PRO—FILING

TYPICAL AND DISORDERED

SPEECH PRODUCTION

IN CHILDREN

DONDERS SERIES

LEENKE VAN HAAFTEN

PROFILING TYPICAL AND DISORDERED SPEECH PRODUCTION IN CHILDREN

USING THE COMPUTER ARTICULATION INSTRUMENT (CAI)

LEENKE VAN HAAFTEN

Profiling typical and disordered speech production in children using the Computer Articulation Instrument (CAI)

ISBN:	978-94-6423-007-9
Cover, concept & design:	Ilse Schrauwers, isontwerp.nl
Typesetting headers:	Leenke Regular (by isontwerp.nl & Autograph, based on Yanone Kaffeesatz)
Typesetting body text:	Noto Sans JP (by Google)
Printed:	ProefschriftMaken.nl

Profiling typical and disordered speech production in children using the Computer Articulation Instrument (CAI)

Proefschrift

ter verkrijging van de graad van doctor
aan de Radboud Universiteit Nijmegen
op gezag van de rector magnificus prof. dr. J.H.J.M. van Krieken,
volgens besluit van het college van decanen
in het openbaar te verdedigen op donderdag 17 december 2020
om 11.30 uur precies

door

Leenke van Haaften
geboren op 6 oktober 1980
te Breda

Promotor: Prof. dr. B.A.M. Maassen (Rijksuniversiteit Groningen)

Copromotoren: Dr. L. van den Engel-Hoek
Dr. B.J.M. de Swart

Manuscriptcommissie: Prof. dr. M.A.A.P. Willemsen
Prof. dr. J.P.M. Fikkert
Dr. R. Jonkers (Rijksuniversiteit Groningen)

FOR THE THINGS WE HAVE TO LEARN BEFORE WE CAN DO THEM, WE LEARN BY DOING THEM ARISTOTLE

TABLE OF CONTENTS

~

ONE

CHAPTER 1

GENERAL INTRODUCTION

Speech, speech development and speech disorders

During the first years of life, infants are in the process of learning to produce the speech sounds in their ambient language. The oral motor, cognitive and linguistic system gradually mature and infants gradually acquire strategies to approximate adult-like speech. These approximations result in systematic speech symptoms such as substitutions, omissions and distortions of sounds (e.g. [pɪn] for /spɪn/; [nana] for /banana/, [tu] for /shoe/). In general, most productions of children resemble the adult-like target production at the age of five (Dodd, 2011). Typical speech sound development can be described as the acquisition of individual speech sounds and the organization of these speech sounds into speech patterns. Children show gradual progress in intelligibility, phonetic or articulatory development and phonemic or phonological development. Some children have persistent difficulty producing words or sounds correctly. These children are diagnosed with a speech delay or a speech sound disorder (SSD).

The term SSD encompasses both expressive phonological problems (knowledge and use of the speech sounds and sound patterns of one's language) and problems with speech production that have motor or physical origins or involve misarticulations such as lisp. Misarticulations are defined as sounds that are produced in a distorted way without losing the contrast with other sounds (Bishop et al., 2017). SSD is a specific subtype of developmental language disorder (DLD), like other subtypes such as lexical semantic impairment, syntactic impairment, pragmatic difficulties, difficulties with auditory conceptualization, or with verbal sequential memory (Bishop et al., 2017; Van Weerdenburg, Verhoeven, & Van Balkom, 2006). Subtypes often co-occur and pure cases hardly ever exists. Children with SSD form a heterogeneous group, showing variability in severity, etiology, proximal causes, speech characteristics, and response to treatment (Dodd, 2014; Ttofari Eecen, Eadie, Morgan, & Reilly, 2018). Prevalence rates for speech delay or SSD ranging from 2.3% to 24.6% are reported (Wren, Miller, Peters, Emond, & Roulstone, 2016). In 2018, up to 23% of the caseloads of Dutch speech language pathologists (SLPs) consisted of children with SSD (Verberne, van den Dool, & Schermer, 2019). The text box below presents a case of a boy with suspected SSD.

Case example
Tom is referred to an SLP by his family doctor at the age of 3;1 years. He produced a variety of speech-like vocalizations as an infant, but no canonical babbling. He expressed his first words at the age of 1;6 years. At this time, there are no concerns about his language comprehension and language expression. He understands verbal instructions in daily situations. His parents report other developmental milestones within normal limits and no familial history of speech language disorders. Audiological assessment showed no significant hearing loss.

Within a known conversational context, his parents can usually understand him. However, without the context, he is difficult to understand, and his speech is unintelligible for unfamiliar listeners. The last few months he gets frustrated when he is not understood.

His speech is characterised by omissions and substitutions of sounds. For example, he produces [tɑp-tɔk] for /kɑp-stɔk/ (Dutch for coat rack). In this example he substitutes word initial /k/ with [t], but /k/ is correctly pronounced word finally. In the second syllable, the consonant cluster /st/ is substituted by a single consonant [t].

Tom's parents are concerned about his speech sound development and wonder whether speech language therapy should be started.

The SLP's role for children like Tom is to collect and analyse information (diagnostic reasoning) to be able to identify an SSD, i.e. speech delay or speech disorder (including diagnostic categorization), in order to make decisions in treatment planning, thereby taking into account the child's circumstances and needs (therapeutic reasoning). These two elements of *clinical reasoning* – or practice decision making (Higgs, Jones, Loftus, & Christensen, 2008) – are especially important in a vulnerable population like children with SSD. They are at risk of limitations and restrictions in several domains of the International Classification of Functioning, Disability and Health (ICF) (WHO, 2007). Children with SSD have difficulties with learning and applying knowledge, communication, mobility, interpersonal interactions and relationships and major life areas like school education and keeping and finding a job (McCormack, McLeod, McAllister, & Harrison, 2009).

Measuring speech development and speech sound disorders

To be able to correctly identify and diagnose children with SSD, it is important to have adequate assessment tools. Because of its important role in clinical decision making, an assessment instrument must provide reliable and valid results. It must fulfill a number of psychometric criteria: (a) evidence of reliability, (b) evidence of validity, and (c) a representative normative sample with an adequate size which is related to generalizability.

Reliability can be defined as the degree to which the measurement is free from measurement error (Mokkink et al., 2010). There are several forms of reliability. To assess SSD two types of reliability were especially described in psychometric reviews of tests (Fabiano-Smith, 2019; Flipsen & Ogiela, 2015; Kirk & Vigeland, 2014). *Test-retest reliability* is the extent to which each constituent task measures the target construct consistently across time, and *interrater reliability* measures the target construct across raters on the same occasion (Mokkink et al., 2010). Test-retest reliability and high interrater reliability of a speech test give confidence that a test is consistent in its ability to measure speech sound production.

Validity can be defined as the degree to which the test truly measures the construct it purports to measure (Mokkink et al., 2010). In most situations, the first step in test construction is aimed at *content* or *face validity*, i.e. the fact that the content of the instrument contains items that are identical to the criterion behaviour. In the context of this manuscript, the question would be whether the test examines speech sound production. The second step in test construction is criterion validity, which refers to how well the scores of the instrument agree with the scores on a gold standard (De Vet, Terwee, Mokkink, & Knol, 2011). Another aspect of validity is *construct validity*, which is defined by the COSMIN (COnsensus-based Standards for the selection of health Measurement INstruments) panel as the degree to which the scores of a measurement instrument correspond with the construct that the instrument is intended to measure (Mokkink et al., 2010).

For the clinical assessment of children with SSD the availability of *normative data* is essential. Normative data yield the criterion to differentiate children with delayed or disordered speech development on the one hand from typically developing children on the other (Dodd, Holm, Hua, & Crosbie, 2003). When evaluating the normative sample of a test, several criteria are important (Fabiano-Smith, 2019; Flipsen & Ogiela, 2015; Kirk & Vigeland, 2014). One criterion is the size of the normative sample. The Dutch Committee on Tests and Testing (COTAN) recommends that each subgroup of a normative sample includes at least 100 children (Evers, Lucassen, Meijer, & Sijtsma, 2010). Second, demographic factors such as socioeconomic status (SES), gender, and geographic region of the normative sample should be representative for the demographic factors of the entire population. Normative data must contain recent information, and information of individuals for whom the test is intended (i.e. a representative portion of the sample must include children with speech sound disorders). Recency of normative data is important because it must represent the current population. As demographics of population change over time, normative data must be updated periodically. McFadden (1996) argued that normative data must contain a proportion of children with SSD to reduce the risk of children at the low end of the normal range being misidentified as having a disorder. Andersson (2005) proposed to include a proportion of 5% children with SSD in a normative sample.

Kirk and Vigeland (2014) published a review of the psychometric properties of norm-referenced tests used in the United States of America to measure phonological error patterns. The six tests included in their review did not meet all of the psychometric properties required of well-designed norm-referenced tests. They especially lacked adequate sample size, contained poor evidence of construct validity, and lacked information about diagnostic accuracy. The review of Flipsen and Ogiela (2015) examined the psychometric characteristics of ten single-word tests of English speech sound production. Although they found improvement in the state-of-the-art for English single-word tests of speech sound production when compared to the review study of McCauley and Swisher in 1984, most tests did not fulfill the total range of psychometric criteria.

How are we assessing SSD in the Netherlands? In clinical practice, various speech assessments are available for the Dutch language. Our research group investigated survey data on common assessments used by Dutch SLPs (Diepeveen, Van Haaften, Terband, De Swart, & Maassen, in press). None of the used assessments are norm-based or provide information about reliability and validity, except for the Articulation subtest of the TAK (subtest Klankarticulatie, Taaltoets Alle Kinderen), a Dutch Language Proficiency Test for All Children (Verhoeven & Vermeer, 2001). This test elicits speech by word imitation, and a correct-incorrect analysis is used. For the other language domains (lexical semantic, syntactic, pragmatic, auditory conceptualization, and verbal sequential memory) sufficient adequate diagnostic tools are available with information about normative data, reliability and validity. An adequate speech assessment test, however, is lacking, despite the fact that several normative studies have been carried out on speech sound development of Dutch-speaking children (Beers, 1995; Fikkert, 1994; Jongstra, 2003; Levelt, 1994; Levelt, Schiller, & Levelt, 2000; Priester & Goorhuis-Brouwer, 2013; Stes, 1977; Van den Berg, Van Severen, Molemans, & Gillis, 2017). These studies collected normative data, but no assessment instruments were developed. Furthermore, these normative data show several limitations. First, in most studies the sample size of the subgroups was not large enough to draw generalizations. Second, no or insufficient information was given with respect to demographic factors such as SES, gender, and geographic region, and third, children with SSD were not included in the normative sample. Thus, assessment procedures and normative data are called for to adequately characterize typical and delayed or disordered speech development.

Most of the Dutch speech assessment instruments make use of picture naming or word imitation. With these elicitation methods it is possible to identify specific speech symptoms. For example, the number and type of sound substitutions or omissions can be defined and phonological error patterns can be identified. Like in the example of Tom, when /kɑp-stɔk/ is produced as [tɑp-tɔk], a substitution of a syllable-initial consonant and a reduction of a consonant cluster can be determined. Also, phonological error patterns, in this case 'fronting' (in the first consonant), can be observed.

Currently available tests yield detailed descriptions of speech symptoms (i.e. speech behaviour), but one of their major shortcomings of is that they do not directly assess the underlying speech production processes involved. Instead of classifying SSD based on symptoms, one should focus on underlying deficits, such as the distinction between processing or representational deficits. Effective differential diagnosis requires a process-oriented approach (Terband, Maassen, & Maas, 2019). Psycholinguistic and psychomotor models form the basis for a process-oriented diagnostic classification system based on the identification of the breakdown in the chain of sequential and parallel speech processes (Baker, Croot, McLeod, & Paul, 2001). According to these models, the speech production process starts with retrieving word forms from the lexicon, which forms the input for phonological encoding, followed by motor planning and motor programming, and finally resulting in motor execution. Terband et al. (2019) give a detailed description of an integrated psycholinguistic model of speech production. They argued that effective diagnosis and treatment planning require a dynamic process-oriented approach, aimed at describing the development of underlying processing deficits to characterise SSD. The main objective of a process-oriented approach is not to categorize, but to give a complete characterization of the speech profile, such that underlying processing deficits can be identified.

In the example of Tom ([tɑp-tɔk] for /kap-stok/), one of the possible interpretations of the substitution of /k/ to [t] is that this substitution is a phonological error and indicates difficulties at the level of phonological encoding. In contrast, an alternative explanation of the substitution of /k/ to [t] is that this substitution is a consequence of difficulties at the level of motor planning and/or programming. According to the speech production model, the selection and sequencing of the same articulatory movement goals (two times /t/) is considered to be easier than two different articulatory movement goals (first /k/ in the initial syllable, and then /t/ in the second syllable). Thus, one speech symptom may present different underlying deficits. This example reveals that researchers and practitioners need to move away from behavioural measures of speech production and use instead measures of underlying processes. This may involve comparing speech behaviour across different conditions (e.g., word repetition vs. nonword repetition; assessing phonological vs. speech motor skills). Insight into the deficits that might be the underlying causes of the child's difficulty requires an extensive analysis of a child's performance on a range of speech tasks that reflect different levels of processing.

A distinction between the level of observed behaviour and underlying cognitive processes is made in a causal model provided by Bishop and Snowling (2004). This model describes four levels: etiology, neurobiology, cognitive processes and observed behaviour. The first, etiological level encompasses both genetic and environmental factors. The second, neurobiological level includes abnormalities of brain structure and function. Genetic factors determine neurobiological factors, and these in turn influence both the third level of cognitive processes and the fourth, behavioural level. Contrariwise, children's behaviour and cognition can affect the neurobiological factors. The description of the underlying cognitive deficits that explain the behavioural deficit (proximal causes) is a step closer to the genetic factors that causes SSD. Underlying deficits represent endophenotypes, that is, a set of deficient processes, arguably closer to the genotype than the observed behaviour itself (Snowling, 2008). Morgan (2013) stated that SLPs can make a significant contribution to defining the endophenotype of SSDs. Knowledge of the genetic bases of SSD will have impact on clinical practice since information on etiology can improve counselling and prognosis.

In summary, assessing SSD requires the availability of an assessment instrument with sufficient psychometric properties. Second, an assessment is required that evaluates different aspects of speech production, so that different aspects can be compared with each other in order to obtain a complete speech profile. Up to the start of this project, no comprehensive test procedures exists in the Netherlands that yield both of these requirements.

Aims and outline of the thesis

The overall aim of this thesis was to characterize the *speech profiles* of typically developing children and children with speech delay or SSD, such that a first, significant step towards process-oriented diagnostics is taken. Insight into the deficits that might be the underlying causes of the SSD requires an extensive analysis of a child's performance on a range of speech tasks that reflect different underlying processes: lemma access, word form selection, phonological encoding, motor planning and motor programming, and motor execution. Based on these premises, the Computer Articulation Instrument (CAI) was developed (Maassen et al., 2019). The aims of this thesis' studies are:

1. To construct a speech production test battery that comprises a range of speech tasks, for the assessment of a comprehensive speech profile.
2. To collect normative data on the development of speech profiles in typically developing children and analyse the psychometric properties of the CAI in terms of interrater and test-retest reliability, and different aspects of construct validity.
3. To provide a comprehensive analysis of speech sound development in Dutch-speaking children based on a large cross-sectional study as a background for interpreting clinical test results.
4. To describe and analyse clinical speech characteristics of speech delay and specific speech deficits in children with diverse neurodevelopmental disorders.

This thesis is divided into two sections. **Part One** focusses on the construction of the CAI and the description of typical speech sound development in Dutch (Aim 1, 2, and 3). In Chapter 2 the rationale of the speech tasks of the CAI are described, the norm study is presented, and an analysis is given of the psychometric properties, comprising interrater and test-retest reliability and two aspects of construct validity. The CAI consists of a battery of four speech production tasks, has a modular structure, and provides an interactive administration and scoring facility. The tasks comprise (1) picture naming (PN), (2) nonword imitation (NWI), (3) word and nonword repetition (WR and NWR) and (4) maximum repetition rate (MRR), thereby covering phonological and speech motor skills. Chapter 3 and Chapter 4 present the results of two studies on typical speech sound development in Dutch. *Chapter 3* gives a detailed description of speech sound development of Dutch speaking typically developing children aged 2;0 to 6;11 years. Speech sound development is described by different parameters: percentage of consonants correct (PCC), percentage of vowel correct (PVC), consonant, vowel and syllabic structure inventories, degrees of complexity (phonemic feature hierarchy) and phonological processes. In *Chapter 4* normative data are provided for the development of maximum repetition rate (MRR) in Dutch-speaking children aged 3;0 to 6;11 years based on a large cross-sectional study.

Part Two of this thesis addresses the first steps of the clinical implementation of the CAI (Aim 4). *Chapter 5* reports on two studies evaluating a third aspect of construct validity of the CAI: known-group validity. In the first study the scores on the picture naming task of the CAI are validated with intelligibility judgments in children diagnosed with speech language impairments. The second study investigates the relation between CAI factors and clinical judgments of severity of the speech disorder by SLPs in children with different types of SSD (phonological disorder, childhood apraxia of speech and unknown diagnosis). Based on these two studies the differential diagnostic power of the resulting speech profiles is determined. *Chapter 6* describes a first prospective study of oral motor, speech, language, literacy, and pragmatic social skills in a large cohort of children with Koolen de Vries syndrome, using standardised tests normed for typical behaviour, among which the CAI, to precisely characterise the communication phenotype associated with this syndrome.

Finally, the general discussion is provided in *Chapter 7*. In this chapter the results of the studies are summarized and discussed in the light of the aims of this thesis. Perspectives on future research are discussed.

References

Andersson, L. (2005). Determining the adequacy of tests of children's language. *Communication Disorders Quarterly, 26*(4), 207-225.

Baker, E., Croot, K., McLeod, S., & Paul, R. (2001). Psycholinguistic models of speech development and their application to clinical practice. *Journal of Speech, Language, and Hearing Research, 44*(3), 685-702.

Beers, M. (1995). *The Phonology of normally developing and language-impaired children.* University of Amsterdam, Amsterdam: IFOTT.

Bishop, D. V. M., & Snowling, M. J. (2004). Developmental dyslexia and specific language impairment: Same or different? *Psychological bulletin, 130*(6), 858-886.

Bishop, D. V. M., Snowling, M. J., Thompson, P. A., et al., A. (2017). Phase 2 of CATALISE: a multinational and multidisciplinary Delphi consensus study of problems with language development: Terminology. *Journal of Child Psychology and Psychiatry, 58*(10), 1068-1080.

