطي خ	batˤ.tˤi:x		batˤ.tˤi:.xa			
Banana	مو ز	mo:z	mo:za	mawz			
Apple	تفاح	tuf.fa:.ħa	tuf.fa:.ħ				
Limon	ليمون	laj.mu:n		laj.mu:.na			
Clementine	يوسفافندي	ju.su.fa.fan.di	ju.suf.ʔa.fan.di	ju.sif.ʔa.fan.di	ju.sa.fan.di	ju.su.fi	
Orange	برتقالة	bur.tu.qa:.la	bur.tu.qa:l				
Carrot	جز ر	ʒuzar	ʒazar	ʒazara	ʒuzara		
Sweet	حالو ة	ħa.la:.wa	ħa.la:.wi.ja:t				
Chicken	دجاج	du.ʒa:ʒ	du.ʒa:.ʒa	di.ʒa:ʒ	da.ʒa:ʒ		

Fish	سمك	sa.mak	sa.ma.ka				
Egg/ Nest	بيض/ع ش	ʕuʃ be:dˤ					be:ðˤ
Spoon	ملعقة	mil.ʕa.ga					mal.ʕa.ga
Tea	شاهي	ʃa:.hi:					ʃa:ji:
Milk	حليب	ħa.li:b					
Gold	دهب	da.hab					ða.hab
Key	مفتاح	muf.ta:ħ					mif.ta:ħ
	محفظة	miħ.fa.dˤa		miħ.fa.zˤa			miħ.fa.ðˤa
Girl dress	فستان	fus.ta:n					fis.ta:n
Men scarf	شماغ	ʃu.ma:ɣ	ʃu.ma:x				ʃi.ma:ɣ
Towel	منشفة	min.ʃa.fa	min.ʃi.fa				man.ʃa.fa
Soap	صابون	sˤa:.bu:n			sˤa.wa:.bi:n		
Crayon	تالوي ن	ʔal.wa:n	ta.la:.wi:n				
Fan	مروح ة	mir.wa.ħa				ma.ru. ħa	mar.wa.ħa
Plaster	لصق	lasˤag			lasˤaga	lasˤga	
Staircase	درج	da.raʒ		da.ra.ʒa			
Garbage	زبالة	zi.ba:.la					zu.ba:.la
pillow	مخدة	mu.xa.da			mu.xad.da		mxa.da
Pyramid	هرم	ha.ram	ʔah.ra:m				
mistake	خطأ	xa.tˤaʔ					ɣa.latˤ
Back	ظهر	dˤa.hir	dˤah.ru				
Corn	ذرة	dura					ðura
Saudi dress	ثو ب	to:b			θo:b		
Triangle	مثلث	mu.θal.laθ		mu.sal.las			
Ear		أذنʔidin			ʔu.ðun		ʔuðn

Appendix J

UHA and MSA consonants phonemes in SHAPA: word–positions and frequency

Phoneme	SIWI	SIWW	SFWW	SFWF	Frequency
/n/	nax.la	batˤ.tˤa:.ni.ja	min.ʃa.fa	ʔal.wa:n	
	naʒ.ma	bal.lo:.na	man.di:l	sˤa.ħin	
	ni.mir	ma.ka.ru:.na	ʕan.ka.bu:t	ħu.sˤa:n	
	na.dˤ.dˤa:.ra	ɣur.fa.ta.no:m	sˤan.du:g	fus.ta:n	
		ʔit.ne:n		baɣ.ba.ɣa:n	
		lu.ba:.na		ʔi.din	
				sˤa:.bu:n	
				laj.mu:n	
				ja.mi:n	
				ʔit.ne:n	
				tˤi:n	
	4	6	4	11	25
/m/	mu:z	sa.mak	ʃam.ʕa	tˤa.ma:.tˤim	
	milħ	laj.mu:n	lam.ba	ha.ram	
	man.di:l	ta.mur	tim.sa:ħ	xu.ʃum	
	ma.la:.bis	ʔaħ.mar	jim.dˤuɣ	fam	
	midˤ.rab	ʃu.ma:ɣ		ga.lam	
	muf.ta:ħ	tˤa.ma:.tˤim		ɣur.fa.ta.no:m	
	mus.taʃ.fa:	ʒaz.ma		tag.wi:m	
	mu.xa.da	naʒ.ma		ħam.ma:m	
	ma.ka.ru:.na	ja.mi:n		ha.ram	
	mak.wa	ni.mir			
	mad.ra.sa	mig.la.mi.ja			
	mu.kaj.jif				
	mil.ʕa.ga				
	mir.wa.ħa				
	mas.ʒid				
	matˤ.bax				
	ma.ri:dˤ				
	mu.ʃutˤ				
	mu.sa.las				
	ma.gasˤ				
	min.ʃa.fa				
	miħ.fa.zˤa				
	mas.baħ				
	mig.la.mi.ja				
	24	11	4	9	48
/b/	ba.ga.ra	sˤa:.bu:n	ʔab.jadˤ	do:.la:b	
	bur.tu.qa.la	zi.ba:.la	ʔib.ti.da:.ʔi:	midˤ.rab	

	bint	ma.la:.bis		ħa.li:b	
	batˤ.tˤi:x	baɣ.ba.ɣa:n		ki.ta:b	
	batˤ.tˤa:.ni.ja	matˤ.bax		da.hab	
	baɣ.ba.ɣa:n	ʕan.ka.bu:t		to:b	
	ba.tˤa:.tˤis	liʕ.ba			
	bal.lo:.na	lam.ba			
	be:t	tˤar.bu:ʃ			
	bir.wa:z	mas.baħ			
	be:dˤ	ʔusˤ.ba:ʕ			
		lu.ba:.na			
	11	12	2	6	31
/t/	tuf.fa:ħa	muf.ta:ħ		ʕan.ka.bu:t	
	ta.mur	mus.taʃ.fa:		ze:t	
	tim.sa:ħ	ki.ta:b		be:t	
	tag.wi:m	ʔib.ti.da:.ʔi:			
		si.ta:.ra			
		ɣur.fa.ta.no:m			
		bur.tu.qa.la			
	4	7	0	3	14
/d/	di:k	mu.xa.da	mad.ra.sa	wa.lad	
	do:.la:b	man.di:l		jad	
	da:.ʔi.ra	sˤan.du:g		mas.ʒid	
	durʒ	ʔib.ti.da:.ʔi:			
	da.raʒ	suʒ.ʒa:.da			
	du.ʒa:ʒ	ha.di.ja			
		hu.du:ʔ			
		war.da			
		ju.su.fa.fan.di			
	6	9	1	3	19
/k/	kur.si:	ʕan.ka.bu:t	mak.wa	sa.mak	
	kalb	ma.ka.ru:.na		ʃub.ba:k	
	ki.ta:b	mu.kaj.jif		di:k	
	3	3	1	3	10
/f/	fi:l	za.ra:.fa	muf.ta:ħ	xa.ru:f	
	fus.ta:n	ʕasˤ.fu:r	dˤif.d̠aʕ	mu.kaj.jif	
	fa.ra:.ʃa	min.ʃa.fa		ʃar.ʃaf	
	fam	miħ.fa.zˤa			
		ɣur.fa.ta.no:m			
		ju.su.fa.fan.di			
		dˤa.fi:.ra			
		ga.fasˤ			
	4	8	2	3	17

/l/	laj.mu:n	ma.la:.bis	mil.ʕa.ga	fi:l	
	liʕ.ba	do:.la:b	ʔal.wa:n	ʒau:.wa:l	
	lam.ba	wa.lad		ri.ʒil	
	lu.ba:.na	ga.lam		ji.ɣas.sil	
	la.zi:z	mig.la.mi.ja		hi.la:l	
	la.sˤag	ħa.la:.wa		man.di:l	
		tˤa:w.la			
		ɣas.sa:.la			
		zi.ba:.la			
		hi.la:l			
		ħa.li:b			
		nax.la			
		bur.tu.qa:.la			
		ta.la:.ta			
	6	14	2	6	28
/w/	wa.lad	ħa.la:.wa			
	war.da	ʔal.wa:n			
	waʒh	mak.wa			
		tag.wi:m			
		bir.wa:z			
	3	5	0	0	89
/ħ/	ħu.sˤa:n	sˤa.ħin	ʔaħ.mar	muf.ta:ħ	
	ħa.li:b	mir.wa.ħa	miħ.fa.zˤa	ji.tˤi:ħ	
	ħa.la:.wa	tuf.fa:ħa		mas.baħ	
	ħam.ma:m			tim.sa:ħ	
	4	3	2	4	13
/ʃ/	ʃub.ba:k	min.ʃa.fa	mus.taʃ.fa:	tˤar.bu:ʃ	
	ʃa:.riʕ	mu.ʃutˤ		ʕe:ʃ	
	ʃams	fa.ra:.ʃa		ʕuʃ	
	ʃu.ma:ɣ	ʃar.ʃaf			
	ʃa:.hi:	xu.ʃum			
	ʃam.ʕa				
	ʃar.ʃaf				
	7	5	1	3	16
/s/	sa:.ʕa	kur.si:	mas.baħ	ra:s	
	suʒ.ʒa:.da	ju.su.fa.fan.di	fus.ta:n	ba.tˤa:.tˤis	
	sa.mak	tim.sa:ħ	mas.ʒid	ma.la:.bis	
	si.ta:.ra	mad.ra.sa	mus.taʃ.fa:	xas	
	4	4	4	4	16
/x/	xe:tˤ	mu.xa.da	nax.la	batˤ.tˤi:x	
	xa.ru:f		ʔax.dˤar	matˤ.bax	

	xi.ja:r			sˤa:.ru:x	
	xu.ʃum				
	xas				
	xal.la:tˤ				
	xa.tˤaʔ				
	7	1	2	3	13
/h/	ha.di.ja	ʃa:.hi:	gah.wa		
	hi.la:l	da.hab			
	hu.du:ʔ	dˤa.hir			
	ha.ram				
	4	3	1	0	8
/r/	ruz	xa.ru:f	war.da		
	ra:s	ʃa:.riʕ	mir.wa.ħa	ʒu.zar	
	ri.ʒil	ʔaz.rag	kur.si:	ta.mur	
	ro:ʒ	sˤa:.ru:x	tˤar.bu:ʃ	ʔaħ.mar	
	raw.dˤa	fa.ra:.ʃa	bir.wa:z	xi.ja:r	
		za.ra:.fa	ɣur.fa	ʔax.dˤar	
		mad.ra.sa	bur.tu.qa.la	ʕasˤ.fu:r	
		ma.ka.ru:.na		ni.mir	
		ba.ga.ra		ʒu.zar	
		ma.ri:dˤ		ʔasˤ.far	
		sˤu:.ra		dˤa.hir	
		du.ra:		ʕa.sˤi:r	
		na.dˤ.dˤa:.ra		sˤa.ɣi:r	
		dˤa.fi:.ra			
		si.ta:.ra			
		ɣa.ra:			
		da:.ʔi.ra			
		ha.ram			
	5	18	7	12	42
/ɣ/	ɣur.fa/ ɣur.fa.ta.no:m	ji.ɣas.sil	baɣ.ba.ɣa:n	ʃu.ma:ɣ	
	ɣa.ra:	ji.ɣu:sˤ		jim.dˤuɣ	
	ɣas.sa:.la	baɣ.ba.ɣa:n			
		sˤa.ɣi:r			
	3	4	1	2	10
/j/	ja.mi:n	ʔab.jadˤ			
	ju.su.fa.fan.di	batˤ.tˤa:.ni.ja			
	jad	mig.la.mi.ja	–		–
		ha.di.ja			
	3	4	0	0	7
/tˤ/	tˤar.bu:ʃ	xa.tˤaʔ	matˤ.bax	mu.ʃutˤ	