De Vet, H. C., Terwee, C. B., Mokkink, L. B., & Knol, D. L. (2011). *Measurement in medicine: a practical guide*: Cambridge University Press.

Diepeveen, S., Van Haaften, L., Terband, H., De Swart, B., & Maassen, B. (in press). Clinical reasoning in the field of SSD: Diagnosis and intervention in SLTs' daily practice. *American Journal of Speech-Language Pathology.*

Dodd, B. (2011). Differentiating speech delay from disorder: does it matter? *Topics in Language Disorders, 31*(2), 96-111

Dodd, B. (2014). Differential diagnosis of pediatric speech sound disorder. *Current Developmental Disorders Reports, 1*(3), 189-196.

Dodd, B., Holm, A., Hua, Z., & Crosbie, S. (2003). Phonological development: a normative study of British English speaking children. *Clinical Linguistics & Phonetics, 17*(8), 617-643.

Evers, A., Lucassen, W., Meijer, R., & Sijtsma, K. (2010). *COTAN Beoordelingssysteem voor de kwaliteit van tests (COTAN Assessment system for the quality of tests)*. Amsterdam: NIP.

Fabiano-Smith, L. (2019). Standardized tests and the diagnosis of speech sound disorders. *Perspectives of the ASHA Special Interest Groups, 4*(1), 58-66.

Fikkert, P. (1994). *On the acquisition of prosodic structure.* University of Leiden, Dordrecht; ICG Printing.

Flipsen, J. 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.

Higgs, J., Jones, M. A., Loftus, S., & Christensen, N. (2008). *Clinical reasoning in the health professions*: Elsevier Health Sciences.

Jongstra, W. (2003). *Variation in reduction strategies of Dutch word-initial consonant clusters.* University of Toronto.

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, 45*(4), 365-377.

Levelt, C. (1994). *On the acquisition of place.* University of Leiden, The Hague: HAG.

Levelt, C., Schiller, N., & Levelt, W. (2000). The acquisition of syllable types. *Language acquisition, 8*(3), 237-264.

Maassen, B., van Haaften, L., Diepeveen, S., Terband, H., van den Engel-Hoek, L., Veenker, T., & De Swart, B. (2019). *Computer Articulation Instrument*. Amsterdam: Boom Uitgevers.

McCauley, R. J., & Swisher, L. (1984). Psychometric review of language and articulation tests for preschool children. *Journal of Speech and Hearing Disorders, 49*(1), 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.

McFadden, T. U. (1996). Creating language impairments in typically achieving children. *Language, Speech, and Hearing Services in Schools, 27*(1), 3-9.

Mokkink, L., Terwee, C., Patrick, D., Alonso, J., Stratford, P., Knol, D., . . . De Vet, H. (2010). The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *Journal of Clinical Epidemiology, 63*(7), 737-745.

Morgan, A. (2013). Speech-language pathology insights into genetics and neuroscience: Beyond surface behaviour. *International Journal of Speech-Language Pathology, 15*(3), 245-254.

Priester, G. H., & Goorhuis-Brouwer, S. M. (2013). Measuring speech sound development: An item response model approach. *International Journal of Pediatric Otorhinolaryngology, 77*, 1469-1473.

Snowling, M. J. (2008). Specific disorders and broader phenotypes: The case of dyslexia. *The Quarterly Journal of Experimental Psychology, 61*(1), 142-156.

Stes, R. (1977). Normatieve studie over articulatieontwikkeling bij Nederlandssprekende kinderen tussen 3-8 jaar (Normative study on articulation development of Dutch speaking children between 3-8 years of age). *Tijdschrift voor Logopedie en Audiologie, 7*, 140-149.

Terband, H., Maassen, B., & Maas, E. (2019). A psycholinguistic framework for diagnosis and treatment planning of developmental speech disorders. *Folia Phoniatrica Et Logopaedica, 71,* 216-227.

Ttofari Eecen, K., Eadie, P., Morgan, A. T., & Reilly, S. (2018). Validation of Dodd's Model for Differential Diagnosis of childhood speech sound disorders: a longitudinal community cohort study. *Developmental Medicine & Child Neurology, 61,* 689-696.

Van den Berg, R., Van Severen, L., Molemans, I., & Gillis, S. (2017). Syllable type development in toddlers acquiring Dutch. *Dutch Journal of Applied Linguistics, 6*(1), 119-146.

Van Weerdenburg, M., Verhoeven, L., & Van Balkom, H. (2006). Towards a typology of specific language impairment. *Journal of Child Psychology and Psychiatry, 47*(2), 176-189.

Verberne, L., van den Dool, J., & Schermer, T. (2019). *Zorg door de logopedist. Jaarcijfers 2018 en trendcijfers 2016-2018.* Utrecht: Nivel.

Verhoeven, L., & Vermeer, A. (2001). *Taaltoets Alle Kinderen*. Arnhem: Citogroep.

WHO. (2007). *Nederlandse vertaling van de International Classification of Functioning, Disability and Health, Children & Youth version: ICF-CY.* Houten: Bohn Stafleu van Loghum.

Wren, Y., Miller, L., Peters, T., Emond, A., & Roulstone, S. (2016). Prevalence and predictors of persistent speech sound disorder at eight years old: Findings from a population cohort study. *Journal of Speech, Language, and Hearing Research, 59*(4), 647-673.

ONE

THE CONSTRUCTION OF THE CAI AND TYPICAL SPEECH SOUND DEVELOPMENT IN DUTCH PART 1

TWO

CHAPTER 2

THE PSYCHOMETRIC EVALUATION OF A SPEECH PRODUCTION TEST BATTERY FOR CHILDREN

THE RELIABILITY AND VALIDITY OF THE COMPUTER ARTICULATION INSTRUMENT

Leenke van Haaften, Sanne Diepeveen, Lenie van den Engel-Hoek, Marianne Jonker, Bert de Swart and Ben Maassen

JOURNAL OF SPEECH, LANGUAGE AND HEARING RESEARCH
2019; 62; 2141-2170

Abstract

Purpose: The aims of this study were to assess the reliability and validity of the Computer Articulation Instrument (CAI), a speech production test battery assessing phonological and speech motor skills in 4 tasks: (1) picture naming, (2) nonword imitation, (3) word and nonword repetition, and (4) maximum repetition rate (MRR).

Method: Normative data were collected in 1,524 typically developing Dutch-speaking children (aged between 2;0 and 7;0 [years;months]). Parameters were extracted on segmental and syllabic accuracy (Tasks 1 and 2), consistency (Task 3), and syllables per second (Task 4). Interrater reliability and test–retest reliability were analyzed using subgroups of the normative sample and studied by estimating intraclass correlation coefficients (ICCs). Construct validity was investigated by determining age-related changes of test results and factor analyses of the extracted speech measures.

Results: ICCs for interrater reliability ranged from sufficient to good, except for percentage of vowels correct of picture naming and nonword imitation and for the MRRs for bisyllabic and trisyllabic items. The ICCs for test–retest reliability were sufficient (picture naming, nonword imitation) to insufficient (word and nonword repetition, MRR) due to larger-than-expected normal development and learning effects. Continuous norms showed developmental patterns for all CAI parameters. The factor analyses revealed 5 meaningful factors: all picture-naming parameters, the segmental parameters of nonword imitation, the syllabic structure parameters of nonword imitation, (non)word repetition consistency, and all MRR parameters.

Conclusion: Its overall sufficient to good psychometric properties indicate that the CAI is a reliable and valid instrument for the assessment of typical and delayed speech development in Dutch children in the ages of 2–7 years.

Introduction

A major task for speech-language therapists (SLTs) is to differentiate children with delayed or disordered speech development from typically developing peers and to determine eligibility for speech services. For such an assessment, they should be able to rely on standardized tests and normative data. However, several reviews that evaluated the content and psychometric characteristics of speech assessments in English and other languages (Flipsen & Ogiela, 2015; Kirk & Vigeland, 2014; McCauley & Strand, 2008; McLeod & Verdon, 2014) concluded that, overall, the diagnostic tests reported on tend to lack fundamental psychometric properties, while sample sizes used for norm referencing were inadequate, and evidence of reliability and validity were poorly described.

Various speech assessments are available for the Dutch language. A survey by our research group (Diepeveen, Van Haaften, Terband, De Swart, & Maassen, submitted) revealed that the vast majority of SLTs (75.8%) in the Netherlands use "LOGO-Art Dutch Articulation Assessment" (Nederlands ArticulatieOnderzoek; Baarda, de Boer-Jongsma, & Jongsma, 2013), with 50.0% (also) using the Dutch version of the Metaphon Screening Assessment (Leijdekker-Brinkman, 2002). The Dutch version of the Hodson Assessment of Phonological Patterns is used by 31.1% (Van de Wijer-Muris & Draaisma, 2000), whereas another 30.3% evaluate a spontaneous speech sample; 22.7% administer the "Dyspraxia Program" similar to the Nuffield Dyspraxia Program (Eurlings-van Deurse, Freriks, Goudt-Bakker, Van der Meulen, & Vries, 1993), 13.6% use oral motor assessments, whereas 9.80% use a qualitative observation based on the Motor Speech Hierarchy framework used for PROMPT therapy: Verbal Motor Production Assessment for Children (Hayden, 2004), and 8.30% use the Articulation subtest (Klankarticulatie subtest) of the TAK (Taaltoets Alle Kinderen), a Dutch Language Proficiency Test for All Children (Verhoeven & Vermeer, 2001), with 4.50% employing their own (custom-made) speech assessments. None of these assessments is norm based or provides information about reliability and validity except for the TAK (Verhoeven & Vermeer, 2001). For the TAK, normative data are provided based on a representative normative group of 807 children with an age range of 4;7–8;3 (years; months), and the manual states that the test's reliability and validity were sufficient to good (Verhoeven & Vermeer, 2006).

Moreover, all tests measure only one aspect of speech production. The production of speech sounds is a complex process that comprises both linguistic (or phonological) and speech motor aspects. Psycholinguistic models of speech production describe speaking as a series of sequential and parallel processes, where the first is the conceptualization of a preverbal message, either from memory or from perception, as occurs in picture naming. The next process is formulating a word or sentence, which is driven by two steps of lexicalization: the selection of a *lemma*, containing meaning and grammatical information, and the corresponding lexeme or word form. The *lexeme* constitutes the input for the next stage, phonological encoding, during which the sequence of speech sounds is specified together with the syllabic and prosodic structures. Syllables are the basic units of the next process: articulomotor planning and programming. The final process is execution, where the articulatory movements are performed, resulting in an acoustic speech signal (Maassen & Terband, 2015). Children with speech production deficits or speech sound disorders (SSDs) can experience problems at the level of lexeme retrieval, phonological encoding, articulomotor planning and programming, and/or execution. Speech assessment should evaluate these different aspects of SSDs to be able to obtain a complete speech profile. The LOGO-Art Dutch Articulation Assessment (Baarda et al., 2013) analyzes speech in terms of substitution errors in initial, medial, and word-final positions (three-position test). The Articulation subtest of the TAK (Verhoeven & Vermeer, 2001) comprises a word imitation test, with each of its 45 items being dichotomously scored (correct or incorrect) without any further analysis of speech errors. The Dutch version of the Metaphon Screening Assessment (Leijdekker-Brinkman, 2002) and Hodson Assessment of Phonological Patterns (Van de Wijer-Muris & Draaisma, 2000) are scored based on phoneme inventories and the analysis of phonological processes. The Dyspraxia program (Eurlings-van Deurse et al., 1993) and the Verbal Motor Production Assessment for Children (Hayden, 2004) assess speech motor abilities such as sequencing. To date, there is no Dutch test that systematically assesses speech performance using a broad set of tasks while providing norm data that allow a speech profile to be compiled. Such a comprehensive speech profile is the first step toward a process-oriented diagnosis in which underlying deficits are identified. Because the available diagnostic tools merely yield a description at the symptom level without assessing other aspects of speech production or providing norm-referenced scores, we developed the Computer Articulation Instrument (CAI).

The CAI is a computer-based speech production test battery consisting of four tasks that we based on a series of studies in children with developmental and acquired SSDs (Nijland, 2003; Thoonen, 1998). It has a modular structure and requires interactive administration. Gauging both phonological and speech motor skills of children in the ages between 2 and 7 years, the tasks comprise (a) picture naming, (b) nonword imitation, (c) word and nonword repetition, and (d) maximum repetition rate (MRR). As demonstrated in Figure 1, picture naming taps into the whole chain of speech processes, from preverbal visual–conceptual processing to lemma access, word form selection, phonological encoding, motor planning, and articulation (motor execution; Maassen & Terband, 2015). During nonword imitation, a child is asked to reproduce nonwords (or nonsense words). In contrast

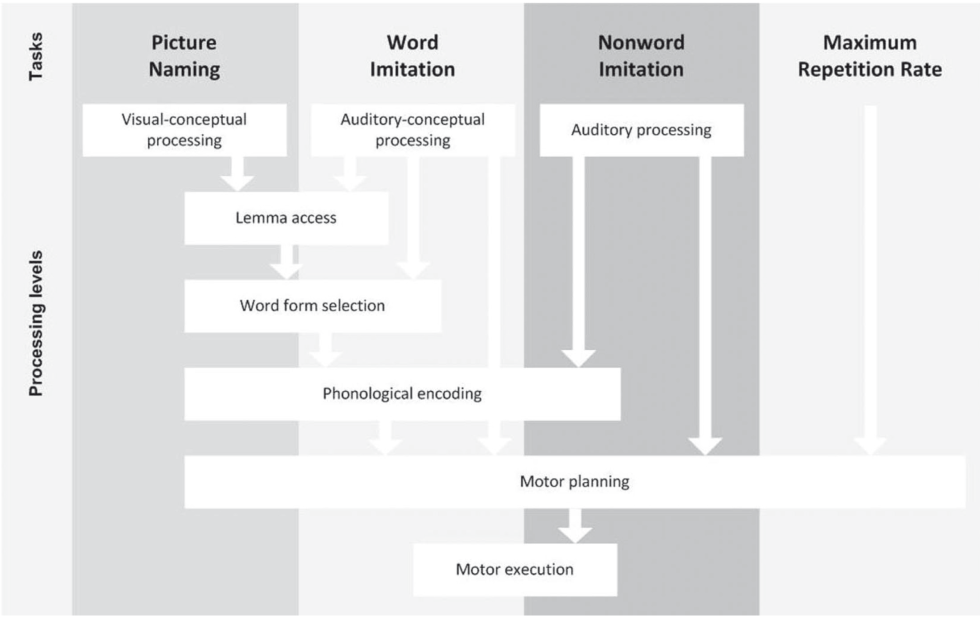

Figure 1. The speech production processes assessed in the four tasks of the Computer Articulation Instrument (based on Maassen & Terband, 2015, Figure 15.2).

to picture naming, a child cannot revert to its lexicon during this task and thus either analyzes the phonological structure of the nonword directly, addressing the phonological encoding system, or follows the auditory-to-motor planning pathway. In word and nonword repetition, a child is asked to repeat a word or nonword five times. This task aims to assess variability in speech production, which occurs when a child uses multiple productions of the same word or nonword. MRR is a pure motor task (articulomotor planning and programming) and does not require any knowledge of words, syllables, or phonemes. In the CAI, the evaluation of speech production is based on phonetic transcriptions and acoustic measurements. Both the tasks and speech analyses are computer implemented. Further explanation of the rationale of the speech tasks and administration procedures are presented in the Method section.

With the CAI, we sought to develop a speech assessment instrument that allows the detection of signs of delay or deviance in several speech production characteristics such that a norm-referenced speech profile for Dutch-speaking children could be obtained. Our ultimate goal for the CAI is to identify and classify children with SSDs. In this article, we will discuss the content of the instrument and the collection of normative data, as well as report on its psychometric properties in terms of interrater and test–retest reliability and its construct validity. Defining reliability as “the degree to which the measurement is free from measurement error” (Mokkink, Terwee, Patrick, et al., 2010), we examined the extent to which each constituent task measures the target construct consistently across time (test–retest) and across raters on the same occasion (interrater; Mokkink, Terwee, Patrick, et al., 2010).

The second aim of this study was to determine the validity of the CAI, which we defined as the degree to which it truly measures the construct it purports to measure (Mokkink, Terwee, Patrick, et al., 2010). In most situations, the first step in test construction is aimed at content or face validity, that is, whether the content of the instrument corresponds with the construct that the instrument is intended to measure. We will demonstrate the content validity of the CAI by the description of the test domain (articulation) and its four speech tasks. The second step in test construction is criterion validity, which refers to how well the scores of the instrument agree with the scores on a gold standard (De Vet, Terwee, Mokkink, & Knol, 2011). In situations in which there is no gold standard, as is the case for speech development in Dutch, one has to fall back on construct validity, which is defined by the COSMIN (COnsensus-based Standards for the selection of health Measurement INstruments) panel as the degree to which the scores of a measurement instrument are consistent with hypotheses (Mokkink, Terwee, Knol, et al., 2010). In our study, we thus investigate two aspects of construct validity.

Because the primary aim of the CAI is to measure speech development, we investigated this first aspect of construct validity by comparing the raw scores or parameters of its tasks in a large sample of typically developing children aged between 2 and 7 years. One of the requirements of a developmental test is that the outcomes show a correlation with age. We hypothesized that the selected parameters, such as the percentage of consonants correct (PCC), percentage of vowels correct (PVC), percentage of cluster reductions, and percentages of particular correctly produced syllable structures, would reflect typical speech development and would thus show a monotonous improvement with age. In order to examine the second aspect of construct or structural validity (De Vet et al., 2011), we conducted factor analyses based on the assumption that clusters of the selected parameters would reflect different aspects of speech production, either within or across tasks, with the parameters' factor structures contributing to the definition of individual speech profiles.

Method

Participants

A total of 1,524 typically developing children aged between 2;0 and 7;0 participated in the normative study. Stratifying for age, we created 14 groups with a range of 4 months for children aged 2;0–5;11 and a range of 6 months for those aged 6;0–6;11. Table 1 summarizes the characteristics of our sample. For the assessment of speech-language development, each age group of a normative sample should contain at least 100 individuals (Andersson, 2005). As Table 1 shows, all our age groups contained more than 100 children, except for the youngest age group (*n* = 72).

The participants were drawn from 47 nurseries and 71 elementary schools located in four different regions of the Netherlands (see Table 2). The nurseries and schools were sent a letter explaining the purpose of the study and inviting them to participate. All parents of the children in the participating nurseries and schools were given an information letter. After obtaining the signed parental consent form, the child was included in the study. To reach the required number of children for each age group in particular geographic regions, assessors randomly selected children from those for whom parental consent had been obtained.

Criteria for inclusion were no hearing loss and Dutch being the spoken language at the nursery or primary school. The parents and teachers of eligible children were asked to complete a questionnaire about the children's development. Another language than Dutch (e.g., Turkish, Arabic, or German) was spoken at home in 3.9% of the participants, with Dutch being the primary language for 96.7% of these children. To ensure the normative sample was representative of the

Table 1. Age, gender, and multilingualism for the 14 age groups of the normative sample.

Age group (years;months)	M_{age} (years;months)	Number of children	Gender (*n*)		Multilingual (*n*)
			Boys	Girls	
2;0–2;3	2;1	72	30	42	2
2;4–2;7	2;5	102	55	47	1
2;8–2;11	2;9	101	46	55	1
3;0–3;3	3;1	104	52	52	3
3;4–3;7	3;5	110	61	49	3
3;8–3;11	3;9	102	57	45	5
4;0–4;3	4;1	100	55	45	1
4;4–4;7	4;5	115	60	55	3
4;8–4;11	4;9	116	56	60	11
5;0–5;3	5;1	121	66	55	12
5;4–5;7	5;5	128	71	57	5
5;8–5;11	5;9	117	64	53	4
6;0–6;5	6;2	117	69	48	5
6;6–6;11	6;8	119	57	62	4
Total		1,524	799	725	60
% of sample		100	52.4	47.6	3.94

Dutch population, we also included children with a history of speech and language difficulties (*n* = 32, 2.2%). The 4- to 7-year-old children were recruited between January 2008 and December 2014; and the children in the younger age group (2–4 years), from March 2011 to April 2015.

Parental socioeconomic status (SES) was based on the social status of the district (zip code area) of the child's nursery or primary school as determined by the Netherlands Institute for Social Research (Knol, Boelhouwer, & Veldheer, 2012). The social status of a district was derived from a number of population characteristics, namely, education, income, and labor market position, with higher scores indicating a higher status for that particular district, with a mean of 0 (see Table 3).

The final sample was representative of the general Dutch population in terms of gender, geographic region, and degree of urbanization (see Tables 1 and 2). For example, in the north of the Netherlands, there are very few intermediately or densely populated areas, which is why all testing in that region was conducted in thinly populated areas.

Table 2. Number of children tested per geographic region and degree of urbanization.

	Degree of urbanization			
Region	**Thinly populated area (index 1.0–2.6)**	**Intermediate density area (index 2.7–4.0)**	**Densely populated area (index 4.1–4.8)**	**Total (%)**
North	128	0	0	128 (8.40)
East	150	212	0	362 (23.8)
South	252	109	0	361 (23.7)
West	341	159	173	673 (44.2)
Total (%)	871 (57.2)	480 (31.5)	173 (11.4)	1,524 (100)

Table 3. Parental socioeconomic status (SES) of the normative sample.

SES	***n***	**% of sample**
< -1	182	11.9
≥ -1 and < 1	1,104	72.4
≥ 1	238	15.6
Total	1,524	100

Material: CAI

Tasks

The CAI consists of four tasks: picture naming, nonword imitation, word and nonword repetition, and MRR. The tasks were administered by (candidate) SLTs specifically trained in the administration of the CAI (for more details, see Procedure section). All utterances were audio-recorded and stored in the CAI database.

Picture naming. Picture naming consists of 60 items. For each item, the child's utterances are compared with the target words. Picture naming is often used for phonological assessment because of its simplicity and ease of administration. Compared to the assessment of conversational speech, a picture-naming task is more efficient and still provides a good index of phonological ability (Wolk & Meisler, 1998).

We used the 50 words of the Dutch revision of Hodson and Paden (1991) Assessment of Phonological Processes–Revised (Van de Wijer-Muris & Draaisma, 2000) that incorporates the full body of vowel, consonant, cluster, and syllable structure combinations of the Dutch language. The syllable shapes of the target words vary from simple to more complex. Because James, Ferguson, and Butcher (2016) suggest that multisyllabic words add value to picture-based speech testing, we decided to add 10 multisyllabic words with all phonemes occurring twice in different positions in different contexts. Comprising 40 one-syllable words, 13 two-syllable words, 6 three-syllable words, and 1 word with four syllables, our task assesses all Dutch phonemes in all possible syllable positions, except for /g/ because this consonant only occurs in loanwords in Dutch (see Appendix A). For the 4- to 7-year-olds, the words are presented in a random order, whereas for the 2- to 4-year-olds, the consonant–vowel–consonant (CVC) words are presented first, followed by the words with more complex syllable structures.

Both seated in front of a computer screen, the SLT asks the child to name what he or she sees on the color pictures that appear consecutively on the screen. Because it was crucial to elicit a sufficiently large speech sample, the computer reads out a sentence with a semantic cue when the child is unable to name the picture spontaneously. When this semantic cue does not elicit the target word, the computer reads out the target word and asks the child to repeat this out loud. It should be noted that, in the latter imitation condition, the lemma and word-form selection processes possibly play a different role than they do in the other two conditions.

Nonword imitation. Poor nonword imitation is widely used as a clinical marker of heritable forms of specific language impairment (Bishop, North, & Donlan, 1996). The capacity to imitate nonwords has been largely attributed to phonological memory (Gathercole, Willis, Baddeley, & Emslie, 1994), but other cognitive and linguistic processes are also involved (Smith, 2006), including speech production (Shriberg et al., 2009; Vance, Stackhouse, & Wells, 2005). We included nonword imitation in the CAI to investigate the underlying processes of phonological encoding and motor programming (Vance et al., 2005). As the child needs to create new motor programs when reproducing nonwords, this task can be used to isolate motor programming skills (Vance et al., 2005).

The task required the children to reproduce prerecorded nonwords, with accompanying color pictures of "nonsense figures" being shown on the computer screen to make the task more attractive, especially for the younger children. To ensure that the pictures of the nonsense figures did not influence a visual processing component in recalling nonwords, we used pictures of unfamiliar nonsense figures (see an example in Figure 2). Forty-seven of the nonwords were derived from the "Dyspraxia Program" (Eurlings-van Deurse et al., 1993), an assessment comparable to that of the Nuffield Dyspraxia Program, and 23 from Scheltinga (1998). We added 10 more nonwords whose syllable structures were based on the words we had added to the picture-naming task. The frequency distribution of phonological features is shown in Appendix A. The full task comprises 29 one-syllable nonwords, 35 nonwords with two syllables, and 16 nonwords with three syllables, with the 2- to 3-year-old children being presented with the full set of 80 items, whereas the older children needed to reproduce a subset of 33 bisyllabic and trisyllabic items. If a child failed to respond to an item, an additional live-voice presentation of the stimulus was given.

Figure 2. Two examples of the pictures of nonsense figures used for the nonword imitation task.

Word and nonword repetition. Speech variability has been associated with certain types of speech disorders, such as childhood apraxia of speech (CAS; Davis, Jakielski, & Marquardt, 1998; Dodd, 1995; Forrest, 2003; Holm, Crosbie, & Dodd, 2007; Iuzzini-Seigel, Hogan, & Green, 2017) and inconsistent phonological disorders (Dodd, 1995). It has also been documented in typically developing 2- and 3-year-olds (Sosa, 2015).

In this task, children are requested to repeat five prerecorded words and as many nonwords five times (without accompanying pictures). Only one model is provided. Both the word and nonword conditions contain 3 two-syllable and 2 three-syllable items with equal, complex consonant structures (CVC-CCVC, CCVC-CVC, CVCC-CCVCC, CVC-CV-CCV, CV-CV-CVC).

MRR. Also known as diadochokinesis, this is one of the most commonly used oral motor assessments in clinical practice. As it is a pure motor task and does not require any knowledge of words, syllables, or phonemes (Maassen & Terband, 2015), the MRR is exploited to differentiate types of SSDs (Lewis, Freebairn, Hansen, Iyengar, & Taylor, 2004; Murray, McCabe, Heard, & Ballard, 2015; Preston & Edwards, 2009; Rvachew, Hodge, & Ohberg, 2005; Shriberg et al., 2010; Thoonen, Maassen, Gabreels, & Schreuder, 1999; Thoonen, Maassen, Wit, Gabreëls, & Schreuder, 1996). MRR is especially useful in the differential diagnosis of children with CAS. CAS is a disorder of speech motor programming and planning (Nijland, 2003), and MRR or diadochokinesis is one of the most important quantitative measures that can differentiate CAS from other types of SSDs (Murray et al., 2015; Thoonen et al., 1996).

The MRR requires the child to produce three monosyllabic sequences (/pa/, /ta/, /ka/), two bisyllabic sequences (/pata/, /taka/), and one multisyllabic sequence (/pataka/) as fast and accurately as possible. We used a similar protocol to that developed by Thoonen et al. (1996). The MRR is calculated as the number of syllables produced per second during the child's fastest correct attempt.

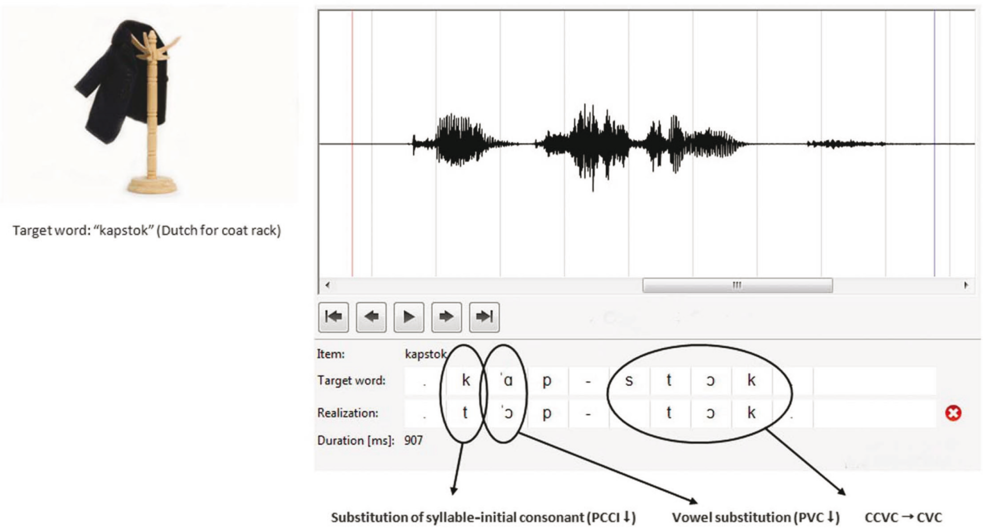

Figure 3. Example of the phonetic transcriptions of a target word and a recorded speech sample used in the Computer Articulation Instrument scoring procedure.

Scoring

The recordings of the children's speech productions were scored by the (student) SLTs that administered the test and took about 30–45 min, depending on the experience of the assessor and the number of speech errors of the child.

Picture naming and nonword imitation (phonetic transcription). Each utterance was transcribed using the Logical International Phonetics Programs (LIPP) software (Oller & Delgado, 2000), which allows for the transcription of International Phonetic Alphabet (IPA) via the traditional keyboard, along with user-designed analysis based on featural characterizations of segments. The assessors phonetically transcribed all speech recordings based on the correct target transcriptions by "editing in" the child's production errors. An example is given in Figure 3. To compare the child's performance with the targets, inventories of productions (or occurrences) of particular syllable structures, syllable-initial and syllable-final consonants, and vowel types as well as error counts were derived automatically based on a set of phonetic analysis rules, which are listed in Table 4. Percentages of correct productions were calculated by dividing the number of correctly produced phonemes or syllable structures by the total number of phonemes or syllable structures elicited in each task: PCC in syllable-initial position (PCCI), PVC, and percentage of correct syllable structure (CVC and consonant–consonant–vowel–consonant [CCVC], respectively). All syllable-initial consonants and all vowels were considered when calculating PCCI and PVC. For PCCI, the percentage of correctly produced consonants was divided by the total number of consonants. Because in our investigations the focus was on phonological and not on phonetic development, both common and uncommon clinical consonant distortions were scored as correct, similar to the PCC-Revised calculation described by Shriberg, Austin, Lewis, McSweeny, and Wilson (1997). The PVC was calculated by dividing the vowels pronounced correctly by the total number of vowels. Error counts of cluster reductions were described as the percentage of reduction of initial consonant clusters from two consonants to one divided by the total number of initial consonant clusters of two consonants (RedClus). In addition, we calculated "Level 4" and "Level 5." As described by Beers (1995), these parameters reflect percentages of the correct production of the two highest phonological complexity levels in typical Dutch phonological development, with Level 4 containing the phonemes /b/, /f/, and /ʊ/ and Level 5 containing the liquids /l/ and /R/, all in syllable-initial position. At least half of the typically developing children in the study by Beers reached Level 5 at the age of 2;6–2;8 (i.e., 75% correct responses).

Table 4. Parameters per speech task.

Task	Parameter	Description
PN	PCCI	Percentage of consonants correct in syllable-initial position
	PVC	Percentage of vowels correct
	Level 5	Percentage of correct consonants /l/ and /R/
	RedClus	Percentage of reduction of initial consonant clusters from two consonants to one
	CCVC	Percentage of correct syllable structure CCVC (C = consonant, V = vowel)
NWI	PCCI	Percentage of consonants correct in syllable-initial position
	PVC	Percentage of vowels correct
	Level 4	Percentage of correct consonants /b/, /f/, and /ʋ/
	Level 5	Percentage of correct consonants /l/ and /R/
	RedClus	Percentage of reduction of initial consonant clusters from two consonants to one
	CVC	Percentage of correct syllable structure CVC
	CCVC	Percentage of correct syllable structure CCVC
WR	PWV word	Proportion of whole-word variability: word repetition
NWR	PWV nonword	Proportion of whole-word variability: nonword repetition
MRR	MRR-pa	Number of syllables per second of sequence /pa/
	MRR-ta	Number of syllables per second of sequence /ta/
	MRR-ka	Number of syllables per second of sequence /ka/
	MRR-pataka	Number of syllables per second of sequence /pataka/
	MRR-pata	Number of syllables per second of sequence /pata/
	MRR-taka	Number of syllables per second of sequence /taka/

Note. PN = picture naming; NWI = nonword imitation; WR = word repetition; NWR = nonword repetition; MRR = maximum repetition rate.

Word and nonword repetition (variability). In the word trials, children had to repeat the following five true words five times: /kɑp-stɔk/, "kapstok," English: coat rack; /vIltstIft/, "viltstift," English: felt-tip pen; /vlix-tʉyx/, "vliegtuig," English: airplane; /pa-Ra-ply/, "paraplu," English: umbrella; and /te-lə-fon/, "telefoon," English: telephone. In the nonword trials, they had to repeat five items with similar structures (/tɛp-skIt/, "tepskit"; /xIlt-stɛxt/, "giltstecht"; /vlʉyx-tix/, "vluigtieg"; /po-Ro-pla/, "poorooplaa"; and /to-li-fan/, "tooliefaan"). For each (non)word, the number of different forms was determined. A production was identified as "different" when at least one of the phonemes of the target word was produced differently or deleted. For example, /airpane/ and /aiplane/ are two forms of the target word /airplane/. Consistency was established by dividing the total number of forms (with a maximum of 25) by the total number of productions (maximum of 25). A score of 1 indicates maximum variability. This parameter is a revised version of the proportion of whole-word variability (PWV), as described by Ingram (2002). Only the trials with at least three productions of a (non)word were used for analysis.

MRR. For each trial, the MRR was calculated as the number of syllables produced per second, resulting in six parameters: MRR-pa, MRR-ta, MRR-ka, MRR-pataka, MRR-pata, and MRR-taka. The fastest correctly produced syllable sequence was used for analysis. To determine the number of syllables and the duration of a trial, the sequence was displayed in a waveform using Praat (Boersma & Weenink, 2016). Only the trials with a minimum of five correct syllables were included in the analysis. Syllable boundaries were determined based on visual and auditory information. The burst of the voiceless plosives was used to localize the onset of a syllable. The first and last syllables were excluded from the analysis. Subsequently, the MRR was calculated by dividing the total number of syllables by the duration of the trial.

Procedure

CAI Administration

The children were tested individually in a quiet room in their own nursery or primary school. The SLT and child were seated side by side at a table on which a laptop computer was placed in a position comfortable for both. Both were wearing headsets, or a speaker and microphone were used. For standardization reasons, the tasks were presented in a fixed sequence: picture naming, word repetition, nonword imitation, nonword repetition, and MRR. Testing took approximately 30 min. Some children were seen twice because of their initial lack of interest or cooperation.

The tasks were administered in the younger age groups (2–4 years) by 14 SLTs and in the older children (4–7 years old) by 110 student SLTs (working in pairs) who were fulfilling their phonetics coursework in the third or fourth (final) year of their program. All were trained in the administration of the CAI by the first two authors, having received precise instructions and training in the scoring procedure (phonetic transcription, consistency evaluation, and MRR). Scoring was performed under the supervision of the first two authors and controlled for reliability. The scoring process was performed by the same SLT or SLT student who administered the test, under supervision of the first two authors. The assessors of the normative study were also used as raters for the reliability study.

Not all children completed all tasks for reasons of shyness or inattentiveness, among other causes. In Table 5, the number of children who completed a task is presented per age group. Incomplete tasks were excluded from the data set. The records of picture naming and nonword imitation were considered incomplete if the number of segments was less than 2 *SD*s below the mean number of segments for the age group. The data for word and nonword repetition were analyzed when a child had produced at least three words or nonwords per trial. For MRR, at least two of the three monosyllabic sequences needed to be correct. Table 5 shows that, from the age of 3;0, more than 60% of the children reached this criterion. Because of the high number of 2- and 3-year-olds not being able to perform the monosyllabic sequences, it was decided to set the lower age boundary for this task at 3;0. Thus, the MRR was calculated based on the data obtained in the children aged 3;0 onwards. After excluding the children who fell more than 2 SDs below the mean for picture naming and nonword imitation, the percentage of children with a history of speech-language difficulties ranged from 1.9% (MRR) to 2.1% (picture naming), similar to the percentage of the whole sample (2.2%).

Table 5. Number of children who completed a task, per task and age group.