	tˤau:.la	tˤa.ma:.tˤim		xe:tˤ	
	tˤi:n	ba.tˤa:.tˤis		xal.la:tˤ	
	tˤa.ma:.tˤim				
	4	3	1	3	11
/dˤ/	dˤa.fi:.ra	ʔax.dˤar	midˤ.rab	ʔab.jadˤ	
	dˤirs	raw.dˤa		be:dˤ	
	dˤif.dˤaʕ	dˤif.dˤaʕ		ma.ri:dˤ	
		jim.dˤuɣ			
	3	4	1	3	11
/sˤ/	sˤu:.ra	ħu.sˤa:n	ʕasˤ.fu:r	ma.gasˤ	
	sˤa:.bu:n	la.sˤag	ʔusˤ.ba:ʕ	ga.fasˤ	
	sˤan.du:g	ʕa.sˤi:r	ʔasˤ.far	ji.ɣu:sˤ	
	sˤa:.ru:x				
	sˤa.ħin				
	sˤu:.ra				
	sˤa.ɣi:r				
	7	3	3	3	16
/z/	zi.ba:.la/zu.ba:la	ʒu.zar	ʔaz.rag	bir.wa:z	
	ze:t		ʒaz.ma	ruz	
	za.ra:.fa			mo:z	
	3	1	2	3	9
/ʔ/	ʔaz.rag	da:.ʔi.ra		laʔ	
	ʔi.din	ʔib.ti.da:.ʔi:		xa.tˤaʔ	
	ʔusˤ.ba:ʕ			hu.du:ʔ	
	ʔib.ti.da:.ʔi:				
	ʔab.jadˤ				
	ʔasˤ.far				
	ʔax.dˤar				
	ʔaħ.mar				
	8	2	0	3	13
/g/	gah.wa	mu.gasˤ	mag.la.mi.ja	ʔaz.rag	
	ga.fasˤ	ba.ga.ra		la.sˤag	
	gird	tag.wi:m		sˤan.du:g	
	ga.lam				
	4	3	1	3	11
/ʕ/	ʕan.ka.bu:t	sa:.ʕa	liʕ.ba	dˤif.dˤaʕ	
	ʕuʃ	mil.ʕa.ga		ʃa:.riʕ	
	ʕasˤ.fu:r	ʃam.ʕa		ʔusˤ.ba:ʕ	
	ʕe:ʃ				
	ʕa.sˤi:r				

	5	3	1	3	12
/ʒ/	ʒau:.wa:l	ri.ʒil	naʒ.ma	da.raʒ	
	ʒu.zar	mas.ʒid		ro:ʒ	
	ʒaz.ma	du.ʒa:ʒ		du.ʒa:ʒ	

	3	3	1	3	10
/θ/	to:b	ta.la:.ta	ʔit.ne:n	mu.sal.las	
	ta.la:.ta	mu.sal.las			
	2	2	1	1	6
/ð/	da.hab	ʔi.din		la.zi:z	
	du.ra	la.zi:z			
	2	2	0	1	5
/ðˤ/	dˤa.hir	miħ.fa.zˤa			
	zˤarf				
	2	1	0	0	3
/q/		bur.tu.qa.la			1

Appendix K

Words with heterosyllabic CC in SHAPA

Heterosyllabic CC	Target word	Glossary	Frequency
1. b.j	ʔab.jadˤ	white	1
2. b.t	ʔib.ti.da:.ʔi:	Primary school	1
3. d.r	mad.ra.sa	school	1
4. dˤ.r	midˤ.rab	Paddle	1
5. f.dˤ	dˤif.dˤaʕ	Frog	1
6. f.t	muf.ta:ħ	Key	1
7. h.w	gah.wa	Coffee	1
8. k.w	mak.wa	Iron	1
9. l.ʕ	mil.ʕa.ga	Spoon	1
10. l. w	ʔal.wa:n	Colours	1
11. m.b	lam.ba	Lamp	1
12. m.dˤ	jim.dˤuɣ	Chews	1
13. m.ʕ	ʃam.ʕa	Candle	1
14. m's	tim.sa:ħ	crocodile	1
15. n.d	man.di:l/ sˤan.du:g/	Tissue/ box	2
16. n.k	ʕan.ka.bu:t	Spider	1
17. n.ʃ	min.ʃa.fa	Towel	1
18. r.d	war.da	flower	1
19. r.f	ɣur.fa	room	1
20. r.s	kur.si:	chair	1
21. r.t	bur.tu.qa.la	orange	1
22. r.w	bir.wa:z/ mir.wa.ħa	Frame/ fan	2
23. r.ʃ	ʃar.ʃaf	Bed sheet	1
24. r.b	tˤar.bu:ʃ	Party hat	1
25. s.b	mas.baħ	Swimming pool	1
26. s.ʒ	mas.ʒid	mosque	1
27. s.t	mus.taʃ.fa:/ fus.ta:n	Hospital/ dress	2
28. sˤ.f	ʕasˤ.fu:r/ ʔasˤ.far	Bride/ yellow	2
29. sˤ.b	ʔusˤ.ba:ʕ	finger	1
30. t.n	ʔit.ne:n	two	1
31. tˤ.b	matˤ.bax	kitchen	1
32. x.dˤ	ʔax.dˤar	green	1
33. xl	nax.la	Palm tree	1
34. z.m	ʒaz.ma	shoos	1
35. z.r	ʔaz.rag	blue	1
36. ħ.f	miħ.fa.zˤa	wallet	1
37. ħ.m	ʔaħ.mar	red	1

38. g.l	mag.la.mi.ja	Pencil case	1
39. g.w	tag.wi:m	calendar	1
40. ɣb	baɣ.ba.ɣa:n	parrot	1
41. ʃ.f	mus.taʃ.fa:	Hospital	1
42. ʒ.m	naʒ.ma	star	1
43. ʕ.b	liʕ.ba	toy	1

Appendix L

Vowels phonemes in the Saudi Hijazi Phonology Assessment items

Vowel	Items					Frequency
/a/	'kalb	batˤ'tˤi:x	ta'la:ta	'daraʒ	'sˤu:ra	
	xa'ru:f	'tamur	ti'ɣassil	si'ta:ra	'madrasa	
	ʕasˤ'fu:r	jusifa'fandi	la'zi:z	'matˤbax	'hadija	
	'bagara	'ʒuzar	ma'ri:dˤ	ħam'ma:m	'luʕba	
	fa'ra:ʃa	'xas	mus'taʃfa:	'makwa	'masʒid	
	za'ra:fa	'dura	'dahab	ɣas'sa:la	suʒ'ʒa:da	
	baɣba'ɣa:n	ħaˈla:wa	'tˤa:ula	'mirwaħa	'ʃarʃaf	
	ʕanka'bu:t	lu'ba:na	nadˤ'dˤɑ:ra	mu'kajjif	'masbaħ	
	'ʔazrag	ħa'li:b	dˤɑ'fi:ra	'lam:ba	'midˤrab	
	'ʔaħmar	'samak	'miħfadˤɑ	'ʃam:ʕa	'haram	138
	'ʔaxdˤɑr	ba'tˤa:tˤis	ma'la:bis	xal'la:tˤ	'xatˤɑʔ	
	'ʔabjadˤ	maka'ro:na	'ʒazma	bal'lo:na	'laʔ	
	'ʔasˤfar	'sˤaħin	tag'wi:m	sˤa'ɣi:r	'gafasˤ	
	'naxla	'milʕaga	'muxada	'walad	xal'la:tˤ	
	'warda	'gahwa	batˤ'tˤɑ:nija	'galam	ʕaˈsˤi:r	
	'ʃams	'fam	'minʃafa	'maglamija		
	'naʒma	'jad	man'di:l	ʔal'wa:n		
	da:'ʔira	ja'mi:n	zi'ba:la	'magasˤ		
	mu'θallaθ	'waʒh	'ɣurfa	'lasˤɑg		
/a:/	fa'ra:ʃa	du'ʒa:ʒ	bir'wa:z			
	ħu'sˤa:n	ʔusˤ'ba:ʕ	do:ˈla:b			
	za'ra:fa	'ra:s	zi'ba:la			
	tim'sa:ħ	ta'la:ta	ʃubˈba:k			
	ˈbaɣbaɣa:n	mus'taʃfa:	siˈta:ra			
	hi'la:l	ʒaw'wa:l	ħamˈma:m			
	da:'ʔira		xal'la:tˤ			38
	tuf'fa:ħ	'sa:ʕɑ	ɣas'sa:la			
	burtu'qa:l	muf'ta:ħ	ʔal'wa:n			
	xi'ja:r	fus'ta:n	ɣɑˈra:			
	tˤɑ'ma:tˤim	ʃu'ma:ɣ	ʔibti'da:ʔi:			
	ħa'la:wa	ma'la:bis	suʒ'ʒa:da			
	lu'ba:na	ki'ta:b	'ʃa:riʕ			
/ɑ/	dˤuf'dˤɑʕ	'miħfadˤɑ	sˤɑn'du:g			
	'ʔaxdˤɑr	'lasˤɑg	tˤɑr'bu:ʃ			
	tˤɑ'ma:tˤim	dˤɑ'fi:ra	'dˤɑrf			15
	'dˤɑhir	sˤɑ'ɣi:r	'rau:dˤɑ			
	'sˤɑħin	'lasˤɑg	'xatˤɑʔ			
/ɑ:/	ba'tˤɑ:tˤis	sˤɑ:ˈbu:n	sˤɑ:'ru:x			5
	nadˤ'dˤɑ:ra	batˤ'tˤɑ:nija				
/u/	dˤuf'dˤɑʕ	lu'ba:na	'muʃutˤ	ʃub'ba:k		
	ħu'sˤa:n	'ʕuʃ	muf'ta:ħ	mu'kajjif		29

	mu'θallaθ	'ruz	fus'ta:n	hu'du:ʔ	
	tuf'fa:ħ	du'ʒa:ʒ	ʃu'ma:ɣ	suʒ'ʒa:da	
	'tamur	'xuʃum	'durʒ		
	jusifa'fandi	ʔusˤ'ba:ʕ	'kursi:		
	burtu'qɑ:l	'jimdˤuɣ	'muxada		
	'dura	mus'taʃfa:	ɣurfat'no:m		
/u:/	xɑ'ru:f	ri'ʒu:l	tˤɑr'bu:ʃ		12
	ʕasˤ'fu:r	ji'ɣu:sˤ	hu'du:ʔ		
	ʕanka'bu:t	'sˤu:ra	sˤ'ɑ:'ru:x		
	laj'mu:n	sˤa:'bu:n	sˤ'ɑn'du:g		
/i/	tim'sa:ħ	'milʕaga	ma'la:bis	'maglamija	43
	'nimir	ri'ʒu:l	ki'ta:b	ʔibti'da:ʔi:	
	'gird	'ʔidin	bir'wa:z	hi'du:ʔ	
	hi'la:l	'dˤirs	batˤ'tˤa:nija	'hadija	
	da:'ʔira	ʔiθ'ne:n	'minʃafa	'liʕba	

	jusifa'fandi	ti'ɣassil	zi'ba:la	'masʒid	
	xi'ja:r	'jimdˤuɣ	si'ta:ra	'midˤrab	
	ba'tˤɑ:tˤis	ji'ɣu:sˤ	'mirwaħa	'ʃa:riʕ	
	'milħ	ji'tˤi:ħ	mu'kajjif		
	'sˤaħin	'miħfadˤɑ	'bint		
/i:/	'fi:l	ja'mi:n	'kursi:	batˤ'tˤi:x	16
	'di:k	la'zi:z	man'di:l		
	ħa'li:b	ma'ri:dˤ	sˤa'ɣi:r		
	'ʃa:hi:	ji'tˤi:ħ	ʔibti'da:ʔi:		
	ʕa'sˤi:r	dˤɑ'fi:ra	'tˤi:n		
/e:/	'be:dˤ	'ze:t	ʔiθ'ne:n		6
	'ʕe:ʃ	'be:t	'xe:tˤ		
/o:/	'mo:z	'ro:ʒ	do:'la:b	bal'lo:na	7
	maka'ro:na	'to:b	ɣurfat'no:m		
/aj/	lai'mu:n				2
	mu'kaijif				
/au:/	'tˤau:la				2
	'rau:dˤɑ				
314					

Appendix M

Participant's Responses Recode Sheet.