Age group (years;months)	Normative sample	PN	NWI	WR	NWR	MRR	
	n	*n*	*n*	*n*	*n*	*n*	Pass (% *n* MRR)
2;0–2;3	72	72	65	42	43	59	18 (30.5)
2;4–2;7	102	101	90	56	65	83	26 (31.3)
2;8–2;11	101	101	96	69	73	88	55 (62.5)
3;0–3;3	104	102	97	70	80	93	68 (73.1)
3;4–3;7	110	107	105	80	87	99	65 (65.7)
3;8–3;11	102	101	98	81	88	97	86 (88.7)
4;0–4;3	100	99	99	85	84	88	77 (87.5)
4;4–4;7	115	111	113	102	103	96	90 (93.8)
4;8–4;11	116	112	113	97	97	94	93 (98.9)
5;0–5;3	121	117	117	101	103	107	103 (96.3)
5;4–5;7	128	128	128	112	116	115	111 (96.5)
5;8–5;11	117	116	116	106	108	105	104 (99.0)
6;0–6;5	117	117	117	102	105	108	108 (100)
6;6–6;11	119	119	118	108	109	110	109 (99.1)
Total	1,524	1,503	1,472	1,211	1,261	1,342	1,113 (82.9)
% of sample	100	98.6	96.6	79.5	82.7	88.1	73.0

Note. PN = picture naming; NWI = nonword imitation; WR = word repetition; NWR = nonword repetition; MRR = maximum repetition rate; Pass = correct production of at least two of the three monosyllabic sequences.

Reliability and Validity Procedure

Interrater reliability and test–retest reliability were determined based on the data sets of subgroups of the total sample. Initially, it was our goal to use 10% of the normative data for these reliability evaluations, compared to other normative studies (Clausen & Fox-Boyer, 2017; Dodd, Holm, Hua, & Crosbie, 2003; Gangji, Pascoe, & Smouse, 2015). However, because of its big volume, we were only able to do so for 4.72%–7.02% of the data. We used a sample size of 63–107 children per parameter (interrater reliability: 67–103 children, test–retest reliability: 63–107 children), thereby far surpassing the sample sizes used in the studies mentioned and complying to the recommended minimum sample size of 50 (De Vet et al., 2011). Giraudeau and Mary (2001) conducted simulation studies, showing that a 95% confidence interval of ± 0.1 is reached with a sample size of 50–100, if the intraclass correlation coefficient (ICC) has a value between .7 and .8. Note that the number of children is not equal for all parameters (see Tables B1–B3 in Appendix B) because not all children completed all test items.

Interrater reliability. Each audio recording was scored independently by two raters: Besides an assessor/rater who had also been involved in the data collection of the normative study, we had an additional, independent rater score all the data selected for the reliability analyses. This rater received the same training as the others but had not taken part in the normative study. Interrater reliability was calculated by comparing the scores of the two raters.

- Picture naming and nonword imitation: The audio recordings of 99 children were randomly selected (6.50% of the full sample with all age ranges being included) and transcribed by 35 (picture naming) and 34 (nonword imitation) raters and compared to the 99 transcriptions of the independent rater.
- Word repetition and nonword repetition: The audio recordings of 72 children were randomly selected (4.72% of the full sample with all age ranges being included) and scored by one rater, whose scores were compared to those of the independent rater.
- MRR: The audio recordings of 103 children were randomly selected (6.76% of the full sample with all age ranges being included) and scored by 33 raters. Their scores were compared to those of the independent rater.

Test–retest reliability. A total of 107 children randomly selected from one nursery and five elementary schools were tested twice (7.02% of the full sample). The subsample included children from all age ranges and geographic regions. The data of 11 of these children were also used in the interrater reliability analysis. To avoid learning effects and effects resulting from natural speech development, Kirk and Vigeland (2014) recommend 1–3 weeks as the preferred time interval between the two tests. In our study, the average interval between the initial test (T1) and the retest (T2) was 3.4 weeks, with a range of 1–13 weeks and a median of 3 weeks, comparable with other studies (Abou-Elsaad, Baz, & El-Banna, 2009; Kirk & Vigeland, 2014; Tresoldi et al., 2015), with 89.4% of the children being retested within 1–5 weeks. All children were in the same age group during the first and second administrations. Both tests were administered by the same assessor, and all four tasks were repeated. Two raters scored the randomized T1 and T2 audio recordings, with the same rater scoring the tests of the same child.

Construct validity. Age trends for all the extracted parameters mentioned in Table 4 were determined, and means per age group were calculated. Next, continuous normalized standard scores were computed based on the model developed by Tellegen and Laros (1993), in which the cumulative proportions of the raw scores across the age groups were simultaneously fitted as a higher-order function of raw score and age. With the exception of MRR, this model was applied separately to the six age groups including those aged 2;0 to 3;11 and the eight groups with children aged 4;0–6;11 for the three speech tasks, mainly because the parameters were not identical for the younger and older age groups. For MRR, the model was applied in one run to the 11 age groups starting from the age of 3;0.

We conducted two factor analyses to determine the component structure of the selected parameters across tasks and the factor scores per task, in which latter scores were obtained such that a test result could also be obtained in cases in which not all tasks had been completed.

Statistical Analyses

Interrater reliability and test–retest reliability were studied by estimating ICCs. For interrater reliability, two-way random-effects models were fitted with the parameter of interest as the dependent variable and the independent variables as random intercepts to allow for different levels per child and rater. No fixed effects were included in the model. For every parameter, the ICC was estimated, and the corresponding 95% confidence intervals were constructed with the bootstrap method. For the test–retest reliability estimation, three-way random-effects models were fitted; a random intercept for time point (T1 or T2) was also included. Again, ICCs were estimated, and bootstrap confidence intervals were constructed.

The Dutch Committee on Tests and Testing (COTAN) considers reliability coefficients below .70 as insufficient, those between .70 and .80 as sufficient, and estimates higher than .80 as good (Evers, Lucassen, Meijer, & Sijtsma, 2010). Besides ICCs, the interrater reliability of the phonetic transcriptions (picture naming and nonword imitation) was examined by calculating point-to-point agreement, with a mean percentage of agreement being reported. For test–retest reliability, Wilcoxon signed-ranks tests were performed to compare the T1 and T2 means.

Means and standard deviations of all parameters were calculated per age group to describe age trends. In addition, error bar graphs were plotted. Parameters were selected based on clinical relevance and monotonous age trends; for these parameters, percentile scores were calculated with the Tellegen and Laros (1993) regression formula containing the raw parameter score, its square, its cube, age, age squared, age cubed, all interaction factors except for the interaction of both cubes, and score cubed with age squared. To see how much variance was explained by the differences between age groups, R-squared values were calculated.

To determine the factor structure of the parameters across tasks and per task, a principal component analysis (PCA) with varimax rotation was conducted on all CAI parameters to identify clusters of items. The Kaiser–Meyer–Olkin (KMO) measure was performed prior to the PCA. Values greater than 0.5 were considered acceptable (Field, 2009). In the PCA, the criterion for eigenvalues to be greater than 1 was used. A factor loading of an absolute value of more than .4 is considered important (Field, 2009). The relationship between tasks was examined with Pearson product–moment correlations.

SPSS Version 20 for Windows (SPSS Inc.) and R Version 3.2.3 were used for all statistical analyses. The estimated ICCs and their bootstrap confidence intervals were computed in RStudio.

Results

Interrater Reliability

Table B1 (see Appendix B) shows the results of the interrater reliability evaluation. For picture naming, the interrater reliability was good for the parameters PCCI, Level 5, RedClus, and CCVC, with ICCs ranging from .80 to .90. The ICC (.59) for PVC was insufficient.

With an ICC of .82, the interrater reliability of the nonword-imitation task was good for RedClus and sufficient for the parameters PCCI, Level 4, Level 5, CVC, and CCVC, with ICCs ranging from .71 to .77. Interrater reliability (ICC of .62) was insufficient for PVC. With .73, the ICC for word repetition (PWV word) was sufficient but insufficient for nonword repetition (PWV nonword: .25).

Interrater reliability was good for the monosyllabic sequences MRR-pa and MRR-ka (ICCs of .81 and .83, respectively) and sufficient for MRR-ta (ICC of .77). With ICCs ranging from .41 to .62, interrater reliability estimates for MRR-pataka, MRR-pata, and MRR-taka were insufficient.

Table B2 shows the results of the point-to-point interrater agreement of the phonetic transcriptions of the picture naming and nonword imitation responses. The agreement of segments (total number of consonants, vowels, and word and syllable boundaries), consonants, and vowels was high for both tasks.

Test–Retest Reliability

Table B3 shows the results of the test–retest reliability analysis. Picture naming showed sufficient reliability for the parameters Level 5, RedCLus, and CCVC (ICCs of .71, .75, and .76, respectively), but this was insufficient for PCCI and PVC (ICCs of .51 and .31, respectively). The nonword-imitation task showed a good test–retest reliability for CVC (ICC of .88) and sufficient estimates for PCCI (ICC of .74), PVC (ICC of .77), and Level 5 (ICC of .73). The other parameters scored insufficient, with ICCs ranging from .41 to .60.

Insufficient test–retest values were found for word and nonword repetition (ICCs of .66 and .39, respectively) as well as for MRR, where ICCs ranged from .18 to .60, except for MRR-pa with a sufficient reliability score (ICC of .70). As Table B3 also shows, in picture naming, the T2 scores for PCCI, PVC, and Level 5 were significantly higher than the T1 scores, as was the case for the nonword-imitation parameters PCCI, PVC, Level 4, Level 5, CVC, and MRR-pataka.

Validity

Age Trends and Continuous Norms

The means and standard deviations of the picture-naming and nonword-imitation parameters are presented in Table B4 for the younger age group (2;0–3;11) and in Table B5 for the older age group (4;0–6;11). Monotonous increases with age were found for all parameters, except for the age range between 4;4 (52 months) and 5;7 (67 months), during which plateaus or only minor increases were observed. Apparently, during this stage, little development takes place.

The means and standard deviations of the parameters of word repetition and nonword repetition are presented in Tables B6 (younger age group) and B7 (older age group). Minor decreases with age were found for the PWV in both tasks. Across age groups, the children were more consistent in producing words than nonwords.

As mentioned in the Method section, the MRR task was completed by children from the age of 3;0 onwards. The mean percentage of children who could produce at least two monosyllabic repetitions was 75.8% at the age of 3;0–3;11, 93.5% at the age of 4;0–4;11, 97.2% at the age of 5;0–5;11, and 99.5% at the age of 6;0–6;11. Table B8 shows steady increases in the rates for monosyllabic, bisyllabic, and trisyllabic sequences with increasing age.

Appendix C shows examples of the modeling results for the main parameters. Criteria to accept the regression model are (a) an increase of percentile scores with increasing raw scores given a particular age and (b) a decrease of percentile scores with increasing age given a particular raw score. In the graph, lines per age group must thus show monotonous increases in percentile scores (*y*-axis) with increasing raw scores (*x*-axis) and should not cross. Because at the tails of the distribution these conditions are not always met, minimum and maximum values were determined for the raw scores such that the model is adequate within these limits. This implies that, within the first to fifth percentile range and within the 95th–100th percentile range, fine discrimination is lost, in which loss is considered clinically acceptable. R-squared values per parameter were all .96 or higher, indicating that the model-predicted percentile score based on raw score and age explained 96% of the variance in the raw scores, which can be considered an excellent match.

Factor Analysis

A PCA with orthogonal rotation (varimax) was conducted on the parameters of all four tasks (see Table 6). The KMO measure verified the sampling adequacy for the analysis (KMO = 0.78), and all Cronbach's alphas were .80 or higher, except for PWV

(.45). The five-component solution explained 60.3% of the variance. All picture-naming parameters are reflected by the third component. The first component mainly reflects the segmental parameters of nonword imitation (NWI-seg); and the fourth component, the nonword imitation cluster-reduction and syllable-structure parameters. The MRR trials form an independent component (Component 2), with loadings on the other components smaller than .1. Adding the PWV did not change the structure of these four components; the PWV measures form an additional component (Component 5) and do not load on the other components. Factor analyses without the MRR parameters and without the PWV measures resulted in an identical structure of picture-naming and nonword-imitation components.

Because in the complete factor analysis there were relatively many missing values, and to obtain composite performance scores, a second series of factor analyses for each task separately was conducted with the same three components being found for picture naming and nonword imitation, explaining 63.9% of the variance. Word and nonword repetition (PWV) and monosyllabic and multisyllabic MRRs both yielded a one-component solution, explaining 64.4% and 46.9% of the variance, respectively. The lower percentage for MRR is partly due to missing values in the multisyllabic sequences, which had to be estimated; in the factor analysis of the three monosyllabic sequences only, 68.4% of the variance was explained.

In order to test any remaining influence of gender and demographic variables on the factor scores, separate analyses of variance were conducted with each of the five factors as the dependent variables and gender (girl, boy), age group (14 categories), and covariate SES (standardized value) as the independent variables. Significant gender effects were found for the factors picture naming, $F(1, 1321) = 8.16$, $p = .004$; MRR, $F(1, 989) = 12.3$, $p < .001$; and consistency of repetition (PWV), $F(1, 1155) = 5.02$, $p = .025$, with the girls performing slightly better on the picture naming and consistency of repetition (PWV) factors and the boys performing slightly better on the MRR. The factor SES was significant for picture naming, $F(1, 1321) = 31.8$, $p < .001$, and consistency of repetition (PWV), $F(1, 1155) = 9.84$, $p = .002$.

Correlation Coefficients

Pearson correlations for the different tasks are shown in Table 7. Weak but significant correlations were found between picture naming and nonword imitation (NWI-seg and syllabic parameters [NWI-syl]) and between picture naming and consistency of repetition (PWV), as well as between nonword imitation (NWI-seg and NWIsyl) and consistency of repetition (PWV). No significant correlations were found between NWI-seg and NWI-syl or between MRR and the other tasks.

Table 6. Summary of the factor analysis results for picture naming (PN), nonword imitation (NWI), word (WR) and nonword (NWR) repetition, and maximum repetition rate (MRR).

		Rotated factor loadings				
		Component				
Task	**Parameter**	**1**	**2**	**3**	**4**	**5**
PN	PCCI	.24	.019	**.75**	-.014	.15
	PVC	.21	-.017	**.67**	-.053	.031
	RedClus	.040	.025	**.77**	.25	-.060
	Level 5	.22	.044	**.72**	-.016	.21
	CCVC	.026	.014	**.80**	.20	-.090
NWI	PCCI	**.87**	.031	.17	.12	.16
	PVC	**.75**	.038	.084	.091	-.011
	RedClus	.21	.023	.15	**.85**	.092
	Level 4	**.76**	.033	.086	.097	.14
	Level 5	**.67**	-.001	.26	.026	.17
	CVC	**.67**	.008	.14	.17	-.17
	CCVC	.21	.014	.098	**.86**	.081
WR	PWV word	.014	.059	.096	.11	**.76**
NWR	PWV nonword	.15	-.051	.033	.028	**.77**
MRR	MRR-pa	.035	**.73**	.002	-.021	-.056
	MRR-ta	-.015	**.77**	.034	.075	.016
	MRR-ka	-.004	**.74**	.024	.046	.010
	MRR-pataka	.076	**.59**	-.004	-.063	.025
	MRR-pata	-.025	**.71**	.026	-.047	-.017
	MRR-taka	.023	**.69**	-.006	.075	.036
Eigenvalues		4.62	2.99	1.89	1.33	1.23
% of variance		23.1	14.9	9.45	6.66	6.16
Cronbach's α		.83	.80	.82	.81	.45

Extraction method: principal component analysis
Rotation method: varimax with Kaiser normalization
Rotation converged in five iterations

Note. PCCI = percentage of consonants correct in syllable-initial position; PVC = percentage of vowels correct; RedClus = percentage of reduction of initial consonant clusters from two consonants to one; Level 5 = percentage of correct consonants /l/ and /R/; CCVC = percentage of correct syllable structure CCVC; Level 4 = percentage of correct consonants /b/, /f/, and /ʋ/; CVC = percentage of correct syllable structure CVC; PWV word = proportion of whole-word variability: word repetition; PWV nonword = proportion of whole-word variability: nonword repetition; MRR-pa = number of syllables per second of sequence /pa/; MRR-ta = number of syllables per second of sequence /ta/; MRR-ka = number of syllables per second of sequence /ka/; MRRpataka = number of syllables per second of sequence /pataka/; MRR-pata = number of syllables per second of sequence /pata/; MRR-taka = number of syllables per second of sequence /taka/.

Discussion

In this study, we report on the psychometric properties of the CAI, a test battery assessing the development of speech production skills, based on the performance of 1,524 children aged between 2 and 7 years, making it the first norm-referenced standardized test designed for process-oriented diagnostics of spoken Dutch. Based on a selection of the scores on its four constituent tasks, namely, picture naming, nonword imitation, word and nonword repetition, and MRR, we examined the interrater and test–retest reliability of the CAI and its construct validity (factor

Table 7. Pearson correlations between tasks.

Task	Correlation	PN	NWI-seg	NWI-syl	PWV	MRR
PN	Pearson correlation	1	.34**	.22**	.15**	.049
	n	1,466	1,351	1,351	1,168	991
NWI-seg	Pearson correlation	.34**	1	-.001	.19**	.045
	n	1,351	1,373	1,373	1,180	980
NWI-syl	Pearson correlation	.22**	-.001	1	.12**	.039
	n	1,351	1,373	1,373	1,180	980
PWV	Pearson correlation	.15**	.19**	.12**	1	.018
	n	1,168	1,180	1,180	1,184	937
MRR	Pearson correlation	.049	.045	.039	.018	1
	n	991	980	980	937	1,012

Note. PN = picture naming; NWI-seg = segmental parameters of nonword imitation; NWI-syl = syllabic parameters of nonword imitation; PWV = proportion of whole-word variability; MRR = maximum repetition rate.

**Correlation of factor scores is significant at the .01 level (two-tailed).

analyses and correlation coefficients) across age groups, with norm values being established for each parameter separately.

Interrater Reliability of the CAI Tasks

The overall findings of the reliability study (interrater and test–retest) were sufficient to good. The majority of the parameters met an ICC minimum of .70.

Phonetic transcription of children's speech using the IPA is widely used, although the reliability of the method is questioned, especially the transcriptions of the speech of young children, and different agreement percentages are described (Preston & Koenig, 2011; Shriberg & Lof, 1991). This variation is due to different factors, such as the experience of the transcriber and environmental conditions. In many speech assessments, especially in those of nonword-imitation performance, rather than phonetic transcription, a correct–incorrect score is used because, in such dichotomous ratings, the interference of variation or measurement errors is less pronounced. We hence were expecting some variation in our phonetic transcriptions and hence the ICCs. Still, the interrater reliability of most of the picture-naming and nonword-imitation parameters was sufficient to good, with the percentages for point-to-point agreement being high (above 80%) for all measures. We think that the high reliability values we obtained are the result of the use of speech recordings and a broad phonetic transcription, with target transcriptions being provided and access to the acoustic signal for verification purposes being easy (Shriberg & Lof, 1991).

The interrater reliability of the PVC was insufficient for both the picture-naming and nonword-imitation tasks. Arguably, the transcription of vowels is more challenging than the transcription of consonants, which may have two causes: (a) There is greater dialectal variation in Dutch vowels than there is in Dutch consonants, and (b) the categorization of vowels is in general more difficult than that of consonants, although phoneme boundaries are less clearly defined (Howard & Heselwood, 2012; Pollock & Berni, 2001). Lower reliability scores for vowel transcriptions have been reported in other studies (Dodd et al., 2003), but others describe no difference in consonant and vowel agreement (Shriberg & Lof, 1991). Noteworthy here is that these latter studies describe interrater reliability in terms of point-to-point agreement only, which was also good for vowels in our evaluation. Whereas point-to-point agreement only reflects agreement, ICCs reflect both agreement and correlation (Shrout & Fleiss, 1979). The calculation of ICC involves dividing the between-subjects variability (BV) by the total variability (similar to analysis of variance). The total variability can be modeled as consisting of BV plus the within-subject variability or—in case of a reliability study—error variance (EV). In formula: ICC = BV / [BV + EV]. This implies that, given a particular value for EV (due to interrater variability or test–retest variability), the ICC is lower if BV is relatively small, as was the case in our normalization data. The PVCs in the typically developing children we tested were high, with little variation (small standard deviation). If the outcomes of the sample show large variability, which we expect to occur when also children with speech disorders are tested, it is easier to distinguish the children from each other despite any measurement errors than when the outcomes differ very little. A next step in the psychometric evaluation of the CAI will be to add children with speech disorders in the reliability study. Because of the important role of PVC in distinguishing typical speech development from SSDs (Forrest, 2003; Iuzzini-Seigel & Murray, 2017), the PVC remains included as a parameter of picture naming and nonword imitation in the CAI.

The ICC for PWV in the word repetition task was sufficient but insufficient for nonword repetition, which may be due to the fact that it is more difficult for raters to judge unfamiliar phonological items than known words. Because the nonword reproductions were not transcribed, the raters had to rely on auditory information only. Here, phonological short-term memory plays a large role, placing higher demands on the rater's working memory capacities. Because of the unfamiliarity of nonwords, it is hypothesized that raters are inclined to listen for more detailed information similarly to what happens with narrow phonetic transcriptions, whereas this manner of transcription proved unreliable (Shriberg & Lof, 1991). In the future, raters are accordingly advised to qualify the differences between target and reproduced nonwords as they do based on broad transcriptions without paying attention to small diacritic differences.

Whereas the few studies that have been conducted previously (Gadesmann & Miller, 2008) describe poor interrater reliability for this task, in our study, the scoring of the monosyllabic items of the MRR was reliable. In most studies, MRR is typically measured by counting syllables and reading the time with a stopwatch. Kent, Kent, and Rosenbek (1987) suggested that interrater reliability would increase if some method of instrumentation were to be used that displayed acoustic waveforms, while standardized instructions and procedures would help reduce variability within and across children. In our study, we indeed used a standardized measurement protocol and an acoustic waveform. The raters judged the speech samples in three steps, supported by a display of the acoustic signal, determining (a) the onset of the second syllable and (b) the onset of the final syllable and then (c) counting the number of syllables, with the duration of a sequence calculated automatically. This procedure potentially explains the high reliability in the scoring of monosyllabic MRRs. In contrast, the interrater reliability for the bisyllabic and trisyllabic items was insufficient. Several factors might underlie this result. First, we noticed that the younger children had difficulties performing the MRR task, likely because they had difficulties understanding the instructions, whereas a large number of children were not able to perform the task at all. Conversely, there was high agreement in the raters' judgments whether the attempts were successful or not. The data of the children who failed to perform the task were not included in the reliability study. If we had, this would have resulted in higher ICCs. Another influencing factor might be that judging whether the sequences of the bisyllabic and trisyllabic items were produced correctly is more difficult than it is for the monosyllabic items because the younger children made more age-specific errors of pronunciation, as Yaruss and Logan (2002) also noted.

Because we used many different raters for our interrater reliability evaluation, our study may be a good reflection of the professional field, with the ICCs we obtained being representative of clinical practice. On the other hand, it has caused more variability in the ratings and lowered the ICCs.

Test–Retest Reliability of the CAI Tasks

Except for the PCCI and PVC, test–retest reliability was sufficient for the other parameters of picture naming, which means that performance on these parameters was stable over time. Because of the important role of PCCI and PVC in the diagnosis of SSDs, as discussed above for interrater reliability, PCCI and PVC remain included as parameters of picture naming.

Comparable to other studies (Gadesmann & Miller, 2008; Gray, 2003), test–retest reliability proved insufficient for most of the parameters of the other three tasks (nonword imitation, word and nonword repetition, and MRR), with the scores having improved significantly the second time round. Four factors seem to be involved. First, as reflected by the normative data, the children's speech variables changed rapidly over time, becoming more accurate especially in the younger age groups (Dodd et al., 2003; McIntosh & Dodd, 2008). The factor "normal speech development" plays an unexpected major role here and impacted the level of the reliability coefficients directly, even within the relatively short space of 3 weeks (mean time interval between T1 and T2: 3.4 weeks; range: 1–13 weeks), while the increase in the raw scores between T1 and T2 of the five children who repeated the tasks 13 weeks after the first administration did not differ from that observed in the other children.

Second, besides a developmental effect, the effect of learning might be more pronounced in tasks assessing imitation and repetition of words and nonwords than it is in picture-naming paradigms. At the first assessment, most of the children will have had no experience with these kinds of tasks, which thus test new skills. During the retest, the children are more familiar with the procedure, which may have boosted their performance.

Third, the tasks were administered by the same assessor at both time points, and especially, the younger children might have felt more at ease with the assessor during the retest, which may have positively influenced their T2 scores. A fourth factor that might have negatively affected the test–retest reliability in our study is that, in most previous studies, nonword imitation was assessed using whole-word scoring procedures, which tend to yield higher reliability results. Because the purpose of our study was to investigate speech production, we chose to use phoneme-by-phoneme scoring (phonetic transcription), which has the disadvantage that scoring will show more variation among raters. Besides the variability likely attributable to developmental and learning effects, also attention and motivation differences among tests may underlie (part of) the variability. The sample size of both reliability studies (interrater and test–retest reliability) was less than 10% of the normative data, and this is a weakness of this study. Our goal was to use 10% of the data, as was recommended in the literature (Clausen & Fox-Boyer, 2017; Dodd et al., 2003; Gangji et al., 2015). However, we used a sample size of 63–107 children per parameter, thereby far surpassing the sample sizes used in the studies mentioned and complying to the recommended minimum sample size of 50 (De Vet et al., 2011; Giraudeau & Mary, 2001).

Another limitation of the study is the presence of children with a history of speech and language difficulties and children with another primary language than Dutch, in the normative sample. Children with a speech-language delay or different linguistic background are less reliable to transcribe (Shriberg & Lof, 1991). However, it is important to include these groups of children in order to achieve a representative sample, as stated by Dodd et al. (2003).

Validity of the CAI

The content validity of the CAI was demonstrated by the description of the constituent tasks and their items. The distribution of the consonants, vowels, clusters, and syllable structures included is representative of the Dutch language. The items of the four tasks each measure different aspects of speech production. Because of the lack of another comprehensive, norm-referenced speech production assessment instrument in Dutch, we were not able to establish the criterion validity of our test battery.

The increase we reported in the mean scores of almost all parameters of the CAI with chronological age supports its construct validity: Speech production abilities improved with increasing age, as was reported in other studies (Abou-Elsaad et al., 2009; Dodd et al., 2003; Lousada, Mendes, Valente, & Hall, 2012; McIntosh & Dodd, 2008; Priester, Post, & Goorhuis-Brouwer, 2011; Tresoldi et al., 2015; Vance et al., 2005). As Tables B4 and B5 show, the largest increase occurred during the preschool years (i.e., in the 2- to 4-year-olds). Because speech development is generally assumed to be completed by the age of 5 years (Dodd, 2011), lower variation is expected when children grow older. Most parameters of picture naming and nonword imitation indeed showed a decrease in standard deviations with increasing age, with the higher standard deviations recorded for the younger age groups, where intersubject variation is typical (Dodd, 1995; Stoel-Gammon & Cooper, 1984). In contrast, standard deviations of the MRR parameters were quite stable for all age groups. Here, with the number of syllables per second increasing, variation did not decrease with age, which indicates that speech or phonological development, as gauged with these two tasks, progresses differently from speech motor development, with the processes representing two different aspects of speech development. We have no clear explanation for the relatively low score on nonword imitation PCCI in those aged 4;0–4;3 (M_{age} = 49 months), except that the children aged 4 years and over were offered more complex material. Because modeling was conducted for the children aged 2;0–3;11 and 4;0–6;11 separately, this decrease in scores should not have complicated the interpretation of the normative scores. Noteworthy is the PWV variability score for word and nonword repetition, which only showed minor decreases with age and little between-subjects variation (Cronbach's α = .448). We propose that, rather than gauging speech development, this component gauges speech pathology, where the normative data can help to differentiate different types of speech disorders.

Continuous normalized standard scores were calculated for all parameters, and all showed an increase of percentile scores with increasing raw scores for all ages. SLTs can thus use these normative data to discriminate between typically developing children and children with a speech delay or potential speech disorder.

Factor analyses revealed five meaningful factors, based on which as many constructs of speech production could be determined. One component represented all picture-naming parameters; a second component, the segmental parameters of nonword imitation; and a third, the nonword-imitation parameters

RedClus and CCVC, both measuring cluster reduction and syllable structure, with the PWV parameters forming an additional, fourth component, best characterized as "(non)word-repetition consistency," and the fifth component reflecting all MRR parameters. Factor analyses for each task separately confirmed the five factors.

Gender-related effects were found for picture naming, MRR, and consistency of repetition (PWV) but not for nonword imitation. The girls performed slightly better on picture naming and PWV, which is comparable to other studies that also reported gender differences in phonological acquisition, with most finding girls to outperform boys on speech accuracy measures (Dodd et al., 2003; Smit, Hand, Freilinger, Bernthal, & Bird, 1991). As to MRR, the boys performed better than the girls and especially so in the younger age groups, a finding also consistent with another study on speech motor performance, where the oral diadochokinetic rates of boys tended to be faster than those recorded for girls (Prathanee, Thanaviratananich, & Pongjanyakul, 2003).

CAI's construct validity is underlined by the correlation between picture naming and nonword segmental accuracy as well as between the picture-naming and nonword-imitation parameters and consistency of repetition (PWV). Monosyllabic and multisyllabic repetition was not correlated with any of the other measures, confirming that MRR is a distinct task.

Future Perspectives

In this study, CAI outcomes were scored based on manual phonetic transcription using LIPP (Oller & Delgado, 2000) and Praat for acoustic measurements (Boersma & Weenink, 2016). To make CAI more user-friendly, in the future version for use in clinical practice (Maassen et al., 2019), all scoring procedures will be automated, where, for instance, the LIPP set of phonetic analysis rules and acoustic waveforms for MRR scoring will have been integrated into the software.

Studies are currently being conducted to determine the role of the CAI in diagnosing speech disorders. We are investigating whether the instrument can differentiate children with typical speech from peers showing signs of an SSD (known-group validity) while allowing pathology profiles to be made for a differential diagnosis, where the subtest scores of children with a suspected deviant speech development are compared with the normative data. Further steps in the implementation of an automated, process-oriented diagnosis of abnormal speech development will include the addition of objective (acoustic) measures of speech production and a process analysis of the outcomes of our assessment battery on the basis of data collected in different speaking conditions (Maassen & Terband, 2015).

A limitation of this study is that the CAI is currently only available in Dutch. However, the CAI has an open structure in that all stimulus materials (spoken instructions, the pictures for the naming task, audio targets for the word and nonword repetition tasks, and audio targets for the MRR task) are stored in separate files. The phonetic transcriptions of the target items in IPA and the rules for analyzing transcribed utterances in relation to the targets are also stored in separate files. A strong asset of the CAI software is the interpreter, which allows for a comprehensive and versatile analysis of transcriptions. This allows the software to be adapted to test the speech development of children speaking other languages than Dutch. When translating the CAI into other languages, new reliability and validity studies should be carried out. In addition, when choosing new test items, the distribution of phonemes of the other language must be taken into account. International collaboration may then contribute to further evaluation and refinement of the instrument.

Conclusion

In this article, we reported the results of a normative study of CAI, a newly developed computer-based speech assessment instrument. The test battery incorporates a picture-naming task and word and nonword reproduction tasks, along with a task assessing MRRs. With these tasks in the CAI, different aspects of speech production can be evaluated and compared with each other in order to obtain a complete speech profile. The analyses of the phonological measurements, syllabic structures, and speech motor skills yielded indices of typical speech development in Dutch-speaking children in the ages between 2 and 7 years, based on which norm-referenced estimates of speech delay were determined.

Reliability and validity evaluations overall yielded sufficient to good values for interrater reliability. ICCs for test–retest reliability were low due to natural development and learning effects but good for construct validity, indicating that the CAI can be used to gauge typical and atypical speech development.

Acknowledgments

The reported research was partly funded by BOOM Test Publishers, Amsterdam, the Netherlands, and the Damsté-Terpstra Fonds, Zeist, the Netherlands. We thank the children and their parents, nurseries, primary schools, speech-language therapists, and speech-language therapy students who/that participated in this project.

References

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.

Andersson, L. (2005). Determining the adequacy of tests of children's language. *Communication Disorders Quarterly, 26*(4), 207–225.

Baarda, D., de Boer-Jongsma, N., & Jongsma, B. (2013). *LOGO-Art. Nederlands Articulatieonderzoek [LOGO-Art Dutch Articulation Assessment].* Dronten, the Netherlands: Uitgeverij LOGO-Start.

Beers, M. (1995). *The phonology of normally developing and language-impaired children.* IFOTT, University of Amsterdam, Amsterdam, the Netherlands.

Bishop, D. V. M., North, T., & Donlan, C. (1996). Nonword repetition as a behavioural marker for inherited language impairment: Evidence from a twin study. *The Journal of Child Psychology and Psychiatry, 37*(4), 391–403.

Boersma, P., & Weenink, D. (2016). *Praat: Doing phonetics by computer* [Computer program]. Version 6.0.21.

Clausen, M. C., & Fox-Boyer, A. (2017). Phonological development of Danish-speaking children: A normative cross-sectional study. *Clinical Linguistics & Phonetics, 31*(6), 440–458.

Davis, B. L., Jakielski, K. J., & Marquardt, T. P. (1998). Developmental apraxia of speech: Determiners of differential diagnosis. *Clinical Linguistics & Phonetics, 12*(1), 25–45.

De Vet, H. C., Terwee, C. B., Mokkink, L. B., & Knol, D. L. (2011). *Measurement in medicine: A practical guide.* New York, NY: Cambridge University Press.

Diepeveen, S., Van Haaften, L., Terband, H., De Swart, B., & Maassen, B. (submitted). *Clinical reasoning in the field of SSD: The what's and why's regarding diagnosis and intervention in SLTs' daily practice.*

Dodd, B. (1995). *Differential diagnosis and treatment of children with speech disorder.* London, United Kingdom: Wiley.

Dodd, B. (2011). Differentiating speech delay from disorder: Does it matter. *Topics in Language Disorders, 31*(2), 96–111.

Dodd, B., Holm, A., Hua, Z., & Crosbie, S. (2003). Phonological development: A normative study of British English-speaking children. *Clinical Linguistics & Phonetics, 17*(8), 617–643.

Eurlings-van Deurse, M., Freriks, A., Goudt-Bakker, K., Van der Meulen, S., & Vries, L. D. (1993). *Dyspraxieprogramma: Therapieprogramma voor kinderen met kenmerken van een verbale ontwikkelingsdyspraxie [Dyspraxia Program: A therapy program for children with childhood apraxia of speech].* Lisse, the Netherlands: Swets & Zeitlinger.

Evers, A., Lucassen, W., Meijer, R., & Sijtsma, K. (2010). *COTAN Beoordelingssysteem voor de kwaliteit van tests [COTAN Assessment system for the quality of tests].* Amsterdam, the Netherlands: NIP.

Field, A. (2009). *Discovering statistics using SPSS.* London, United Kingdom: Sage.

Flipsen, J. 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*, 166–178.

Forrest, K. (2003). Diagnostic criteria of developmental apraxia of speech used by clinical speech-language pathologists. *American Journal of Speech-Language Pathology, 12*(3), 376–380.

Gadesmann, M., & Miller, N. (2008). Reliability of speech diadochokinetic test measurement. *International Journal of Language & Communication Disorders, 43*(1), 41–54.

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 & Communication Disorders, 50*(2), 151–164.

Gathercole, S. E.,Willis, C. S., Baddeley, A. D., & Emslie,H. (1994). The children's test of nonword repetition: A test of phonological working memory. *Memory, 2*(2), 103–127.

Giraudeau, B., & Mary, J. (2001). Planning a reproducibility study: How many subjects and how many replicates per subject for an expected width of the 95 per cent confidence interval of the intraclass correlation coefficient. *Statistics in Medicine, 20*(21), 3205–3214.

Gray, S. (2003). Diagnostic accuracy and test–retest reliability of nonword repetition and digit span tasks administered to preschool children with specific language impairment. *Journal of Communication Disorders, 36*(2), 129–151.

Hayden, D. (2004). PROMPT: A tactually grounded treatment approach to speech production disorders. In I. Stockman (Ed.), *Movement and action in learning and development: Clinical implications for pervasive developmental disorders* (pp. 255–297). San Diego, CA: Elsevier-Academic Press.

Hodson, B. W., & Paden, E. P. (1991). *A phonological approach to remediation: Targeting intelligible speech.* Austin, TX: Pro-Ed.

Holm, A., Crosbie, S., & Dodd, B. (2007). Differentiating normal variability from inconsistency in children's speech: Normative data. *International Journal of Language & Communication Disorders, 42*(4), 467–486.