The examiner Initial: ________ **Child Code:** ____________________ **The School Name:** ______________ **Testing Date:** ___ / ___ / ____

The target word			Hejazi IPA	Mode of stimulation		IPA for the child pronunciation	Consonants position		Stress patterns	Phonological patterns
	English	Arabic								
1.	Dog	كلب	kalb	S AQ D	1st Im Wh. Im		I: /k/ SI:	SF: F:		
2.	Sheep	خروف	xa.ru:f	S AQ D	1st Im Wh. Im		I: /x/ SI: /r/	SF: F: /f/		
3.	Frog	ضفدع	dˁif.dˁaʕ	S AQ D	1st Im Wh. Im		I: /dˁ/ SI:	SF: /f/ F: /ʕ/		
4.	Bird	عصفور	ʕasˁ.fu:r	S AQ D	1st Im Wh. Im		I: /ʕ/ SI: /f/	SF /sˁ/ F		
5.	Butterfly	فراشة	fa.ra:.ʃa	S AQ D	1st Im Wh. Im		I: /f/ SI	SF F		
6.	Horse	حصان	ħu.sˁa:n	S AQ D	1st Im Wh. Im		I: /ħ/ SI: /sˁ/	SF F		
7.	Cow	بقرة	ba.ga.ra	S AQ D	1st Im Wh. Im		I: /b/ SI: /g/	SF F		
8.	Elephant	فيل	fi:l	S AQ D	1st Im Wh. Im		I: /f/ SI:	SF F: /l/		
9.	Giraffe	زرافة	za.ra:.fa	S AQ D	1st Im Wh. Im		I: /z/ SI: /f/	SF F		
10.	Crocodile	تمساح	tim.sa:ħ	S AQ D	1st Im Wh. Im		I: /t/ SI	SF /m/ F		
11.	Rooster	ديك	di:k	S AQ D	1st Im Wh. Im		I: /d/ SI	SF F: /k/		
12.	Tiger	نمر	ni.mir	S AQ D	1st Im Wh. Im		I: /n/ SI	SF F		
13.	Monkey	قرد	gird	S AQ D	1st Im Wh. Im		I: /g/ SI	SF F		
14.	Parrot	بغبغان	baɣ.ba.ɣa:n	S AQ D	1st Im Wh. Im		I: / SI: /ɣ/	SF /ɣ/ F		
15.	Spider	عنكبوت	ʕan.ka.bu:t	S AQ D	1st Im Wh. Im		I: /ʕ/ SI: /k/	SF F: /t/		
16.	Blue	أزرق	ʔaz.rag	S AQ D	1st Im Wh. Im		I: /ʔ/ SI	SF:/z/ F: /g/		

Spontaneous response, AQ: Alternative question, D: sematic description, 1st Im: first sound imitation, Wh Im: whole word imitation.
I: Initial word position, SI: Syllable initial within the word, SF: Syllable final within the word, F: Final word position.

Appendix N

The pronunciations variants among the children of the single–naming task across all the participants

Column1	Column2	Column3	Column4	Column5	Column6	Column7	Column8	Column9	Column10	Column11
ˈkalb	kalba	ʔalˈkalb	ʔalˈkalb							
xɑˈruːf	ʔalxɑˈruːf									
dˤuf.dˤɑʕ	ˈsulħifa	dˤif.dɑʕ	dˤif.dˤɑʕ	dˤuf.dɑʕ	ðˤuf.dɑʕ	dˤuf.dˤɑʕ	dˤuf.dˤɑʕa			
ʕɑsˤ.fuː.r	ʕɑsˤ.fuː.ra	ʕɑsˤɑː.fiːr								
ħu.sˤaːn	ħi.sˤaːn									
ˈbagɑra	ˈbagɑr	ba.qa.ra	ba.qar							
fiːl	ʔalˈfiːl									
diːk	ʔadˈdiːk									
ni.mir	na.mir	ʔan.ni.mir	ʔan.na.mir							
gird	girda	qird								
ˈbaɣbaɣaːn	ˈbabbaɣaːʔ	ˈbaɣbaɣaːʔ								
ʕan.ka.buːt	ʔalʕɑnkaˈbuːt	ʕan.ka.buːta								
ʔaz.rag	ˈʔazraq									
ʔaħ.mar	ʔalˈʔaħmar									
ʔax.dˤar	ʔalˈʔaxdˤɑr	ʔax.ðˤar								
ʔasˤ.far	ʔalˈʔasˤfar									
ʔab.jadˤ	ʔalˈʔabjadˤ									
naxla	naxiːl	ʔannaxla								
warda	ward									
ʃams	ʔaʃʃams									
naʒ.ma	ˈnaʒim									
hi.laːl	ʔalhi.laːl									

S: spontaneous response, AQ: Alternative question, D: sematic description, 1st Im: first sound imitation, Wh Im: whole word imitation.
I: Initial word position, SI: Syllable initial within the word, SF: Syllable final within the word, F: Final word position.

da:.ʔi.ra	do'we:ra	muˈdauwar								

mu.θal.laθ	mu.sal.las	mu.tal.lat	mu.θa.la:.θa:t							
laj.muːn	laj.muː.na									
batˤ.tˤiːx	batˤ.tˤiːxa									
ˈmoːz	ˈmoːza	mawz	mawza							
tuf.faː.ħa	tif.faːħ	tuf.faːħ								
ta.mur	ˈtamra	tamr								
ju.si.fa.fan.di	ˈjusiffandi	ˈjusifʔafandi	ˈjusufafandi	ˈʔafandi	ju.sif.ʔa.fan.di					
bur.tu.qaː.la	bur.tu.qaːl									
ʒu.zar	ˈʒazara	ʒa.zar	ʒu.zara							
xi.ja:r	xi.ja:.ra									
du.ra	ðu.ra									
ħa.la:.wa	ħaˈlaːwija:t	ħaˈlawja:t								
lu.ba:.na	li.ba:.n	li.ba:.na	lu.ba:.n							
ħa.liːb	ʔalħa.liːb									
beːdˤ	bajdˤ	bajdˤa	be:dˤa	beː.dˤa:t	ʔalbeːdˤa					
ʕuʃ	ʔalʕuʃb	ʕiʃ	ʕiʃb	ˈʕuʃŭb						
ʕeːʃ	ʔalʕeːʃ									
ruz	ʔar.ruz									
du.ʒa:ʒ	da.ʒa:.ʒ	da.ʒa:.ʒa	di.ʒa:ʒ	di.ʒa:ʒa	du.ʒa:.ʒa	ʔaddaˈʒa:ʒ	ʔaddu.ʒa:ʒ			
sa.mak	sa.ma.ka	ʔassa.ma.ka								
ba.tˤa:.tˤis	ba.tˤa:.tˤa									
ma.ka.ro:.na	maʕ.ka.ro:.na									
ze:t	ˈzajt	ʔazzajt	ʔazze:t							
milħ	ʔalmilħ									

sˤa.ħin	ˈsˤaħan	sˤa.ħn	ʔasˤa.ħin								
mil.ʕa.ga	ˈmilʕɑqa	mal.ʕa.ga	ʔalˈmilʕɑga								
ʃa:.hi:	ʔaʃʃa:.hi:										

gah.wa	ˈqahwa										
ʕa.sˤi:r	ʕa.sˤi:r burtuqa:l										
xu.ʃum	ˈxuʃmi	ˈxuʃmu	xa.ʃim	xaʃ.mi							
fam	ˈfammana	fammak	fammi	fammu	ʔalfam						
jad	ˈjaddak	jadde:n	jaddi	jaddu	ʔaljad	jaddana					
ja.mi:n	biljami:n	jumˈna	ʔalˈjumna	ʔaljami:n							
ri.ʒil	ri.ʒi:li	ri.ʒlu	ri.ʒu:l	ri.ʒu:li	riˈʒlak	riˈʒu:laha	riʒ.la.na	riʒl	riʒli	ʔarriˈʒu:l	
ʔi.ðin	ˈʔuðni	ʔalʔi.din	ʔi.dni	ʔi.dini	ʔi.ðini	ʔi.dni	ʔi.dnu	ʔu.dun	ʔu.ðun		
ʔusˤ.ba:ʕ	ʔaˈsˤa:ˈbiʕ	ʔal.ʔusˤ.ba:ʕ	ʔasˤa:.biʕ	ʔasˤa:ˈbi:ʕ	ʔisˤ.ba:ʕi						
waʒh	ˈwaʒhi	ˈwaʒhik	waʒhana	waʒhu	ʔal.waʒh						
ra:s	ˈra:sana	ra:si	ra:su	raʔs	ʔarˈra:s	ʔarraʔs					
dˤa.hir	ˈðˤahr	ˈdˤahri	ˈdˤahru	ˈðˤahru	dˤaˈhar	ðˤa.hir	ðˤaˈhar	ʔadˤdˤa.hir	ʔadˤdˤaˈhar		
ʔit.ne:n	ʔit.ne:na	ʔiθˈna:n	ʔiθ.ne:n								
ta.la:ta	θa.la:.θa										
ti.ɣa.sil	ji.ɣa.sil	ni.ɣa.sil	ti.ɣa.sil.ha	ʔi.ɣa.sil							
jim.dˤuɣ	ˈjimdˤuɣha	ˈnimdˤuɣ	ˈtimdˤuɣ	ʔam.dˤuɣ	ʔim.dˤuɣ						
la.zi:z	la.ði:ða	la.zi:.za	la.ði:ð								
ji.ɣu:sˤ	bi.ɣu:sˤ	ja.ɣu:sˤ	ʔi.ɣu:sˤ								
ma.ri:dˤ	marˈdˤa:n										
ji.tˤi:ħ	ħajtˤi:ħ	tˤa:.ji:ħ	tˤa:ħ								
mus.taʃ.fa:	mis.taʃ.fa:	ʔal.mus.taʃ.fa:									
be:t	bajt	be:.ta.ha	be:tana:	ʔal.be:t							

da.hab	ða.hab	ʔadda.hab	ʔaðða.hab							
ʒaw.wa:l	ʒaw.wa:la:t	ʒaw.wa:li	ʒawˈwa:lik	ʔaʒʒaw.wa:l						
nadˤ.dˤa:.ra	naðˤ.ðˤa:.ra	nadˤ.dˤa:.ra:t	ʔannadˤ.dˤa:.ra							
mu.ʃutˤ	ˈmiʃtˤ									
ro:ʒ	ʔarˈro:ʒ									

sa:.ʕa	ˈsa:ʕɑ:t	ʔasˈsa:ʕɑ								
muf.ta:ħ	mif.ta:ħ	ʔalmuf.ta:ħ								
miħ.fa.dˤa	miħ.fa.ðˤa	ˈmaħfadˤɑ	ˈmaħfaðˤɑ	ˈmiħfazˤɑ	ˈmaħfazˤɑ	ʔal.miħ.fa.ðˤa				
fus.ta:n	fis.ta:n	ʔalfis.ta:n								
to:b	θo:b	ˈθawb								
ʃu.ma:ɣ	ʃi.ma:ɣ	ʃi.ma:x	ʃu.ma:x	ʔaʃʃi.ma:ɣ	ʔaʃʃu.ma:ɣ					
ma.la:.bis	ʔalmala:bis									
ʒaz.ma	ˈʒazmatu	ˈʒizam	ʔaʒˈʒazma							
xe:tˤ	ˈxajtˤ	ˈxiju:tˤ	ˈxajja:tˤ							
ki.ta:b	ku.tub	ʔalkiˈta:b								
tag.wi:m	taq.wi:m									
sˤu:.ra	ˈsˤu:rati	ˈsˤuwar	ʔasˤˈsˤu:ra	ʔasˤsˤuwar						
bir.wa:z	ʔal.bir.wa:z									
ʔaddurʒ	ʔaddurʒ									
do:.la:b	ʔaddo:.la:b									
kur.si:	ʔalˈkursi:									
tˤa:w.la	ˈtˤa:wila									
mu.xa.da	ˈmixada	ma.xa.da	ʔalˈmaxada	ʔalmuxada						
min.ʃa.fa	man.ʃa.fa	ʔalˈminʃafa								
man.di:l	ma.na:.di:l									