Howard, S., & Heselwood, B. (2012). The contribution of phonetics to the study of vowel development and disorders. In M. Ball, J. Damico, M. Ball, & F. Gibbon (Eds.), *Handbook of vowels and vowel disorders* (pp. 61–112). Abingdon, United Kingdom: Routledge.

Ingram, D. (2002). The measurement of whole-word productions. *Journal of Child Language, 29*(4), 713–733.

Iuzzini-Seigel, J., Hogan, T. P., & Green, J. R. (2017). Speech inconsistency in children with childhood apraxia of speech, language impairment, and speech delay: Depends on the stimuli. *Journal of Speech, Language, and Hearing Research, 60*(5), 1194–1210.

Iuzzini-Seigel, J., & Murray, E. (2017). Speech assessment in children with childhood apraxia of speech. *Perspectives of the ASHA Special Interest Groups, 2*(2), 47–60.

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.

Kent, R. D., Kent, J. F., & Rosenbek, J. C. (1987). Maximum performance tests of speech production. *Journal of Speech and Hearing Disorders, 52*(4), 367–387.

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, 45*(4), 365–377.

Knol, F., Boelhouwer, J., & Veldheer, V. (2012). *Statusontwikkeling van wijken in Nederland 1998–2010 [Neighbourhood status development in the Netherlands 1998–2010].* The Hague, the Netherlands: Institute for Social Research.

Leijdekker-Brinkman, W. (2002). Metaphonbox (Dutch version of the Metaphon Screening Assessment). Lisse, the Netherlands: Swets & Zeitlinger B.V.

Lewis, B. A., Freebairn, L. A., Hansen, A. J., Iyengar, S. K., & Taylor, H. G. (2004). School-age follow-up of children with childhood apraxia of speech. *Language, Speech, and Hearing Services in Schools, 35*(2), 122–140.

Lousada, M., Mendes, A. P., Valente, A. R., & Hall, A. (2012). Standardization of a phonetic–phonological test for European–Portuguese children. *Folia Phoniatrica et Logopaedica, 64*(3), 151–156.

Maassen, B., & Terband, H. (2015). Process-oriented diagnosis of childhood and adult apraxia of speech (CAS and AOS). In M. A. Redford (Ed.), *The handbook of speech production* (1st ed., pp. 331–350). Hoboken, NJ: Wiley.

Maassen, B., van Haaften, L., Diepeveen, S., Terband, H., van den Engel-Hoek, L., Veenker, T., & De Swart, B. (2019). *Computer Articulation Instrument.* Amsterdam, the Netherlands: Boom Uitgevers.

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.

McIntosh, B., & Dodd, B. J. (2008). Two-year-olds' phonological acquisition: Normative data. *International Journal of Speech-Language Pathology, 10*(6), 460–469.

McLeod, S., & Verdon, S. (2014). A review of 30 speech assessments in 19 languages other than English. *American Journal of Speech-Language Pathology, 23*(4), 1–16.

Mokkink, L. B., Terwee, C. B., Knol, D. L., et al. (2010). The COSMIN checklist for evaluating the methodological quality of studies on measurement properties: A clarification of its content. *BMC Medical Research Methodology, 10*(22), 1–8.

Mokkink, L. B., Terwee, C. B., Patrick, D. L., et al. (2010). The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *Journal of Clinical Epidemiology, 63*(7), 737–745.

Murray, E., McCabe, P., Heard, R., & Ballard, K. J. (2015). Differential diagnosis of children with suspected childhood apraxia of speech. *Journal of Speech, Language, and Hearing Research, 58*, 43–60.

Nijland, L. (2003). *Developmental apraxia of speech: Deficits in phonetic planning and motor programming.* Wageningen, the Netherlands: Ponson & Looijen.

Oller, D., & Delgado, R. (2000). Logical International Phonetics Program (Version 2.02) [Computer software]. Miami, FL: Intelligent Hearing Systems.

Pollock, K. E., & Berni, M. C. (2001). Transcription of vowels. *Topics in Language Disorders, 21*(4), 22–40.

Prathanee, B., Thanaviratananich, S., & Pongjanyakul, A. (2003). Oral diadochokinetic rates for normal Thai children. *International Journal of Language & Communication Disorders, 38*(4), 417–428.

Preston, J. L., & Edwards, M. L. (2009). Speed and accuracy of rapid speech output by adolescents with residual speech sound errors including rhotics. *Clinical Linguistics & Phonetics, 23*(4), 301–318.

Preston, J. L., & Koenig, L. L. (2011). Phonetic variability in residual speech sound disorders, exploration of subtypes. *Topics in Language Disorders, 31*(2), 168–184.

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.

Rvachew, S., Hodge, M., & Ohberg, A. (2005). Obtaining and interpreting maximum performance tasks from children: A tutorial. *Journal of Speech-Language Pathology and Audiology, 29*(4), 146–157.

Scheltinga, F. (1998). *Nonword repetition test.* Universiteit van Amsterdam & CED Groep Rotterdam, Rotterdam, Netherlands.

Shriberg, L. D., Austin, D., Lewis, B. A., McSweeny, J. L., & Wilson, D. L. (1997). The percentage of consonants correct (PCC): Metric extensions and reliability data. *Journal of Speech, Language, and Hearing Research, 40*(4), 708–722.

Shriberg, L. D., Fourakis, M., Hall, S. D., et al. (2010). Extensions to the speech disorders classification system (SDCS). *Clinical Linguistics & Phonetics, 24*(10), 795–824.

Shriberg, L. D., & Lof, G. L. (1991). Reliability studies in broad and narrow phonetic transcription. *Clinical Linguistics & Phonetics, 5*(3), 225–279.

Shriberg, L. D., Lohmeier, H. L., Campbell, T. F., Dollaghan, C. A., Green, J. R., & Moore, C. A. (2009). A nonword repetition task for speakers with misarticulations: The syllable repetition task (SRT). *Journal of Speech, Language, and Hearing Research, 52*(5), 1189–1212.

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin, 86*(2), 420–428.

Smit, A. B., Hand, L., Freilinger, J. J., Bernthal, J. E., & Bird, A. (1991). The Iowa articulation norms project and its Nebraska replication. *Journal of Speech and Hearing Research, 55*(4), 779–798.

Smith, B. (2006). Precautions regarding nonword repetition tasks. *Applied Psycholinguistics, 27*(4), 584–587.

Sosa, A. V. (2015). Intraword variability in typical speech development. *American Journal of Speech-Language Pathology, 24*(1), 24–35.

Stoel-Gammon, C., & Cooper, J. A. (1984). Patterns of early lexical and phonological development. *Journal of Child Language, 11*(2), 247–271.

Tellegen, P., & Laros, J. (1993). The construction and validation of a nonverbal test of intelligence: The revision of the Snijders-Oomen tests. European Journal of Psychological Assessment, 9(2), 147–157.

Thoonen, G. (1998). *Developmental apraxia of speech in children: Quantitative assessment of speech characteristics.* Enschede, the Netherlands: Print Partners Ipskamp.

Thoonen, G., Maassen, B., Gabreels, F., & Schreuder, R. (1999). Validity of maximum performance tasks to diagnose motor speech disorders in children. *Clinical Linguistics & Phonetics, 13*(1), 1–23.

Thoonen, G., Maassen, B., Wit, J., Gabreëls, F., & Schreuder, R. (1996). The integrated use of maximum performance tasks in differential diagnostic evaluations among children with motor speech disorders. *Clinical Linguistics & Phonetics, 10*(4), 311–336.

Tresoldi, M., Ambrogi, F., Favero, E., Colombo, A., Barillari, M. R., Velardi, P., & Schindler, A. (2015). Reliability, validity and normative data of a quick repetition test for Italian children. *International Journal of Pediatric Otorhinolaryngology, 79*(6), 888–894.

Van de Wijer-Muris, I., & Draaisma, Y. (2000). *Cursusreader 'Fonologische stoornissen,' version October 2000 (Dutch version of the Hodson Assessment of Phonological Patterns).* Unpublished document.

Vance, M., Stackhouse, J., & Wells, B. (2005). Speech-production skills in children aged 3–7 years. *International Journal of Language & Communication Disorders, 40*(1), 29–48.

Verhoeven, L., & Vermeer, A. (2001). *Taaltoets alle kinderen (Dutch language proficiency test for all children).* Arnhem, the Netherlands: Citogroep.

Verhoeven, L., & Vermeer, A. (2006). *Verantwoording taaltoets alle kinderen (Dutch language proficiency test for all children).* Arnhem, the Netherlands: Citogroep.

Wolk, L., & Meisler, A. W. (1998). Phonological assessment: A systematic comparison of conversation and picture naming. *Journal of Communication Disorders, 31*(4), 291–313.

Yaruss, S. J., & Logan, K. J. (2002). Evaluating rate, accuracy, and fluency of young children's diadochokinetic productions: A preliminary investigation. *Journal of Fluency Disorders, 27*(1), 65–86.

Appendix A

Frequency distributions of the phonological features of picture naming and nonword imitation.

		Number of syllable-initial features		
		Picture naming	Nonword imitation	
Class	Feature		2–3 years	4–7 years
Consonants	p	4	9	7
	b	2	5	3
	t	9	5	3
	d	3	7	5
	k	3	6	4
	g	0	0	0
	ŋ	0	0	0
	m	3	9	7
	n	2	8	6
	l	6	9	7
	R	4	6	4
	f	6	14	10
	v	3	6	4
	s	5	12	8
	z	3	3	1
	ʃ	2	1	1
	ʒ	1	0	0
	j	3	3	1
	x	3	7	5
	h	2	4	2
	ʋ	4	4	2
Vowels	i	8	22	3
	y	2	4	1
	e	4	8	0
	ø	4	0	0
	a	7	14	5
	o	5	15	5
	u	4	7	0
	I	9	16	16
	ɛ	3	20	16
	ɑ	11	16	11
	ʉ	2	11	11
	ə	12	1	1
	ɔ	9	11	11
Diphthongs	ɛi	5	0	0
	ɑu	2	1	1
	ʉy	3	1	1
Syllable structures	V	3	16	1
	CV	17	42	7
	VC	0	15	0
	CVC	40	66	66
	CCV	3	1	1
	CVCC	6	2	2
	CCVC	15	3	3
	CCVCC	3	2	2
	CCCVC	3	0	0
	Initial consonant clusters	23	6	6

Appendix B

Data of the reliability and validity studies

Table B1. Intraclass correlation coefficients for the interrater reliability per task.

Task	Parameter	*n*	ICC	CI lb	CI ub
PN	PCCI	99	.80	.74	.88
	PVC	99	.59	.45	.72
	Level 5	99	.84	.80	.91
	RedClus	99	.90	.87	.95
	CCVC	99	.85	.80	.92
NWI	PCCI	98	.77	.70	.87
	PVC	98	.62	.50	.75
	Level 4	98	.75	.68	.85
	Level 5	98	.76	.69	.86
	RedClus	96	.82	.76	.90
	CVC	96	.76	.68	.85
	CCVC	94	.71	.62	.82
WR	PWV word	72	.73	.65	.85
NWR	PWV nonword	67	.25	.05	.47
MRR	MRR-pa	100	.81	.75	.89
	MRR-ta	95	.77	.70	.87
	MRR-ka	103	.83	.77	.90
	MRR-pataka	67	.41	.23	.62
	MRR-pata	90	.62	.51	.77
	MRR-taka	91	.52	.38	.67

Note. ICC = intraclass correlation coefficient; CI lb= confidence interval, lower bound; CI ub= confidence interval, upper bound; PN = picture naming; PCCI = percentage of consonants correct in syllable-initial position; PVC = percentage of vowels correct; Level 5 = percentage of correct consonants /l/ and /R/; RedClus = percentage of reduction of initial consonant clusters from two consonants to one; CCVC = percentage of correct syllable structure CCVC; NWI = nonword imitation; Level 4 = percentage of correct consonants /b/, /f/, and /ʋ/; CVC = percentage of correct syllable structure CVC; WR = word repetition; PWV word = proportion of whole-word variability: word repetition; NWR = nonword repetition; PWV nonword = proportion of whole-word variability: nonword repetition; MRR = maximum repetition rate; MRR-pa = number of syllables per second of sequence /pa/; MRR-ta = number of syllables per second of sequence /ta/; MRR-ka = number of syllables per second of sequence /ka/; MRR-pataka = number of syllables per second of sequence /pataka/; MRR-pata = number of syllables per second of sequence /pata/; MRR-taka = number of syllables per second of sequence /taka/.

Table B2. Point-to-point agreement for interrater reliability for phonetic transcription.

		% Agreement		
Task	***n***	**Segments**	**Consonants**	**Vowels**
PN	100	95.7	95.5	95.8
NWI	97	94.7	95.1	94.4

Note. PN = picture naming; NWI = nonword imitation.

Table B3. Intraclass correlation coefficients and descriptives (means and standard deviations) for test–retest reliability per task.

						T1		T2		
Task	**Parameter**	***n***	**ICC**	**CI lb**	**CI ub**	***M***	***SD***	***M***	***SD***	**Sig.**[a]
PN	PCCI	106	.51	.39	.67	97.5	4.25	98.6	3.38	.002
	PVC	106	.31	.15	.52	97.9	3.47	98.9	2.44	.000
	Level 5	106	.71	.63	.81	95.6	13.2	97.6	9.94	.043
	RedClus[b]	105	.75	.68	.86	-3.42	8.49	-2.93	10.4	.10
	CCVC	105	.76	.69	.85	95.1	11.5	96.3	12.4	.055
NWI	PCCI	107	.74	.63	.88	89.1	9.70	92.6	7.76	.000
	PVC	107	.77	.69	.89	94.2	6.59	96.5	5.33	.000
	Level 4	107	.52	.37	.72	90.4	11.8	93.4	9.00	.007
	Level 5	107	.73	.65	.84	87.7	16.0	91.3	13.4	.001
	RedClus[b]	105	.41	.26	.59	-5.71	9.76	-6.57	13.8	.51
	CVC	106	.88	.84	.94	93.9	10.4	96.1	8.59	.000
	CCVC	105	.60	.50	.76	87.0	21.4	84.9	25.9	.71
WR	PWV word	86	.66	.56	.82	0.26	0.051	0.25	0.059	.65
NWR	PWV nonword	75	.39	.21	.64	0.32	0.093	0.32	0.076	.69
MRR	MRR-pa	65	.70	.61	.85	4.63	0.73	4.69	0.64	.66
	MRR-ta	63	.53	.35	.75	4.50	0.61	4.52	0.61	.75
	MRR-ka	63	.60	.47	.79	4.25	0.55	4.34	0.54	.29
	MRR-pataka	54	.18	-.070	.36	3.91	0.86	4.35	0.83	.004
	MRR-pata	59	.46	.28	.68	4.48	0.86	4.71	0.93	.072
	MRR-taka	54	.43	.25	.67	4.63	0.88	4.57	0.70	.81

Note. ICC = intraclass correlation coefficient; CI lb= confidence interval lower bound; CI ub= confidence interval upper bound; PN = picture naming; PCCI = percentage of consonants correct in syllable-initial position; PVC = percentage of vowels correct; Level 5 = percentage of correct consonants /l/ and /R/; RedClus = percentage of reduction of initial consonant clusters from two consonants to one; CCVC = percentage of correct syllable structure CCVC; NWI = nonword imitation; Level 4 = percentage of correct consonants /b/, /f/, and /ʋ/; CVC = percentage of correct syllable structure CVC; WR = word repetition; PWV word = proportion of whole-word variability: word repetition; NWR = nonword repetition; PWV nonword = proportion of whole-word variability: nonword repetition; MRR = maximum repetition rate; MRR-pa = number of syllables per second of sequence /pa/; MRR-ta = number of syllables per second of sequence /ta/; MRR-ka = number of syllables per second of sequence /ka/; MRR-pataka = number of syllables per second of sequence /pataka/; MRR-pata = number of syllables per second of sequence /pata/; MRR-taka = number of syllables per second of sequence /taka/.
[a]Statistical significance of the T1 and T2 differences were calculated using the Wilcoxon signed-ranks test. The significance level was set at.05.
[b]The percentage cluster reductions (RedClus) was inverted (multiplied by -1) such that higher values reflected a more accurate performance.

Table B4. Descriptives (means and standard deviations) of the extracted parameters for the picture-naming and nonword- imitation tasks for the younger age groups (2;0–3;11).

Age range (years; months)	**2;0–2;3**		**2;4–2;7**		**2;8–2;11**		**3;0–3;3**		**3;4–3;7**		**3;8–3;11**	
	M	***SD***	***M***	***SD***	***M***	***SD***	***M***	***SD***	***M***	***SD***	***M***	***SD***
PN												
PCCI	76.0	14.1	81.4	14.1	87.7	10.0	90.9	7.8	92.0	7.2	93.4	6.3
PVC	87.6	9.2	89.8	7.6	93.3	4.9	94.9	4.1	95.2	3.8	96.6	3.1
RedClus[a]	-49.6	26.1	-34.6	26.3	-20.0	21.5	-13.2	17.1	-11.1	16.2	-7.1	14.5
Level 5	49.7	27.2	59.1	27.2	69.6	24.7	76.7	21.6	81.6	18.8	86.1	18.1
CCVC	39.7	28.0	56.9	28.4	75.4	24.9	82.0	19.3	84.4	18.5	88.6	17.2
NWI												
PCCI	67.2	12.7	72.1	11.9	77.3	10.3	79.7	10.3	83.7	9.1	84.6	10.1
PVC	81.8	9.2	86.5	6.7	89.3	6.4	91.9	5.6	93.5	4.8	94.1	4.9
RedClus[a]	-54.5	29.1	-42.2	28.1	-35.1	26.5	-26.3	21.4	-23.1	23.1	-13.5	16.6
Level 4	26.0	36.0	45.6	40.1	55.6	35.6	60.9	33.9	64.0	32.9	74.9	28.4
Level 5	63.1	17.6	68.4	17.0	75.4	14.9	76.3	15.8	81.5	12.7	82.5	13.8
CVC	93.0	7.3	93.5	5.2	96.0	4.4	96.0	4.0	97.7	3.1	97.4	4.1
CCVC	66.7	47.7	72.2	45.1	77.8	41.9	85.0	35.9	85.9	35.0	94.1	23.7

Note. PN = picture naming; PCCI = percentage of consonants correct in syllable-initial position; PVC = percentage of vowels correct; RedClus = percentage of reduction of initial consonant clusters from two consonants to one; Level 5 = percentage of correct consonants /l/ and /R/; CCVC = percentage of correct syllable structure CCVC; NWI = nonword imitation; Level 4 = percentage of correct consonants /b/, /f/, and /ʊ/; CVC = percentage of correct syllable structure CVC.

[a]The percentage of cluster reductions (RedClus) was inverted (multiplied by -1) such that higher values reflect a more accurate performance.

Table B5. Descriptives (means and standard deviations) of the extracted parameters for the picture-naming and nonword-imitation tasks for the older age groups (4;0–6;11).

Age range (years; months)	**4;0–4;3**		**4;4–4;7**		**4;8–4;11**		**5;0–5;3**		**5;4–5;7**		**5;8–5;11**		**6;0–6;5**		**6;6–6;11**	
	M	***SD***	***M***	***SD***	***M***	***SD***	***M***	***SD***	***M***	***SD***	***M***	***SD***	***M***	***SD***	***M***	***SD***
PN																
PCCI	95.2	5.2	96.7	3.3	96.9	3.4	96.8	3.7	97.0	5.8	97.9	3.3	98.2	2.5	98.4	2.2
PVC	97.0	3.9	97.9	2.8	98.0	2.2	97.7	3.1	97.6	5.5	98.5	2.4	98.4	2.3	98.6	1.8
RedClus[a]	-4.8	7.1	-3.2	5.9	-2.8	6.8	-3.0	6.2	-2.3	6.2	-1.7	5.1	-1.7	3.3	-0.8	2.0
Level 5	91.9	12.4	91.9	13.6	95.3	8.9	93.6	11.0	95.5	9.2	96.5	10.7	97.3	5.7	98.1	4.9
CCVC	92.6	10.6	95.3	8.0	95.2	8.1	94.4	9.9	94.9	9.4	96.8	7.0	96.9	5.5	97.5	4.5
NWI																
PCCI	82.5	10.1	87.2	9.8	86.9	9.6	87.7	6.9	89.0	7.2	89.6	7.4	90.9	6.2	91.8	7.0
PVC	90.5	7.5	92.4	6.5	93.0	6.7	93.5	4.8	94.4	4.5	94.3	4.7	96.1	3.8	94.8	5.7
RedClus[a]	-12.8	13.7	-8.0	12.2	-9.3	12.9	-7.7	11.8	-6.6	11.9	-6.8	11.8	-6.4	10.5	-4.2	8.5
Level 4	79.2	15.0	87.0	12.0	85.8	14.3	87.8	11.8	90.0	9.9	88.9	10.3	92.3	8.6	92.8	9.6
Level 5	80.3	16.2	84.3	16.3	84.1	15.7	87.2	12.8	86.3	12.9	87.9	14.6	89.4	11.7	89.7	11.2
CVC	91.1	8.4	92.6	7.4	92.7	7.6	93.3	5.7	94.4	5.5	95.0	5.0	96.6	4.0	95.2	5.8
CCVC	74.9	25.7	82.1	22.3	79.8	25.7	83.0	25.2	85.6	21.7	84.1	24.4	86.1	21.6	90.8	16.2

Note. PN = picture naming; PCCI = percentage of consonants correct in syllable-initial position; PVC = percentage of vowels correct; RedClus = percentage of reduction of initial consonant clusters from two consonants to one; Level 5 = percentage of correct consonants /l/ and /R/; CCVC = percentage of correct syllable structure CCVC; NWI = nonword imitation; Level 4 = percentage of correct consonants /b/, /f/, and /ʋ/; CVC = percentage of correct syllable structure CVC.
[a]The percentage of cluster reductions (RedClus) was inverted (multiplied by -1) such that higher values reflect a more accurate performance

Table B6. Descriptives (means and standard deviations) of the extracted parameters for the word and nonword repetition tasks for the younger age groups (2;0–3;11).

PWV		**2;0–2;3**	**2;4–2;7**	**2;8–2;11**	**3;0–3;3**	**3;4–3;7**	**3;8–3;11**
WR	*M*	0.30	0.31	0.30	0.30	0.28	0.28
	SD	0.11	0.09	0.07	0.09	0.09	0.08
NWR	*M*	0.36	0.35	0.33	0.34	0.33	0.30
	SD	0.11	0.10	0.12	0.11	0.11	0.07

Note. PWV = proportion of whole-word variability; WR = word repetition; NWR = nonword repetition.

Table B7. Descriptives (means and standard deviations) of the extracted parameters for the word and nonword repetition tasks for the older age groups (4;0–6;11).

PWV		4;0–4;3	4;4–4;7	4;8–4;11	5;0–5;3	5;4–5;7	5;8–5;11	6;0–6;5	6;6–6;11
WR	*M*	0.24	0.24	0.24	0.23	0.23	0.23	0.23	0.23
	SD	0.05	0.05	0.05	0.04	0.05	0.05	0.04	0.04
NWR	*M*	0.27	0.28	0.27	0.28	0.27	0.26	0.26	0.25
	SD	0.07	0.08	0.07	0.08	0.07	0.07	0.07	0.06

Note. PWV = proportion of whole-word variability; WR = word repetition; NWR = nonword repetition.

Table B8. Maximum repetition rates (MRRs) per sequence and age group (number of syllables per second).

Sequence		3;0–3;3	3;4–3;7	3;8–3;11	4;0–4;3	4;4–4;7	4;8–4;11	5;0–5;3	5;4–5;7	5;8–5;11	6;0–6;5	6;6–6;11
MRR-pa	*M*	4.00	4.02	4.08	4.29	4.54	4.57	4.64	4.80	4.82	4.92	5.01
	SD	0.54	0.49	0.54	0.69	0.50	0.64	0.61	0.57	0.61	0.51	0.58
MRR-ta	*M*	3.88	3.99	4.06	4.15	4.41	4.44	4.44	4.69	4.69	4.85	4.93
	SD	0.57	0.58	0.62	0.65	0.55	0.60	0.60	0.54	0.61	0.67	0.61
MRR-ka	*M*	3.61	3.75	3.85	3.98	4.15	4.21	4.34	4.35	4.42	4.50	4.60
	SD	0.51	0.56	0.53	0.56	0.57	0.58	0.51	0.45	0.49	0.50	0.55
MRR-pataka	*M*	3.48	3.70	3.64	3.74	3.93	3.96	4.09	4.14	4.37	4.39	4.55
	SD	0.71	0.98	0.89	0.76	0.77	0.97	0.82	0.88	0.89	0.98	0.93
MRR-pata	*M*	4.11	3.93	4.04	4.37	4.42	4.47	4.49	4.71	4.58	4.77	4.80
	SD	0.72	0.61	0.75	0.91	0.78	1.02	0.73	0.75	0.83	0.90	0.82
MRR-taka	*M*	3.82	4.04	4.09	4.28	4.37	4.47	4.37	4.54	4.74	4.63	4.96
	SD	0.72	0.85	0.82	0.75	0.80	0.78	0.75	0.61	0.77	0.74	0.78

Note. MRR-pa = number of syllables per second of sequence /pa/; MRR-ta = number of syllables per second of sequence /ta/; MRR-ka = number of syllables per second of sequence /ka/; MRR-pataka = number of syllables per second of sequence /pataka/; MRR-pata = number of syllables per second of sequence /pata/; MRR-taka = number of syllables per second of sequence /taka/.

Appendix C

Examples of the modeling results for the main parameters of the computer articulation instrument (mean age in months).

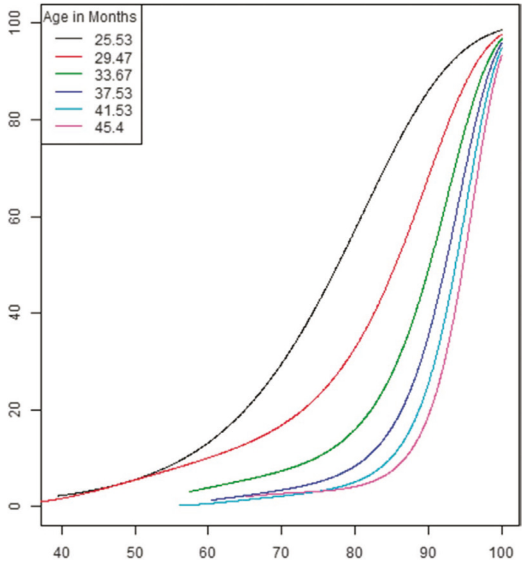

Figure 1. Continuous norms per age group for picture naming percentage of consonants correct in syllable-initial position, age 2;0–3;11.

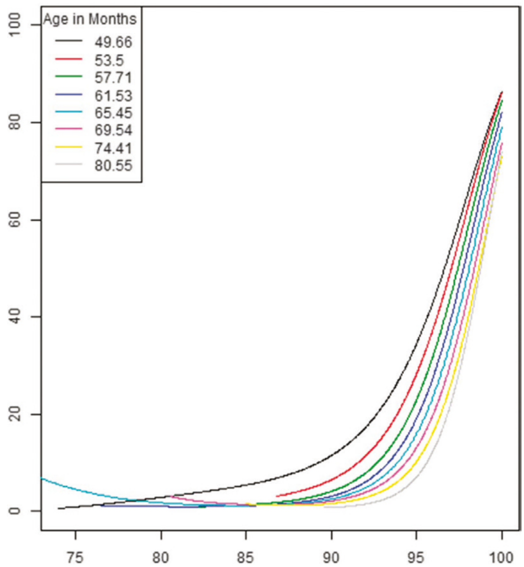

Figure 2. Continuous norms per age group for picture naming percentage of consonants correct in syllable-initial position, age 4;0–6;11.

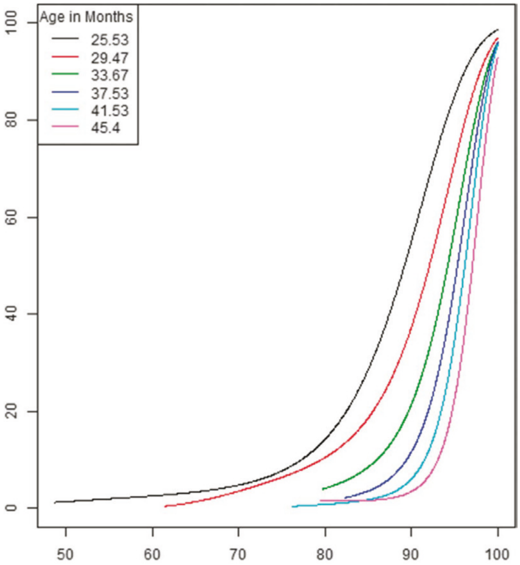

Figure 3. Continuous norms per age group for picture naming percentage of vowels correct, age 2;0–3;11.

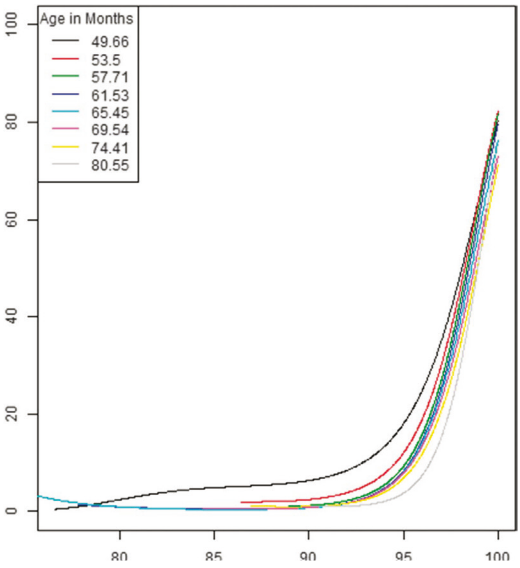

Figure 4. Continuous norms per age group for picture naming percentage of vowels correct, age 4;0–6;11.

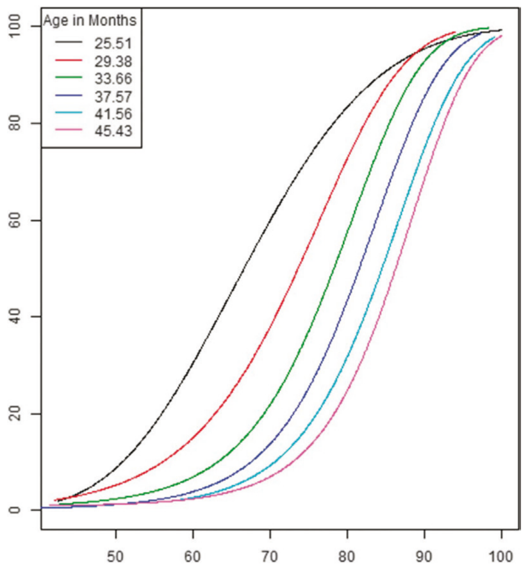

Figure 5. Continuous norms per age group for nonword imitation percentage of consonants correct in syllable-initial position, age 2;0–3;11.

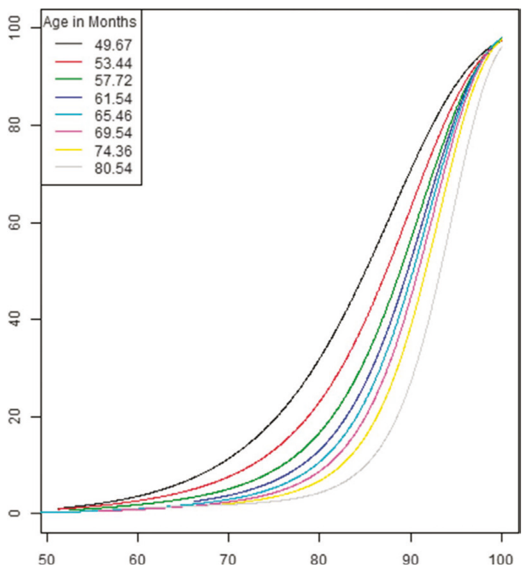

Figure 6. Continuous norms per age group for nonword imitation percentage of consonants correct in syllable-initial position, age 4;0–6;11.

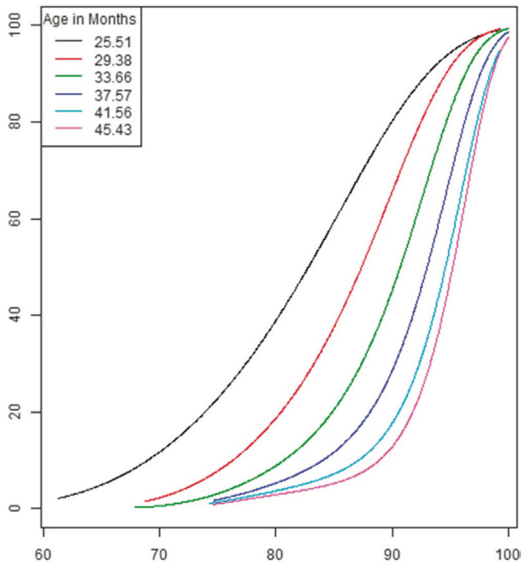

Figure 7. Continuous norms per age group for nonword imitation percentage of vowels correct, age 2;0–3;11.

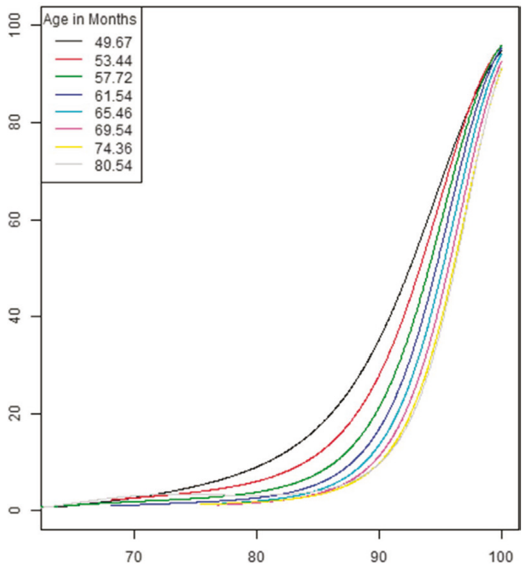

Figure 8. Continuous norms per age group for nonword imitation percentage of vowels correct, age 4;0–6;11.

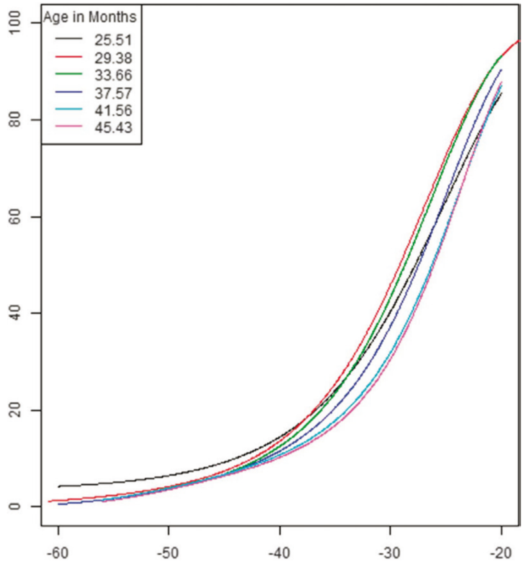

Figure 9. Continuous norms per age group for word repetition; proportion whole-word variability, age 2;0–3;11.

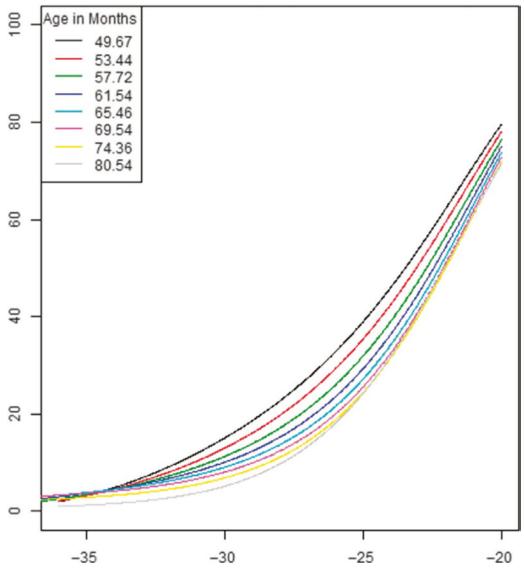

Figure 10. Continuous norms per age group for word repetition; proportion whole-word variability, age 4;0–6;11.