sˁa:.bu:n	ˈsˁa:bu:na	ʔasˁsˁa:.bu:n								
zi.ba:.la	zu.ba:la	ʔazzi.ba:.la	ʔazzu.ba:.la							
ɣur.fat.no:m	ˈɣurfatano:m	ɣur.fa	ʔalˈɣurfa							
da.raʒ	ˈdaraʒa	ʔadˈdaraʒ								
si.ta:.ra	si.ta:r	ʔas.si.ta:ra								
matˁ.bax	titˁ.bux	matˁ.bax								
ħam.ma:m	ʔalħam.ma:m									

xal.la:tˁ	xal.la:.tˁa	ʔalxal.la:tˁ								
mak.wa	ˈkawwa:ja	mik.wa	ʔalˈmakwa	ʔalˈmikwa						
ɣas.sa:.la	ʔalɣasˈsa:la									
mir.wa.ħa	ˈmarwaħa									
mu.kaj.jif	mi.kaj.jif	ʔalmukajjif								
lam:.ba	lam:ba:t	ʔallam:ba								
ʃam:.ʕa	ʃu.mu:ʕ	ʃi.mu:ʕ	ʔaʃˈʃam:ʕɑ							
tˁar.bu:ʃ	tˁar.bu:ʃi									
bal.lo:.na	ballo:na:t									
sˁa.ɣi:r	sˁa.ɣi:ra									
bint	ʔalbint									
wa.l:ad	ʔalwalad	ʔaw.la:d								
ga.lam	ʔăgˈla:m	qalam	ʔalga.lam							
mag.la.mi.ja	ˈmaglama	ˈmaqlamija	ˈmiglamija							
ʔal. wa:n	ʔalʔal. wa:n	taˈla:wi:n								
ma.gasˁ	ma.qasˁ	miˈqasˁ	ʔalmugasˁ	mu.gasˁ	muˈqasˁ	maqsˁ				
ɣa.ra:	ʔalɣa.ra:									
la.sˁag	ˈlasˁaga	ˈlasˁaq	ˈlasˁaqa	ʔalˈlasˁag	ʔalˈlasˁaq					

ðˤarf	dˤarf	ˈzˤarf	ʔaðˤˈðˤarf							
mad.ra.sa	madrasati	ma.da:ris	ʔalˈmadrasa							
ʔib.ti.da:.ʔi:	ʔalˈʔibtidaːʔiː									
rauː.dˤa	ʔarˈrauːdˤɑ	ʔarˈrauːdˤɑ	rauː.ðˤa							
hu.duːʔ	hi.du:ʔ									
ha.di.ja	ha.da:.ja									
liʕ.ba	ˈʔalˈʕaːb	luʕ.ba	ʔalˈliʕba							
mas.ʒid	ʔalˈmasʒid									
suʒ.ʒaː.da	si.ʒ.ʒa:da	ʔassuʒ.ʒaː.da	suʒ.ʒaː.d							
ʃar.ʃaf	ʔaʃʃar.ʃaf									
mas.baħ	mis.baħ									
tˤiːn	ʔatˤˈtˤiːn									
midˤ.rab	ˈmadˤrab	maðˤ.rab	miðˤ.rab	ʔalˈmidˤrab						
sˤa:.ruːx	ʔassˤˤa:.ruːx									
ha.ram	ʔalharam									
sˤan.duːg	sˤan.duːq									
xa.tˤaʔ	ɣɑlatˤ	ʔalxa.tˤaʔ								
ga.fasˤ	qa.fasˤ	ʔalˈqɑfasˤ								
ʃaː.riʕ	ʔaʃˈʃaːriʕ	ˈʃawaːriʕ								

Appendix O

The participant's response sheet
Imitation Task

examiner Initial: ____________Child Code: __________The School Name: _________Testing Date: ___/___/_____ Child Age: __________

The Phones	Child responses
/ʔ/	
/b/	
/t/	
/θ/	
/ʒ/	
/ħ/	
/χ/	
/d/	
/ð/	
/ɾ/	
/ʀ/	
/z/	
/s/	
/ʃ/	
/sˤ/	
/dˤ/	
/tˤ/	
/ðˤ/	
/ʕ/	
/ɣ/	
/f/	
/q/	
/g/	
/k/	
/l/	
/m/	
/n/	
/h/	
/w/	
/j/	
/a:/	
/i:/	
/u:/	

Appendix P

Phonological error patterns that identified for this study

Error pattern	Type	Definition
Substitution	1. Fronting	Fronting the postalveolar fricatives (ʃ/ʒ) to the alveolar/dental fricatives (s/z/θ/ð)
		Fronting the alveolar fricatives (s/z) to the dental fricatives (θ/ð)
		Fronting the velar stop (k/g) to alveolar stop (t/d)
	2. Backing	Backing of the alveolar fricatives (s/z) to postalveolar fricatives (ʃ/ʒ)
		Backing of the velar fricatives (x/ɣ) to pharyngeal fricatives (ħ/ʕ)
	3. Lateralization	Substitute the alveolar trill /r/ to the alveolar lateral approximate /l/
	4. Stopping of the fricatives	Stop the labiodental fricatives /f/ to bilabial stop /b/
		Stop the alveolar and postalveolar fricatives (s, sˤ, z, ʃ, ʒ) to alveolar stops (t/d)
		Stop the velar fricatives (x/ɣ) to the velar stop (k/g).
	5. Glottal replacement	Substitute any consonants to the glottal stop /ʔ/
	6. Replacement to /h/	Substitute any consonants to the glottal fricative /h/
	7. De–emphasisation	Substitute any emphatic consonants to its unemphatic counterpart: (tˤ, dˤ, sˤ, ðˤ, zˤ) substitute to (t, d, s, ð, z) respectively.
	8. De–nasalization	Substitute the nasal sound (n/m) to unnasal sound (d/b) respectively.
	9. Gemination	Geminate the consonants on the syllables boundires. For exapmle: /war.da/→[wad.da]; substitute SFWW consonant to the SIWW consonant.
Syllable structures	10. Final consonants deletion (FCD)	Delete the word-final singleton consonants.
	11. Syllable final consonant deletion (SFC del)	Deletion of the syllable final consonant within the word.

	12. Weak syllable deletion (WSD)	Deletion of unstressed weak syllable.
	13. Word-final Consonants clusters (CCs) reduction	Deletion one of the consonants of the cluster.
	14. Epenthesis	Vowel insertion between the two consonants of the cluster.
	15. Metathesis	Change the order /position of the consonants or syllables.
Assimilation	16. Assimilation	One sound changed to another sound that is similar or exactly the same as another sound in the word (including total or partial assimilation).
	17. Voicing	Unvoiced sound changed to voiced sound
	18. Devoicing	Voiced sound changed to unvoiced sound
Vowels	19. Vowel substation	Substitute any vowels in the word with another vowel
	20. Vowel deletion	Delete the vowel in the syllable which result in CC

Appendix Q

Number and percentage of spontaneous responses across age groups

Test items	5;6-5;11	5;0-5;5	4;6-4;11	4;0-4;5	3;6-3;11	3;0-3;5	2;6-2;11	Average% of Sp. Responses
n	33	37	31	39	38	27	30	
ˈkalb	32	36	25	37	33	23	17	85.50
xɑˈruːf	21	26	12	24	18	10	9	49.79
ˈdˤufdˤɑʕ	29	35	24	31	19	6	7	62.13
ʕɑsˤˈfuːr	28	31	26	28	29	15	17	73.26
faˈraːʃa	32	35	27	35	32	11	15	77.62
ħuˈsˤaːn	32	36	29	33	31	19	15	82.05
ˈbagɑrɑ	31	35	23	30	15	10	7	62.78
ˈfiːl	29	32	20	30	23	11	13	65.77
zaˈraːfa	25	31	20	28	19	9	10	58.93
timˈsaːħ	32	31	18	29	18	6	9	58.97
ˈdiːk	7	8	7	10	2	0	2	14.71
ˈnimir	26	22	15	22	13	5	3	43.68
ˈgird	26	29	23	28	16	6	2	53.45
baɣbaˈɣaːʔ	19	25	15	9	4	2	2	31.60
ʕɑnkaˈbuːt	29	27	18	16	12	10	6	49.79
ˈʔazrag	33	35	25	29	20	10	8	66.56
ˈʔaħmar	33	37	29	33	26	11	9	73.90
ˈʔaxdˤɑr	32	36	27	31	16	9	7	65.66
ˈʔasˤfar	23	31	24	31	19	9	5	58.63
ˈʔabjadˤ	33	37	26	28	23	8	8	67.50
ˈnaxla	11	14	7	7	2	0	1	17.19
ˈwarda	31	34	29	30	31	22	21	84.19
ˈʃams	33	35	31	35	33	15	17	83.34
ˈnaʒma	32	34	26	31	14	6	4	60.66
hiˈlaːl	5	1	3	2	1	0	0	5.04
ˈdaːʔira	32	34	26	32	18	6	4	62.53
muˈθallaθ	26	24	15	24	11	3	4	43.85
lajˈmuːn	32	36	28	32	21	16	19	77.79
batˤˈtˤiːx	23	27	17	22	12	8	4	46.92
ˈmoːz	33	36	29	36	36	19	23	89.28
tufˈfaːħ	33	36	30	38	36	23	23	92.59
ˈtamur	33	35	31	32	29	16	11	78.41
ˈjusufi	23	22	17	8	4	2	2	32.73
burtuˈqɑːl	33	37	28	32	24	19	19	81.32
ˈʒazar	31	34	28	30	34	9	13	74.17
xiˈjaːr	20	28	22	28	30	10	11	61.67
ˈxas	22	29	12	16	5	4	1	36.58
tˤɑˈmaːtˤim	29	32	24	32	24	12	14	69.73
ˈðura	29	32	23	38	23	9	7	66.17
ħalaːwiˈjaːt	32	35	30	37	35	24	27	93.46
luˈbaːn	26	25	19	2	13	6	2	39.41
ħaˈliːb	33	36	30	10	34	21	23	80.52
ˈbeːdˤ	33	37	31	27	32	18	20	83.82
ˈʕuʃ	10	21	10	26	2	0	0	27.32
ˈʕeːʃ	15	17	13	37	9	6	2	40.11
ˈruz	30	37	23	36	28	17	15	77.72
duˈʒaːʒa	27	32	18	28	21	11	7	59.64
ˈsamaka	33	37	31	3	35	23	19	78.33

baˈtˤaːtˤis	33	35	28	28	36	24	24	88.62
makaˈroːna	25	28	29	34	23	14	15	70.65
ˈzeːt	7	8	5	38	0	1	1	23.35
ˈmilħ	28	34	23	36	20	9	3	62.74
ˈsˤaħin	31	34	31	26	34	21	17	82.35
ˈmilʕɑga	33	37	30	35	35	22	22	90.49
ˈʃaːhiː	33	37	30	21	29	20	21	81.57
ˈgahwa	25	27	19	31	21	13	6	58.99
ʕaˈsˤiːr	20	33	25	37	31	17	16	74.74
ˈxuʃum	32	32	25	1	25	8	4	53.63
ˈfam	20	30	25	38	27	13	14	69.38
ˈjad	31	37	30	33	37	20	23	89.06
jaˈmiːn	9	13	8	12	3	3	0	19.71
riˈʒuːl	33	36	31	16	37	25	20	84.99
ˈʔuðun	33	35	30	15	32	13	14	72.69
ʔusˤˈbaːʕ	27	21	12	1	8	2	1	30.23
ˈwaʒh	25	26	22	26	18	7	0	50.99
ˈraːs	17	20	10	28	11	5	7	40.06
ˈdˤirs	1	8	5	32	0	0	0	17.55
ˈðˤahar	32	33	28	26	22	6	4	62.37
ʔitˈneːn	31	34	25	0	21	14	9	68.28
θaˈlaːθa	33	36	29	6	17	19	9	74.37
tiˈɣassil	27	32	22	3	22	12	3	60.12
ˈjimdˤuɣ	0	0	1	7	0	0	1	0.94
laˈziːz	5	7	0	6	2	1	1	8.82
jiˈɣuːsˤ	5	4	3	17	3	0	0	7.32
maˈriːdˤ	11	16	10	36	2	1	2	20.71
ˈtˤaːħ	33	36	28	0	31	14	11	77.11
ˈmustaʃfaː	24	25	13	37	17	4	1	41.24
ˈbeːt	31	37	30	38	38	19	17	87.52
ˈðahab	5	12	1	31	6	0	0	9.51
ʒawˈwaːl	33	35	29	5	37	26	28	96.08
naðˤˈðˤaːra	33	37	30	26	36	25	22	93.92
ˈmuʃutˤ	33	31	28	37	28	18	11	76.17
dˤɑˈfiːra	6	9	2	37	4	0	0	10.33
ˈroːʒ	28	36	28	19	25	20	13	74.62
ˈsaːʕɑ	33	37	31	33	36	22	23	92.90
mufˈtaːħ	32	37	30	27	37	24	22	93.03
ˈmiħfaðˤɑ	29	21	15	11	13	5	0	42.43
fusˈtaːn	33	36	28	32	35	21	16	85.43
ˈθoːb	31	29	22	27	26	14	2	63.15
ʃuˈmaːɣ	16	17	9	14	3	1	0	23.69
maˈlaːbis	32	32	27	34	22	18	11	73.77
ˈʒazma	29	29	28	1	31	13	13	71.63
ˈxeːtˤ	24	21	17	24	7	2	0	35.15
kiˈtaːb	32	31	29	2	29	12	7	72.59
tagˈwiːm	0	2	0	26	0	0	0	1.14
ˈsˤuːra	27	29	20	33	15	9	1	52.14
birˈwaːz	1	0	0	38	1	0	0	1.54
ˈdurʒ	23	24	22	38	20	10	6	54.92
doːˈlaːb	32	35	26	17	29	16	12	76.88
ˈkursiː	32	37	31	6	38	23	21	93.16
ˈtˤauːla	31	35	29	21	32	22	16	85.87
ˈmuxada	31	33	22	35	24	12	5	60.65

batˤˈtˤaːnija	13	11	7	30	8	7	1	22.49
ˈminʃafa	29	27	21	35	16	10	4	53.93
manaːˈdiːl	32	37	31	2	35	24	19	90.51
sˤaːˈbuːn	32	37	27	36	35	24	17	85.89
ziˈbaːla	31	36	28	25	25	18	10	76.73
ˈɣurfa	10	13	7	24	4	0	1	15.29
ˈdaraʒ	32	35	29	36	28	17	11	79.04
ʃubˈbaːk	30	31	27	31	17	14	6	63.58
siˈtaːra	25	25	18	6	15	9	1	48.44
ˈmatˤbax	32	36	31	21	31	17	10	81.00
ħamˈmaːm	22	30	26	27	30	18	11	70.85
xɑlˈlaːtˤ	13	12	5	15	3	1	0	16.42
ˈmakwa	29	29	20	15	18	5	5	52.82
ɣasˈsaːla	30	31	22	26	23	11	5	62.20
ˈmarwaħa	28	30	23	21	19	9	2	53.02
muˈkajjif	29	26	17	0	14	7	3	46.32
ˈlamːba	29	30	22	38	23	14	2	61.17
ˈʃamːʕɑ	28	30	25	15	19	9	1	55.66
tˤarˈbuːʃ	4	3	0	34	0	2	0	3.95
balˈloːna	33	37	31	37	36	24	23	94.33
sˤaˈɣiːr	13	25	20	31	22	16	5	49.47
ˈbint	33	36	30	4	36	23	13	86.72
ˈwalad	33	36	30	22	37	21	21	90.95
ˈgalam	32	34	25	34	28	17	7	73.08
ˈmaglamija	10	11	4	10	2	2	0	13.69
ʔalˈwaːn	30	33	21	7	18	12	13	63.14
ˈmagasˤ	33	36	31	3	31	19	15	84.14
ˈɣɑraː	26	23	16	32	5	4	1	35.64
ˈlasˤag	16	18	12	0	7	4	0	26.72
ˈðˤarf	17	12	4	2	1	0	0	15.31
ˈmadrasa	33	32	23	2	27	11	6	67.79
ʔibtiˈdaːʔiː	2	2	0	37	0	0	0	1.64
ˈrauːdˤɑ	2	4	1	31	3	1	1	6.10
huˈduːʔ	1	8	0	16	0	0	1	4.73
ˈhadija	33	34	30	17	35	20	10	83.66
ˈliʕba	26	30	25	10	27	19	8	70.09
ˈmasʒid	31	26	21	28	12	8	3	49.54
suʒˈʒaːda	23	19	15	5	12	1	0	35.84
ˈʃarʃaf	10	15	5	4	4	3	1	19.65
ˈmasbaħ	29	34	24	18	25	17	10	70.52
ˈtˤiːn	11	11	6	2	2	0	1	14.83
ˈmidˤrab	12	14	2	18	3	0	0	14.11
ˈsˤaːruːx	30	27	17	9	9	3	2	43.76
ˈharam	10	4	2	6	0	0	0	7.53
sˤanˈduːg	25	24	11	4	10	4	1	38.47
ˈxɑtˤaʔ	6	5	12	25	4	0	0	14.86
ˈlaʔ	7	9	8	8	3	1	3	15.48
ˈgɑfasˤ	15	12	4	4	0	0	1	14.91
ˈʃaːriʕ	28	31	23	23	22	6	2	56.61

Appendix R

Table 1

Test items named spontaneously by >50% of children (N=97 items)

Test items	5;6–5;11	5;0–5;5	4;6–4;11	4;0–4;5	3;6–3;11	3;0–3;5	2;6–2;11	Total	% of Sp. Responses
1. ˈkalb	32	36	25	38	33	23	17	204	86.81
2. xɑˈruːf	21	26	12	25	18	10	9	121	51.49
3. ˈdˤufdˤɑʕ	29	35	24	32	19	6	7	152	64.68
4. ʕɑsˤˈfuːr	28	31	26	29	29	15	17	175	74.47
5. faˈraːʃa	32	35	27	36	32	11	15	188	80.00
6. ħuˈsˤaːn	32	36	29	34	31	19	15	196	83.40
7. ˈbagɑrɑ	31	35	23	31	15	10	7	152	64.68
8. ˈfiːl	29	32	20	31	23	11	13	159	67.66
9. zaˈraːfa	25	31	20	29	19	9	10	143	60.85
10. timˈsaːħ	32	31	18	30	18	6	9	144	61.28
11. ˈgird	26	29	23	29	16	6	2	131	55.74
12. ʕɑnkaˈbuːt	29	27	18	17	12	10	6	119	50.64
13. ˈʔazrag	33	35	25	29	20	10	8	160	68.09
14. ˈʔaħmar	33	37	29	34	26	11	9	179	76.17
15. ˈʔaxdˤɑr	32	36	27	31	16	9	7	158	67.23
16. ˈʔasˤfar	23	31	24	31	19	9	5	142	60.43
17. ˈʔabjadˤ	33	37	26	28	23	8	8	163	69.36
18. ˈwarda	31	34	29	31	31	22	21	199	84.68
19. ˈʃams	33	35	31	36	33	15	17	200	85.11
20. ˈnaʒma	32	34	26	32	14	6	4	148	62.98

21. ˈdaːʔira	32	34	26	32	18	6	4	152	64.68
22. lajˈmuːn	32	36	28	33	21	16	19	185	78.72
23. ˈmoːz	33	36	29	36	36	19	23	212	90.21
24. tufˈfaːħ	33	36	30	38	36	23	23	219	93.19
25. ˈtamur	33	35	31	33	29	16	11	188	80.00
26. burtuˈqɑːl	33	37	28	32	24	19	19	192	81.70
27. ˈʒazar	31	34	28	31	34	9	13	180	76.60
28. xiˈjaːr	20	28	22	29	30	10	11	150	63.83
29. tˤɑˈmaːtˤim	29	32	24	33	24	12	14	168	71.49
30. ˈðura	29	32	23	30	23	9	7	153	65.11
31. ħaˈlaːwa	32	35	30	38	35	24	27	221	94.04
32. ħaˈliːb	33	36	30	39	34	21	23	216	91.91
33. ˈbeːdˤ	33	37	31	38	32	18	20	209	88.94
34. ˈruz	30	37	23	28	28	17	15	178	75.74
35. duˈʒaːʒa	27	32	18	27	21	11	7	143	60.85
36. ˈsamaka	33	37	31	38	35	23	19	216	91.91
37. baˈtˤaːtˤis	33	35	28	37	36	24	24	217	92.34
38. makaˈroːna	25	28	29	28	23	14	15	162	68.94
39. ˈmilħ	28	34	23	29	20	9	3	146	62.13
40. ˈsˤaħin	31	34	31	35	34	21	17	203	86.38
41. ˈmilʕɑga	33	37	30	39	35	22	22	218	92.77
42. ˈʃaːhiː	33	37	30	37	29	20	21	207	88.09
43. ˈgahwa	25	27	19	26	21	13	6	137	58.30

44. ʕaˈsˤiːr	20	33	25	35	31	17	16	177	75.32
45. ˈxuʃum	32	32	25	22	25	8	4	148	62.98
46. ˈfam	20	30	25	31	27	13	14	160	68.09
47. ˈjad	31	37	30	38	37	20	23	216	91.91
48. riˈʒuːl	33	36	31	39	37	25	20	221	94.04
49. ˈʔuðun	33	35	30	34	32	13	14	191	81.28
50. ˈðˤahar	32	33	28	26	22	6	4	151	64.26
51. ʔitˈneːn	31	34	25	29	21	14	9	163	69.36
52. θaˈlaːθa	33	36	29	33	17	19	9	176	74.89
53. tiˈɣassil	27	32	22	27	22	12	3	145	61.70
54. ˈtˤaːħ	33	36	28	32	31	14	11	185	78.72
55. ˈbeːt	31	37	30	37	38	19	17	209	88.94
56. ʒawˈwaːl	33	35	29	38	37	26	28	226	96.17
57. naðˤˈðˤaːra	33	37	30	39	36	25	22	222	94.47
58. ˈmuʃutˤ	33	31	28	32	28	18	11	181	77.02
59. ˈroːʒ	28	36	28	26	25	20	13	176	74.89
60. ˈsaːʕɑ	33	37	31	38	36	22	23	220	93.62
61. mufˈtaːħ	32	37	30	38	37	24	22	220	93.62
62. fusˈtaːn	33	36	28	34	35	21	16	203	86.38
63. ˈθoːb	31	29	22	28	26	14	2	152	64.68
64. maˈlaːbis	32	32	27	33	22	18	11	175	74.47
65. ˈʒazma	29	29	28	28	31	13	13	171	72.77
66. kiˈtaːb	32	31	29	35	29	12	7	175	74.47

67. ˈsˤuːra	27	29	20	25	15	9	1	126	53.62
68. ˈdurʒ	23	24	22	27	20	10	6	132	56.17
69. doːˈlaːb	32	35	26	34	29	16	12	184	78.30
70. ˈkursiː	32	37	31	39	38	23	21	221	94.04
71. ˈtˤauːla	31	35	29	39	32	22	16	204	86.81
72. ˈmuxada	31	33	21	18	24	12	5	144	61.28
73. ˈminʃafa	29	27	21	22	16	10	4	129	54.89
74. manˈdiːl	32	37	31	36	35	24	19	214	91.06
75. sˤaːˈbuːn	32	37	27	31	35	24	17	203	86.38
76. ziˈbaːla	31	36	28	35	25	18	10	183	77.87
77. ˈdaraʒ	32	35	29	37	28	17	11	189	80.43
78. ʃubˈbaːk	30	31	27	26	17	14	6	151	64.26
79. ˈmatˤbax	32	36	31	37	31	17	10	194	82.55
80. ˈmakwa	29	29	20	22	18	5	5	128	54.47
81. ɣasˈsaːla	30	31	22	28	23	11	5	150	63.83
82. ˈmarwaħa	28	30	23	16	19	9	2	127	54.04
83. ˈlamːba	29	30	22	27	23	14	2	147	62.55
84. ˈʃamːʕɑ	28	30	25	22	19	9	1	134	57.02
85. balˈloːna	33	37	31	39	36	24	23	223	94.89
86. ˈbint	33	36	30	35	36	23	13	206	87.66
87. ˈwalad	33	36	30	38	37	21	21	216	91.91
88. ˈgalam	32	34	25	34	28	17	7	177	75.32

89. ʔalˈwaːn	30	33	21	23	18	12	13	150	63.83
90. ˈmagasˤ	33	36	31	3	31	19	15	168	71.49
91. ˈmadrasa	33	32	23	32	27	11	6	164	69.79
92. ˈhadija	33	34	30	38	35	20	10	200	85.11
93. ˈliʕba	26	30	25	32	27	19	8	167	71.06
94. ˈmasʒid	31	26	21	17	12	8	3	118	50.21
95. ˈmasbaħ	29	34	24	29	25	17	10	168	71.49
96. ˈʃaːriʕ	28	31	23	26	22	6	2	138	58.72

Table 2

Test items named spontaneously by fewer than 50% of children (54 items)

Test items	5;6–5;11	5;0–5;5	4;6–4;11	4;0–4;5	3;6–3;11	3;0–3;5	2;6–2;11	Total	% of Sp. Responses
1. ˈdiːk	7	8	7	10	2	0	2	36	15.32
2. ˈnimir	26	22	15	22	13	5	3	106	45.11
3. baɣbaˈɣaːʔ	19	25	15	9	4	2	2	76	32.34
4. ˈnaxla	11	14	7	7	2	0	1	42	17.87
5. hiˈlaːl	5	1	3	2	1	0	0	12	5.11
6. muˈθallaθ	26	24	15	24	11	3	4	107	45.53
7. batˤˈtˤiːx	23	27	17	22	12	8	4	113	48.09
8. ˈjusufi	23	22	17	8	4	2	2	78	33.19

9. ˈxas	22	29	12	16	5	4	1	89	37.87
10. luˈbaːn	26	25	19	18	13	6	2	109	46.38
11. ˈʕuʃ	10	21	10	2	2	0	0	45	19.15
12. ˈʕeːʃ	15	17	13	11	9	6	2	73	31.06
13. ˈzeːt	7	8	5	3	0	1	1	25	10.64
14. jaˈmiːn	9	13	8	1	3	3	0	37	15.74
15. ʔusˤˈbaːʕ	27	21	12	12	8	2	1	83	35.32
16. ˈwaʒh	25	26	22	16	18	7	0	114	48.51
17. ˈraːs	17	20	10	15	11	5	7	85	36.17
18. ˈdˤirs	1	8	5	1	0	0	0	15	6.38
19. ˈjimdˤuɣ	0	0	1	0	0	0	1	2	0.85
20. laˈziːz	5	7	0	6	2	1	1	22	9.36
21. jiˈɣuːsˤ	5	4	3	3	3	0	0	18	7.66
22. maˈriːdˤ	11	16	10	8	2	1	2	50	21.28
23. ˈmustaʃfaː	24	25	13	17	17	4	1	101	42.98
24. ˈðahab	5	12	1	0	6	0	0	24	10.21
25. dˤɑˈfiːra	6	9	2	5	4	0	0	26	11.06
26. ˈmiħfaðˤɑ	29	21	15	20	13	5	0	103	43.83
27. ʃuˈmaːɣ	16	17	9	12	3	1	0	58	24.68
28. ˈxeːtˤ	24	21	17	14	7	2	0	85	36.17

29. tagˈwiːm		2	0	1	0	0	0	3	1.28
30. birˈwaːz	1	0	0	2	1	0	0	4	1.70
31. batˤˈtˤaːnija	13	11	7	6	8	7	1	53	22.55
32. ˈɣurfa	10	13	7	2	4	0	1	37	15.74
33. siˈtaːra	25	25	18	24	15	9	1	117	49.79
34. xɑlˈlaːtˤ	13	12	5	6	3	1	0	40	17.02
35. muˈkajjif	29	26	17	15	14	7	3	111	47.23
36. tˤarˈbuːʃ	4	3	0	0	0	2	0	9	3.83
37. sˤaˈɣiːr	13	25	20	16	22	16	5	117	49.79
38. ˈmaglamija	10	11	4	4	2	2	0	33	14.04
39. ˈɣɑraː	26	23	16	32	5	4	1	107	45.53
40. ˈlasˤag	16	18	12	7	7	4	0	64	27.23
41. ˈðˤarf	17	12	4	3	1	0	0	37	15.74
42. ʔibtiˈdaːʔiː	2	2	0	0	0	0	0	4	1.70
43. ˈrauːdˤɑ	2	4	1	3	3	1	1	15	6.38
44. huˈduːʔ	1	8	0	2	0	0	1	12	5.11
45. suʒˈʒaːda	23	19	15	18	12	1	0	88	37.45
46. ˈʃarʃaf	10	15	5	10	4	3	1	48	20.43
47. ˈtˤiːn	11	11	6	5	2	0	1	36	15.32
48. ˈmidˤrab	12	14	2	4	3	0	0	35	14.89
49. ˈsˤaːruːx	30	27	17	18	9	3	2	106	45.11
50. ˈharam	10	4	2	2	0	0	0	18	7.66
51. sˤanˈduːg	25	24	11	19	10	4	1	94	40.00
52. ˈxɑtˤaʔ	6	5	12	9	4	0	0	36	15.32

53. ˈlaʔ	7	9	8	6	3	1	3	37	15.74
54. ˈgɑfasˤ	15	12	4	4	0	0	1	36	15.32

Appendix S

Items named by imitation by more than 50% of children

Test items	5;6-5;11	5;0-5;5	4;6-4;11	4;0-4;5	3;6-3;11	3;0-3;5	2;6-2;11	Mean
hiˈlaːl	39.39	51.35	64.52	74.36	92.11	92.59	86.67	71.57
ˈjimdˤuɣ	42.42	64.86	70.97	82.05	89.47	85.19	56.67	70.23
ˈðˤarf	18.18	48.65	58.06	79.49	94.74	92.59	66.67	65.48
ˈtˤiːn	18.18	24.32	41.94	71.79	86.84	81.48	63.33	55.41
ˈmidˤrab	12.12	37.84	45.16	64.10	65.79	70.37	63.33	51.25
ˈgɑfasˤ	21.21	37.84	45.16	61.54	86.84	85.19	56.67	56.35
ˈdˤirs	60.61	59.46	67.74	89.74	97.37	92.59	76.67	77.74
tagˈwiːm	90.91	81.08	90.32	94.87	92.11	96.30	70.00	87.94
birˈwaːz	72.73	89.19	80.65	76.92	78.95	92.59	63.33	79.19
tˤarˈbuːʃ	60.61	59.46	77.42	89.74	81.58	88.89	76.67	76.34
ˈharam	45.45	70.27	80.65	87.18	92.11	85.19	73.33	76.31

Percentages of items named by imitation with more than 50% of participants across the age groups (n=11 items). Although /ˈjimdˤuɣ/ ‘chewing’ was the least item named spontaneously, /ˈdˤirs/ ‘molar tooth’ was the most item named by whole word imitation.

Appendix T

Table 1– Test items scored lower than 80% in picture–naming agreement measure (N=14 items)

Test items		**5;6–5;11**	**5;0–5;5**	**4;6–4;11**	**4;0–4;5**	**3;6–3;11**	**3;0–3;5**	**2;6–2;11**	**Mean**
1.	**bayba'ɣaːʔ**	57.58	59.46	87.10	79.49	86.84	88.89	90.00	78.48
2.	**ˈmoːz**	84.85	94.59	96.77	69.23	68.42	70.37	50.00	76.32
3.	**ˈjusufi**	51.52	78.38	58.06	69.23	76.32	77.78	76.67	69.71
4.	**ˈʒazar**	72.73	100.00	90.32	79.49	63.16	77.78	70.00	79.07
5.	**ˈbeːdˤ**	72.73	75.68	64.52	58.97	42.11	77.78	46.67	62.63
6.	**ˈʕuʃ**	51.52	59.46	74.19	74.36	89.47	92.59	96.67	76.89
7.	**daˈʒaːʒa**	75.76	75.68	77.42	74.36	76.32	77.78	76.67	76.28
8.	**ˈsamaka**	69.70	62.16	58.06	64.10	36.84	48.15	33.33	53.19
9.	**jaˈmiːn**	78.79	54.05	100.00	89.74	71.05	77.78	76.67	78.30
10.	**ˈwaʒh**	84.85	78.38	77.42	84.62	76.32	96.30	60.00	79.70
11.	**laˈziːz**	69.70	70.27	100.00	71.79	81.58	74.07	60.00	75.35
12.	**ˈtˤaːħ**	45.45	40.54	51.61	58.97	78.95	85.19	70.00	61.53
13.	**ˈlasˤag**	84.85	89.19	90.32	71.79	52.63	55.56	56.67	71.57
14.	**ˈmasʒid**	90.91	64.86	83.87	61.54	73.68	66.67	73.33	73.55

Note: The highlighted items are items with high items difficulty score.

Appendix U

Table 1

Intra–rater reliability points to point agreement for IPA phonetic transcription

	2;6–2;11	3;0–3;5	3;6–3;11	4;0–4;5	4;6–4;11	5;0–5;5	5;6–5;11	Average
Mean	97.93	97.57	98.99	99.13	98.93	99.00	99.23	98.68
SD	0.90	0.67	0.77	0.23	0.06	0.87	0.25	0.54
Min	97.00	97.00	98.10	99.00	98.90	98.00	99.00	98.14
Max	98.8	98.3	99.5	99.4	99	99.6	99.5	99.16

Table 2

Inter-rater reliability point to point agreement for IPA phonetic transcription

	2;6–2;11	3;0–3;5	3;6–3;11	4;0–4;5	4;6–4;11	5;0–5;5	5;6–5;11	Average
Mean	97.18	97.75	98.33	98.63	98.47	98.47	98.67	98.21
SD	1.06	1.29	1.10	0.29	0.42	1.08	0.58	0.83
Min	95.60	96.40	96.70	98.30	98.00	97.70	98.00	97.24
Max	97.90	99.00	99.00	99.00	98.80	99.70	99.00	98.91

Table 3

Intra–rater reliability percentage agreement for phonological patterns identified (using ≥4 cut-off criterion)

	2;6–2;11	3;0–3;5	3;6–3;11	4;0–4;5	4;6–4;11	5;0–5;5	5;6–5;11	Average
Mean	97.43	91.02	94.87	97.43	96.15	97.43	98.72	96.15
SD	2.22	2.22	5.88	4.45	6.66	4.45	2.22	4.01
Min	96.15	88.46	88.46	92.3	88.46	92.3	96.15	91.75
Max	100	92.3	100	100	100	100	100	98.90

Table 4

Inter–rater reliability percentage agreement for phonological patterns identified (using ≥4 cut-off criterion)

	2;6–2;11	3;0–3;5	3;6–3;11	4;0–4;5	4;6–4;11	5;0–5;5	5;6–5;11	Average
Mean	94.42	89.33	93.47	91.50	90.68	86.33	83.57	89.90
SD	0.03	0.04	0.08	0.03	0.05	0.07	0.11	0.06
Min	91.23	83.70	84.51	88.28	86.02	82.00	73.00	84.11
Max	97.35	93.30	98.35	93.68	96.70	94.00	95.70	95.58

Appendix V

Table 1

Consonant phonetic inventory of UHA and MSA phones for the single-word naming task

Phone		2;6–2;11	3;0–3;5	3;6–3;11	4;0–4;5	4;6–4;11	5;0–5;5	5;6–5;11
N Shared		30	27	38	39	31	37	33
1. **b**		100.00	100.00	100.00	100.00	100.00	100.00	100.00
2. **t**		100.00	100.00	100.00	100.00	100.00	100.00	100.00
3. **tˤ**		56.67	81.48	94.74	100.00	100.00	100.00	96.67
4. **d**		100.00	100.00	100.00	100.00	100.00	100.00	100.00
5. **dˤ**		36.67	77.78	89.47	94.87	93.55	100.00	100.00
6. **k**		86.67	100.00	94.74	100.00	96.77	100.00	100.00
7. **ʔ**		100.00	100.00	100.00	100.00	100.00	100.00	100.00
8. **f**		100.00	100.00	100.00	100.00	100.00	100.00	100.00
9. **s**		93.33	96.30	97.37	97.44	93.55	100.00	100.00
10. **sˤ**		50.00	88.89	86.84	100.00	90.32	97.30	90.00
11. **z**		83.33	85.19	81.58	94.87	87.10	97.30	96.67
12. **ʃ**		80.00	70.37	78.95	87.18	87.10	94.59	90.00
13. **ʒ**		60.00	55.56	71.05	82.05	87.10	91.89	80.00
14. **x**		63.33	74.07	78.95	92.31	100.00	100.00	93.33
15. **ɣ**		43.33	66.67	81.58	94.87	96.77	97.30	96.67
16. **ħ**		100.00	100.00	100.00	100.00	100.00	100.00	100.00
17. **ʕ**		70.00	96.30	94.74	100.00	100.00	100.00	100.00
18. **h**		96.67	100.00	100.00	100.00	100.00	100.00	100.00
19. **m**		100.00	100.00	100.00	100.00	100.00	100.00	100.00
20. **n**		100.00	100.00	100.00	100.00	100.00	100.00	100.00
21. **l**		100.00	100.00	100.00	100.00	100.00	100.00	100.00
22. **ɾ**		100.00	100.00	100.00	100.00	100.00	100.00	100.00
23. **r**		86.67	85.19	89.47	97.44	100.00	97.22	100.00
24. **w**		100.00	100.00	100.00	100.00	100.00	100.00	100.00
25. **j**		100.00	100.00	100.00	100.00	96.77	100.00	100.00
UHA								
26. ɡ		90.00	96.30	92.11	100.00	100.00	100.00	100.00
MSA								
27. q	*	20.00	40.74	47.37	84.62	89.66	86.49	93.33
28. θ	*	60.71	74.07	70.27	84.62	77.42	75.68	83.33
29. ð	*	100.00	100.00	100.00	96.15	96.00	93.10	100.00
30. ðˤ	*	83.33	90.91	93.75	100.00	90.91	92.86	100.00

Key: Mastery level (90–100%) Acquisition level (75–89%) Customary (50–74%) Not acquired (< 50%)

* = the phone occurs at least once by each child in an age group, due to the limited frequency of these phones within the word– list.

Table 2

Vowels phonetic inventory of UHA and MSA phones for the single-word naming task

Phone		2;6–2;11	3;0–3;5	3;6–3;11	4;0–4;5	4;6–4;11	5;0–5;5	5;6–5;11
N Shared		30	27	38	39	31	37	33
Vowels								
i		100.00	100.00	100.00	100.00	100.00	100.00	100.00
iː		100.00	100.00	100.00	100.00	100.00	100.00	100.00
u		100.00	100.00	100.00	100.00	100.00	100.00	100.00
uː		96.67	100.00	100.00	100.00	100.00	100.00	100.00
eː		100.00	100.00	100.00	100.00	100.00	100.00	100.00
oː		93.33	100.00	100.00	100.00	100.00	100.00	100.00
a		100.00	100.00	100.00	100.00	100.00	100.00	100.00
aː		100.00	100.00	100.00	100.00	100.00	100.00	100.00
a͡iː	*	55.56	59.26	81.58	89.74	93.55	97.30	93.33
a͡uː	*	69.23	96.30	97.30	100.00	100.00	100.00	96.67
ɑ		100.00	100.00	100.00	100.00	100.00	100.00	100.00
ɑː	*	83.33	74.07	89.47	94.87	100.00	100.00	100.00

Key: Mastery level (90-100%). Acquisition level (75-89%) Customary (50-74%). Not acquired (< 50%)

Table 3

Phonetic inventory of the phones prone to phonetic distortions for the single word naming task

Phone	2;6-2;11	3;0-3;5	3;6-3;11	4;0-4;5	4;6-4;11	5;0-5;5	5;6-5;11
N	30	27	38	39	31	37	33
s	100.00	100.00	100.00	100.00	100.00	100.00	100.00
sˤ	53.33	96.30	89.47	100.00	96.77	100.00	100.00
z	93.33	88.89	89.47	94.87	96.77	97.30	100.00
ʃ	86.67	77.78	84.21	87.18	90.32	100.00	93.33
ʒ	76.67	62.96	76.32	87.18	87.10	97.22	83.33
ɾ	100.00	100.00	100.00	100.00	100.00	100.00	100.00
r	93.33	85.19	100.00	100.00	100.00	100.00	100.00
l	100.00	100.00	100.00	100.00	100.00	100.00	100.00

Key: Mastery level (90-100%). Acquisition level (75-89%) Customary (50-74%). Not acquired (< 50%)

Appendix W

Table 1

Consonants phonetic inventory of UHA and MSA phones for the stimulability task

Phone	2;6–2;11	3;0–3;5	3;6–3;11	4;0–4;5	4;6–4;11	5;0–5;5	5;6–5;11
N	22	22	34	33	31	37	33
b	95.45	95.45	100.00	96.97	100.00	100.00	100.00
t	90.91	100.00	97.06	96.97	100.00	100.00	100.00
tˤ	50.00	86.36	85.29	93.94	96.77	91.89	100.00
d	72.73	86.36	100.00	93.94	90.32	97.30	93.94
dˤ	31.82	63.64	76.47	81.82	87.10	91.89	93.94
k	80.95	90.91	94.12	96.97	90.32	94.59	100.00
ʔ	77.27	95.45	94.12	96.97	87.10	91.89	96.97
f	95.24	95.45	97.06	96.97	90.32	100.00	100.00
s	54.55	68.18	82.35	81.82	83.87	86.49	75.76
sˤ	40.91	72.73	85.29	75.76	77.42	78.38	78.79
z	59.09	63.64	76.47	75.76	80.65	83.78	78.79
ʃ	36.36	54.55	82.35	78.79	77.42	78.38	87.88
ʒ	27.27	36.36	64.71	72.73	74.19	83.78	81.82
x	63.64	81.82	79.41	87.88	93.55	100.00	100.00
ɣ	42.86	68.18	79.41	93.94	90.32	97.30	96.97
ħ	72.73	90.91	94.12	96.97	96.77	100.00	100.00
ʕ	63.64	81.82	88.24	87.88	100.00	97.30	96.97
h	100.00	95.45	100.00	93.94	96.77	100.00	100.00
m	100.00	100.00	100.00	96.97	100.00	100.00	100.00
n	95.00	95.45	91.18	96.97	90.32	94.59	93.94
l	95.00	95.45	97.06	96.97	93.55	100.00	100.00
ɾ	50.00	77.27	76.47	87.88	83.87	81.08	96.97
r	9.09	45.45	35.29	54.55	74.19	62.16	84.85
w	95.00	100.00	100.00	100.00	100.00	100.00	100.00
j	100.00	100.00	100.00	100.00	100.00	100.00	100.00
UHA							
g	66.67	86.36	88.24	96.97	90.32	100.00	96.97
MSA							
q	14.29	40.91	50.00	78.79	87.10	81.08	93.94
θ	4.55	45.45	44.12	57.58	70.97	81.08	93.94
ð	22.73	36.36	52.94	57.58	64.52	78.38	96.97
ðˤ	22.73	45.45	52.94	63.64	58.06	83.78	90.91

Key: Mastery level (90-100%). Acquisition level (75-89%) Customary (50-74%). Not acquired (< 50%)

Table 2

Vowels phonetic inventory of UHA and MSA phones for the stimulability task

Phone	2;6–2;11	3;0–3;5	3;6–3;11	4;0–4;5	4;6–4;11	5;0–5;5	5;6–5;11
N Shared	22	22	34	33	31	37	33
Vowels							
i:	90.00	100.00	100.00	100.00	100.00	100.00	100.00
u:	90.00	100.00	100.00	100.00	100.00	100.00	100.00
a:	100.00	100.00	100.00	100.00	100.00	100.00	100.00

Key: Mastery level (90-100%). Acquisition level (75-89%) Customary (50-74%). Not acquired (< 50%)

Table 3

Phonetic inventory of the phones prone to phonetic distortions for the stimulability task

Phone	2;6–2;11	3;0–3;5	3;6–3;11	4;0–4;5	4;6–4;11	5;0–5;5	5;6–5;11
N	22	22	34	33	31	37	33
s	95.45	90.91	91.18	93.94	90.32	91.89	96.97
sˤ	50.00	77.27	88.24	87.88	83.87	86.49	90.91
z	72.73	77.27	82.35	87.88	87.10	91.89	90.91
ʃ	63.64	68.18	82.35	81.82	90.32	91.89	93.94
ʒ	45.45	45.45	67.65	72.73	77.42	83.78	84.85
ɾ	59.09	77.27	76.47	87.88	90.32	89.19	96.97
r	9.09	45.45	35.29	57.58	80.65	67.57	90.91
l	100.00	100.00	97.06	96.97	96.77	100.00	100.00

Key: Mastery level (90-100%). Acquisition level (75-89%) Customary (50-74%). Not acquired (< 50%)

Appendix X

Table 1

Word-final consonant clusters (CCs) inventory in SHAPA across age groups

Age groups	2;6–2;11	3;0–3;5	3;6–3;11	4;0–4;5	4;6–4;11	5;0–5;5	5;6–5;11
N / F–CC	30	27	38	39	31	37	33
1. lb	66.67	77.78	84.21	100.00	100.00	97.30	100.00
2. lħ	30.00	62.96	57.89	76.92	90.32	83.78	84.85
3. nt	56.67	74.07	65.79	82.05	87.10	91.89	90.91
4. ms	46.67	55.56	78.95	87.18	80.65	83.78	90.91
5. rd	20.00	59.26	50.00	69.23	77.42	81.08	90.91
6. rf	30.00	44.44	65.79	79.49	77.42	81.08	87.88
7. rs	16.67	40.74	47.37	64.10	74.19	75.68	78.79
8. rʒ	16.67	29.63	34.21	33.33	61.29	72.97	60.61
9. ʒh	3.33	18.52	18.42	30.77	35.48	51.35	51.52

Key: Mastery level (90-100%). Acquisition level (75-89%) Customary (50-74%). Not acquired (< 50%)

Table 2

The heterosyllabic consonant cluster inventory for single-word naming task across age groups

Heterosyllabic CC	Frequency	2;6-2;12	3;0-3;5	3;6-3;11	4;0-4;5	4;6-4;11	5;0-5;5	5;6-5;11
b.j	1	70.00	77.78	94.74	94.87	93.55	100.00	100.00
b.t	1	10.00	29.63	39.47	46.15	70.97	78.38	90.91
d.r	1	13.33	33.33	36.84	66.67	67.74	78.38	81.82
dˤ.r	1	0.00	22.22	47.37	53.85	48.39	45.95	45.45
f.dˤ	1	16.67	22.22	52.63	48.72	51.61	70.27	78.79
f.t	1	30.00	55.56	73.68	79.49	96.77	97.30	96.97
h.w	1	63.33	62.96	84.21	92.31	96.77	100.00	93.94
k.w	1	43.33	55.56	81.58	92.31	93.55	94.59	93.94
l.ʕ	1	30.00	59.26	78.95	84.62	96.77	94.59	90.91
l. w	1	53.33	55.56	84.21	82.05	93.55	94.59	93.94
m.b	1	56.67	66.67	86.84	94.87	93.55	100.00	100.00
m.dˤ	1	10.00	29.63	52.63	64.10	61.29	62.16	69.70
m.ʕ	1	33.33	77.78	76.32	79.49	87.10	94.59	96.97
m΄s	1	46.67	66.67	76.32	84.62	80.65	91.89	100.00
n.d	2	63.33	70.37	89.47	100.00	100.00	97.30	100.00
n.k	1	36.67	33.33	47.37	87.18	80.65	89.19	93.94
n.ʃ	1	16.67	7.41	26.32	43.59	64.52	83.78	72.73
r.d	1	40.00	40.74	55.26	66.67	70.97	89.19	90.91
r.f	1	26.67	44.44	57.89	79.49	77.42	83.78	90.91
r.s	1	43.33	29.63	34.21	64.10	54.84	83.78	75.76
r.t	1	6.67	14.81	23.68	28.21	61.29	72.97	81.82

r.w	2	36.67	51.85	57.89	66.67	83.87	86.49	93.94
r.ʃ	1	13.33	25.93	31.58	35.90	70.97	81.08	75.76
r.b	1	30.00	44.44	65.79	74.36	83.87	78.38	81.82
s.b	1	33.33	62.96	81.58	82.05	80.65	89.19	87.88
s.ʒ	1	3.33	0.00	2.63	10.26	9.68	24.32	12.12
s.t	2	46.67	74.07	81.58	94.87	90.32	97.30	96.97
sˤ.f	2	36.67	44.44	65.79	87.18	80.65	97.30	96.97
sˤ.b	1	20.00	48.15	55.26	69.23	80.65	97.30	78.79
t.n	1	30.00	37.04	55.26	82.05	80.65	94.59	81.82
tˤ.b	1	10.00	37.04	36.84	58.97	64.52	86.49	75.76
x.dˤ	1	3.33	22.22	42.11	53.85	54.84	62.16	57.58
xl	1	50.00	55.56	60.53	76.92	83.87	83.78	84.85
z.m	1	26.67	51.85	55.26	64.10	64.52	81.08	72.73
z.r	1	20.00	22.22	31.58	41.03	54.84	78.38	66.67
ħ.f	1	26.67	37.04	55.26	69.23	93.55	89.19	90.91
ħ.m	1	63.33	70.37	89.47	89.74	96.77	97.30	96.97
g.l	1	23.33	37.04	73.68	74.36	87.10	83.78	90.91
g.w	1	33.33	44.44	42.11	69.23	74.19	72.97	72.73
ɣb	1	13.33	22.22	39.47	71.79	67.74	70.27	63.64
ʃ.f	1	36.67	18.52	44.74	58.97	70.97	86.49	72.73
ʒ.m	1	26.67	25.93	42.11	53.85	64.52	78.38	60.61
ʕ.b	1	36.67	55.56	71.05	74.36	93.55	89.19	84.85

Key: Mastery level (90-100%). Acquisition level (75-89%) Customary (50-74%). Not acquired (< 50%)

Frequency= the frequency of occurrences of the clusters in the single-word naming task.

Table 3

Geminated consonants inventory in SHAPA across age groups

Age groups	**2;6–2;11**	**3;0–3;5**	**3;6–3;11**	**4;0–4;5**	**4;6–4;11**	**5;0–5;5**	**5;6–5;11**
N Gemin.	30	27	38	39	31	37	33
1. bb	76.67	88.89	100.00	100.00	100.00	100.00	100.00
2. dˤdˤ	23.33	14.81	60.53	48.72	54.84	51.35	45.45
3. ff	83.33	81.48	94.74	100.00	100.00	100.00	100.00
4. jj	76.67	81.48	94.74	97.44	96.77	100.00	100.00
5. ll	83.33	100.00	100.00	100.00	100.00	100.00	100.00
6. mm	83.33	92.59	97.37	92.31	100.00	100.00	100.00
7. ss	70.00	81.48	89.47	94.87	90.32	100.00	100.00
8. tˤtˤ	30.00	59.26	65.79	94.87	83.87	94.59	96.97
9. ʒʒ	10.00	14.81	23.68	43.59	51.61	56.76	63.64

Key: Mastery level (90-100%). Acquisition level (75-89%) Customary (50-74%). Not acquired (< 50%)

Appendix Y

Developmental phonological patterns identified in SHA-speaking children between age of 2;6– 5;11

Age group Pattern		2;6–2;11	3;0–3;5	3;6–3;11	4;0–4;5	4;6–4;11	5;0–5;5	5;6–5;11
		n= 30	n=27	n=38	n=39	n=31	n=37	n=33
		% (n)	% (n)	% (n)	% (n)	% (n)	% (n)	% (n)
Systemic simplifications								
Fronting of /ʃ ʒ/→ [s z θ ð]	≥4	60 (18)	70 (19)	61(23)	41(16)	35 (11)	27 (10)	36 (12)
/ʃa:.riʕ/ → [sa:.riʕ] (street)	≥6	47(14)	63 (17)	58 (22)	38 (15)	23 (7)	16 (6)	27 (9)
Fronting of /s, z/ □[θ, ð]	≥4	27 (8)	26 (7)	21 (8)	18 (7)	29 (9)	16 (6)	15 (5)
/ze:t/□ [ðe:t] **(oil)**	≥6	27 (8)	15 (4)	16 (6)	15 (6)	23 (7)	14 (5)	
Fronting /k, g/□ [t/d]	≥4	17 (5)						
/kalb/□[talb**] (dog)**	≥6	17(5)						
Backing /x, ɣ/□ [ħ/ʕ]	≥4	30 (9)	26 (7)	26 (10)	13 (5)			
/xas/□[ħas] **(lettuce)**	≥6	10 (3)	26 (7)	21 (8)	10 (4)			
Stopping of Fricatives	≥4	57 (17)	48 (13)	61 (23)	28 (11)			
/sa.mak/□[ta.mak] **(fish)**	≥6	40 (12)	37 (10)	34 (13)	15 (6)			
Lateralization /r/□[l]	≥4	80 (24)	56 (15)	42 (16)	41 (16)	19 (6)	19 (7)	
/ra:s/□[la:s] **(head)**	≥6	70 (21)	52 (14)	37 (14)	26 (10)	16 (5)	14 (5)	
Glottal replacement to /ʔ/	≥4	77 (23)	48 (13)	45 (17)	36 (14)	10 (3)		
/nax.la/□[ʔax.la] **(palmtree)**	≥6	50 (15)	37 (10)	37 (14)	15 (6)			
Replacement to /h/	≥4	47 (14)	30 (8)	24 (9)				
/ħa.la:.wa/□[ha.la:.wa] **(candy)**	≥6	40 (12)	22 (6)	13 (5)				
De–emphasisation	≥4	100 (30)	93 (25)	79 (30)	51 (20)	45 (14)	19 (7)	
/ tˤɑˈma:tˤim /□[ta.ma:tim] **(tomato)**	≥6	100 (30)	81 (22)	63 (24)	38 (15)	35 (11)	16 (6)	
Denasalization	≥4	13 (4)						
/mak.wa/□[bak.wa] **(iron)**	≥6							
Metathesis	≥4	77 (23)	52 (14)	50 (19)	21(8)	10 (3)		
/ʔib.ti.da:.ʔi:/ □[daʔibˈda:ʔi:] **(primary school)**	≥6	63 (19)	41 (11)	21 (8)	13 (5)			
Assimilation	≥4	87 (26)	81 (22)	76 (29)	49 (19)	13(4)	19 (7)	27 (9)
/ʔasˤ.far/□[ʔar.far] **(yellow)**	≥6	73 (22	52 (14)	55 (21)	15 (7)		11(4)	
Heterosyllabic CC□ Gemination	≥4	50 (15)	63 (17)	26 (10)	10 (4)			
/ʕan.ka.bu:t/□[ʕak.ka.bu:t] **(spider)**	≥6	33 (10)	30 (8)	16 (6)				
Devoicing	≥4	77 (23)	70 (19)	63 (24)	44 (17)	29 (9)	16 (6)	24 (8)
/ruz/□[rus] **(rice)**	≥6	70 (21)	63 (17)	39 (15)	28 (11)	23 (7)	11 (4)	15 (5)
Voicing	≥4	23 (7)	19 (5)	13 (5)				
/ta.mur/□[da.mur] **(date)**	≥6	10 (3)						
	≥4		15 (4)	11 (4)	15 (6)	16 (5)	30 (11)	33 (11)

/dˤ/□[ðˤ] /be:dˤ/□ [be:ðˤ] **(egg)**	≥6				10 (4)		22 (8)	
Vowel substitution	≥4	100 (30)	89 (24)	87 (33)	72 (28)	65 (20)	54 (20)	58 (19)
/ma.gasˤ/□[mɛgasˤ] **(scissor)**	≥6	97 (29)	89 (24)	79 (30)	49 (19)	42 (13)	41 (15)	33 (11)
Vowel deletion	≥4	13 (4)	11 (3)					
/ˈminʃafa/□[minʃaf] **(towel)**	≥6							
Structural simplifications								
Weak Syllable deletion	≥4	87 (26)	52 (14)	47 (18)	23 (9)			
/mu.kaj.jif/□[kaj.jif] **(air conditioner)**	≥6	67 (20)	41 (11)	29 (11)	10 (4)			
Syllable coda WW deletion	≥4	70 (21)	52 (14)	32 (12)				
/kur.si:/ □[ku.si:] **(chair)**	≥6	53 (16)	37 (10)	16 (6)				
Word-final CC reduction	≥4	30 (9)	19 (5)	13 (5)				
/gird/□[gid] **(monkey)**	≥6							
Word-final consonants deletion	≥4	53 (16)	52 (14)	32 (12)	31 (12)	10 (3)		
/ni.mir/□[ni.mi] **(tiger)**	≥6	33 (10)	41(11)	16 (6)	13 (5)			
Word initial consonant deletion	≥4	33 (10)	33 (9)	11 (4)				
/ħi.sˤa:n/□[isˤa:n**] (horse)**	≥6	20 (6)	15 (4)					
Reduplication	≥4	40 (12)	15 (4)					
/ matˤ.bax/ □[makmak] (kitchen)	≥6	13(4)						
Phonetic distortions								
Distortions of fricatives	≥4	97 (29)	89 (24)	89 (34)	54 (21)	61 (19)	59 (22)	52 (17)
/minʃafa/□[minçafa] **(towel)**	≥6	90 (27)	78 (21)	82 (31)	28 (11)	28 (16)	49 (18)	42 (14)
Distortions of /r/	≥4	70 (21)	33 (9)	39 (15)	54 (21)	45 (14)	27 (10)	
/riʒil/□[ɹiʒil] **(leg)**	≥6	63 (19)	22 (6)	21 (8)	41 (16)	35 (11)	24 (9)	