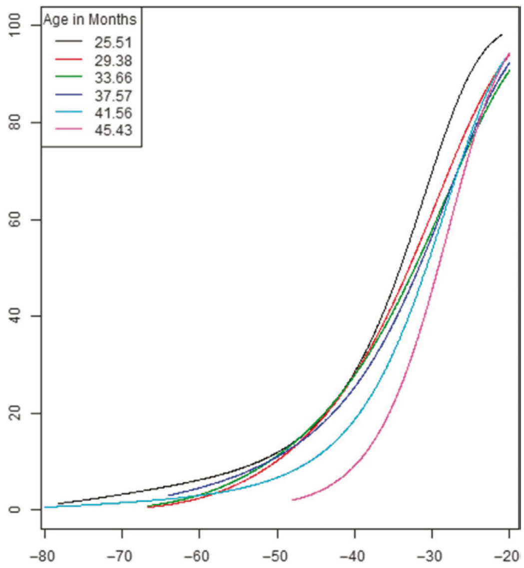

Figure 11. Continuous norms per age group for nonword repetition; proportion whole-word variability, age 2;0–3;11.

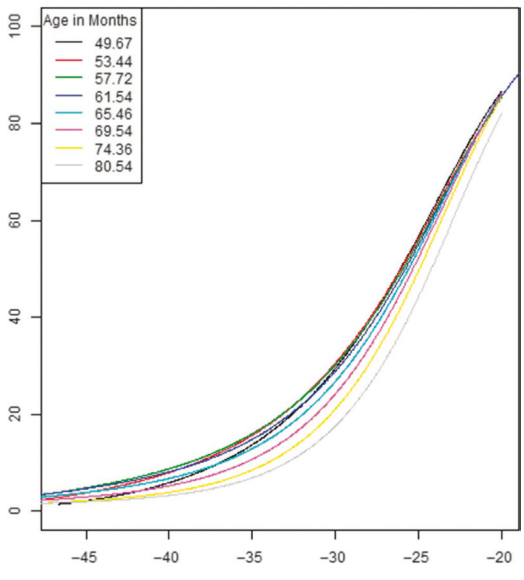

Figure 12. Continuous norms per age group for nonword repetition; proportion whole-word variability, age 4;0–6;11.

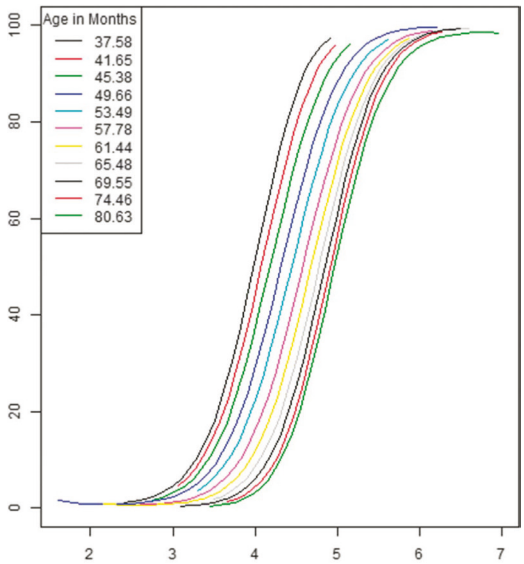

Figure 13. Continuous norms per age group for number of syllables per second of sequence /pa/, age 3;0–6;11.

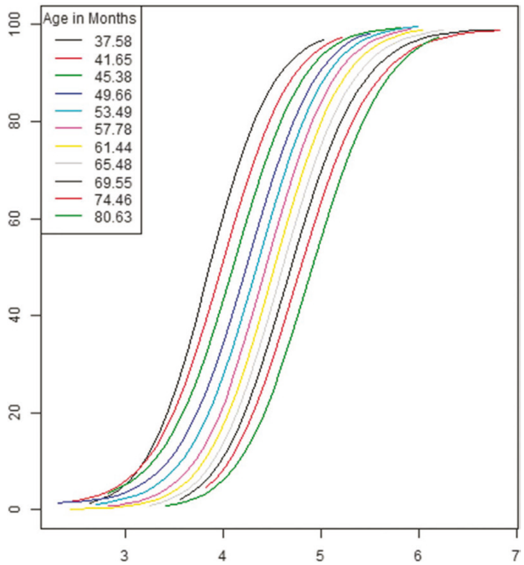

Figure 14. Continuous norms per age group for number of syllables per second of sequence /taka/, age 3;0–6;11.

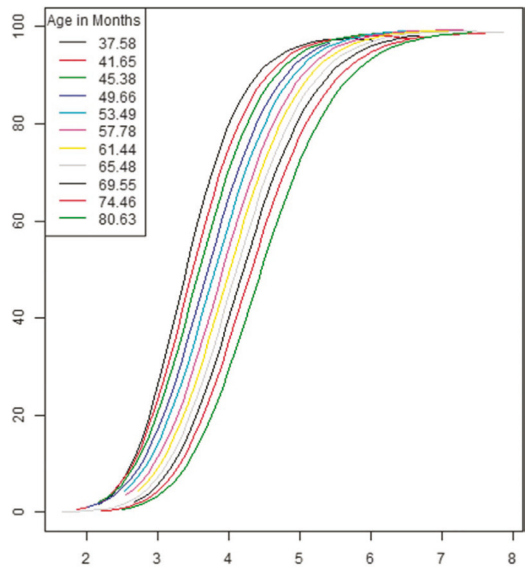

Figure 15. Continuous norms per age group for number of syllables per second of sequence /pataka/, age 3;0–6;11.

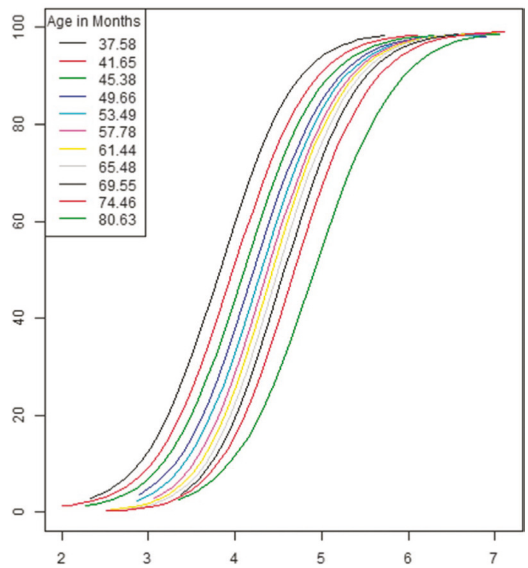

Figure 16. Continuous norms per age group for number of syllables per second of sequence /taka/, age 3;0–6;11.

2

THREE

CHAPTER 3

SPEECH SOUND DEVELOPMENT

IN TYPICALLY DEVELOPING 2- TO 7-YEAR-OLD DUTCH-SPEAKING CHILDREN:

A NORMATIVE CROSS-SECTIONAL STUDY

Leenke van Haaften, Sanne Diepeveen, Lenie van den Engel-Hoek,
Bert de Swart and Ben Maassen

ACCEPTED FOR PUBLICATION IN
INTERNATIONAL JOURNAL OF LANGUAGE & COMMUNICATION DISORDERS

Abstract

Background: Dutch is a West-Germanic language spoken natively by around 24 million speakers. Although studies on typical Dutch speech sound development have been conducted, norms for phonetic and phonological characteristics of typical development in a large sample with a sufficient age range are lacking.

Aim: To give a detailed description of the speech sound development of typically developing Dutch-speaking children from 2 to 7 years.

Methods & Procedures: A total of 1,503 typically developing children evenly distributed across the age range of 2;0-6;11 years participated in this normative cross-sectional study. The picture-naming task of the Computer Articulation Instrument (CAI) was used to collect speech samples. Speech development was described in terms of (1) percentage consonants correct-revised (PCC-R) and percentage vowels correct (PVC), (2) consonant, vowel, and syllabic structure inventories, (3) degrees of complexity (phonemic feature hierarchy) and (4) phonological processes.

Outcomes & Results: A two-way mixed ANOVA confirmed a significant increase in the number of PCC-R and PVC between the ages of 2;0 and 6;11 years ($p < .001$). The consonant inventory was found to be complete at 3;7 years of age for the syllable-initial consonants, with the exception of the voiced fricatives /v/ and /z/, and the liquid /r/. All syllable-final consonants were acquired before the age of 4;4 years. At the age of 3;4 years, all children had acquired a complete vowel inventory and at the age of 4;7 years they produced most syllable structures correctly, albeit that the syllable structure CCVCC was still developing. All phonological contrasts were produced correctly at 3;8 years of age. Children in the younger age groups used more phonological simplification processes than the older children and by the age of 4;4 years, all had disappeared, except for the initial cluster reduction from three to two consonants and the final cluster reduction from two to one consonant.

Conclusions & Implications: This paper describes a large normative cross-sectional study of Dutch speech sound development which, in clinical practice, can help Dutch speech language pathologists to differentiate children with delayed or disordered speech development from typically developing children.

Introduction

Typical speech sound development can be described as the acquisition of individual speech sounds and the organization of these speech sounds into speech patterns, encompassing both the phonetic (i.e. articulatory) and the phonological (i.e. phonemic) development. The term 'phonetic' refers to speech sound production, that is, articulatory skills, whereas the term 'phonemic' refers to speech sound use and function, and thus the organization of the speech sound system (Dodd, 2003). Speech sound production requires physiological movements to be made such that speech sounds can be recognized, in other words, movements that cause the production of the main features of recognizable sounds (place, manner, voice). In the process of phonetic acquisition, a distinction can be made between phonetic development prior to word learning and phonetic development in words (Winitz, 1969), where the first process has a physiological basis in that the child learns sounds falling within and outside the context of its ambient language. The phonetic development in words, however, comprises the acquisition of movements by which the relevant features of place, manner, and voice can be produced in a continuous phonetic context, and may be less of a physiological process in the sense that it involves a stable sound-meaning relationship (Winitz, 1969). Phonological development is characterized by the increase of phonological contrasts and the decrease of simplification processes. In clinical descriptions, the systematic differences between adult target sounds and children's realizations are described in terms of simplification processes, which can be defined as typical error patterns children produce during speech development. These simplifications involve substitution processes, where one sound is systematically substituted for another sound, assimilation processes, when a sound becomes the same or similar to another sound in the word, or syllable structure processes that affect the syllabic structure of a word. Simplification processes occur as the result of natural limitations and capacities of human speech production and perception (Dodd et al., 2003), where children try to solve these limitations by approaching the problematic target sounds or sound sequences of the target adult word with sounds that are already incorporated in their phonological system (Beers, 1995).

One of the theoretical approaches that explains the intertwinement of phonetic and phonological development is the Articulatory Phonology model (Namasivayam et al., 2020). This model describes a perspective that is based on the notion of an articulatory "gesture" that serves as a unit of phonological contrast and characterization of the resulting articulatory movements. Following this model, measuring speech in words or context involves both phonetics and phonology. Consistent production of a speech sound in context, indicates both an articulatory (phonetic) and phonological mastery of this speech sound.

A phonetic inventory of speech sounds in words catalogues those speech sounds that a child can produce in initial, medial, and final positions in syllables or words. Over and above such a phonetic inventory, one can conduct a phonological analysis, where error patterns are identified that characterize the mismatches between a child's production and adult target form in terms of simplification processes. A hierarchical analysis in terms of contrastive features (e.g., /p/ vs. /k/ or /p/ vs. /b/) provides indications regarding the child's organization of its phonological system, with, among other features, [dorsal] contrasts being required to distinguish /k/ from /p/ and [voice] to distinguish /p/ and /b/ (Ingram and Ingram, 2001). This phonological inventory thus describes the system of contrasts a child can produce.

In recent years, many studies have been carried out to investigate typical speech sound development in different languages, among which are Putonghua (Modern Standard Chinese) (Hua and Dodd, 2000), British English (Dodd et al., 2003), Maltese (Grech and Dodd, 2008), Québécois French (MacLeod et al., 2011), isiXhosa (Maphalala et al., 2014), Malay (Phoon et al., 2014), Swahili (Gangji et al., 2015), Setswana (Mahura and Pascoe, 2016), Haitian Creole (Archer et al., 2017), Danish (Clausen and Fox-Boyer, 2017), South African English (Pascoe et al., 2018), and Italian (Tresoldi et al., 2018). Providing a cross-linguistic review of children's consonant acquisition, McLeod and Crowe (2018) concluded that in all languages five-year-old children have acquired most consonants, with individual languages differing only in the specific consonants that have not yet been mastered at that age.

Dutch phonetics and phonology

A range of studies have examined the typical speech sound development of Dutch-speaking children (Beers, 1995; Fikkert, 1994; Jongstra, 2003; Levelt, 1994; Levelt et al.; 2000, Priester and Goorhuis-Brouwer, 2013; Stes, 1977; Van den Berg et al., 2017). Dutch is a West-Germanic language and the majority language in the Netherlands and parts of Belgium, as well as in Suriname, Aruba and the Dutch Antilles. It is spoken natively by around 24 million speakers (Rys et al., 2017), with 16% speaking more than one other language, which mainly includes English, French, German, and Frisian (Fernhout et al., 2011). Of note here is that Dutch children typically learn English from the age of 10 years. English has long been a compulsory subject in all types of Dutch secondary education and since 1986 in the two final years of primary education.

The 19 consonants of Dutch and four additional consonants in parentheses are presented in Table 1. All consonants can occur in syllable-initial position, except for /ŋ/. Any consonant can occur in word-final position, except for voiced plosives, voiced fricatives, and /h/. The consonants /c, ʃ, ʒ, ɲ/ only occur in loanwords and/or as allophones (e.g. *jasje* [jɑʃ-ʃə] 'jacket'). The 16 vowels in Dutch can be divided into a set of long vowels /i, y, u, e, ø, o, a/, a set of short vowels /ɪ, ɛ, ɔ, ʉ, ɑ/, a reduced vowel /ə/, and three diphthongs /ɑu, ɛi, ʉy/ (Mennen et al., 2006). Long vowels, diphthongs, and the schwa can occur in syllable- and word-final position, as in *knie* [kni] 'knee' and *vrij* [vrɛi] 'free', whereas short vowels cannot occur at the end of a syllable or word, e.g. *kapstok* [kɑp-stɔk] 'coat rack'. The height classification for Dutch vowels shows two high vowels /i, u/, four high mid vowels /e, ɪ, o ɔ/, one low mid vowel /ɛ/, and two low vowels /a, ɑ/ (Levelt, 1994). In Dutch, like in English, a syllable consists of a vowel, from zero to three consonants in syllable-initial position, and from zero to four consonants in syllable-final position ($C^{0\text{-}3}VC^{0\text{-}4}$) (Collins and Mees, 2003), e.g. *strand* [strɑnt] 'beach' and *herfst* [hɛrfst] 'autumn'.

Table 1. The consonants of Dutch.

Place of articulation	Manner of articulation				
	Plosives	**Fricatives**	**Nasals**	**Liquids**	**Glides**
Bilabial	p, b		m		
Labiodental		f, v			w
Alveolar	t, d	s, z	n	l, r	
Post alveolar	(c)	(ʃ), (ʒ)	(ɲ)		
Palatal					j
Velar	k, (g)	x	ŋ		
Glottal		h			

Note. Four additional consonants are presented in parentheses because they only occur in loanwords and/or as allophones

Typical Dutch speech sound development

One of the first studies of typical speech sound development in Dutch was performed by Stes, who, in 1977, had 480 children aged between 3 and 10 years complete a single-word-naming task. This study was focused on the phonetic acquisition of vowels, consonants, and consonant clusters, yielding a phonetic inventory of speech sounds in Dutch words. Determining the age of acquisition (75% of the children) and age of mastery (90% of the children), he showed that all vowels were already present at the age of three years and that at around the age of four most consonants were correctly produced by 75% of the children, with an exception for /s/ and /r/. More recently, Priester and Goorhuis-Brouwer (2013) also used a picture-naming task to chart the phonetic acquisition of speech sounds in 1,035 typically developing Dutch children between the ages of 3;8 and 6;3 years. They observed that all children older than 4;3 years pronounced most sounds (single consonants and consonant clusters) correctly.

So far, only one study looked into the typical speech sound development of Dutch-speaking children in phonological terms. Besides phonetic acquisition, Beers (1995) studied the acquisition of phonological contrasts and occurrence of phonological processes in 90 children aged between 1;3 and 4;0 years using samples of spontaneous speech. The normative data from this study are still used by clinicians to determine whether a child's speech pattern is age-appropriate, delayed, or deviant. Beers (1995) analysed the order of acquisition of Dutch consonants in syllable-initial position and found that the children aged between 1;3 and 1;8 years had acquired the consonants /p/, /t/, /m/, /n/ and /j/, reflecting the use of the contrastive features 'sonorant', 'labial', and 'coronal'. Around age 1;9 and 1;11 years, children were able to produce the consonant /k/ correctly, thereby showing they had acquired the contrastive 'dorsal' feature. Between the ages of 2;0 and 2;2 years, the children acquired the contrast 'continuant', as indicated by the correct production of the continuants /s/, /x/, and /h/. Between 2;3 and 2;5 years, children were able to pronounce /b/, /f/, and /w/ correctly, indicating that the contrastive features 'front', 'round', and 'voice' had been mastered. The children aged between 2;6 and 2;8 years had learned to use the contrasts 'nasal', 'lateral', and 'rhotic', as was shown by the correct production of the liquids /l/ and /r/. To summarize, Dutch children were able to use all contrasts correctly at 2;8 years of age. Based on this sequence of acquisition, Beers proposed a 5-level phonemic feature hierarchy, which is presented in Table 2.

Table 2. Degrees of Complexity of phonological contrasts of Dutch syllable-initial consonants described by Beers (1995).

Degree of Complexity	Contrastive feature	Segments
Degree 1	Sonorant, labial, coronal	/p/, /t/, /m/, /j/, /n/
Degree 2	Dorsal	/k/
Degree 3	Continuant	/s/, /x/, /h/
Degree 4	Front, round	/b/, /f/, /w/
Degree 5	Lateral, rhotic, nasal	/l/, /r/

Exploring simplification processes in the same sample, Beers (1995) noted that typically developing Dutch children aged between 1;3 and 1;11 years commonly used the syllable structure processes of cluster reduction, final consonant deletion, weak syllable deletion, reduplication and assimilation, and the substitution processes of (de)voicing, fronting, gliding, stopping, and vocalization. Simplifications such as reduplication and final consonant deletion, and assimilation processes showed a sharp decline in their occurrence between the ages of 2;0 and 2;5 years, while the occurrence of cluster reduction and weak syllable deletion decreased between 2;6

and 3;0 years. Only the substitution process of gliding continued to be used until the age of 4;0 years.

A year earlier, Levelt (1994) had reported on the mean percentage of vowels correct (PVC) for Dutch-speaking children, finding that the high vowels /i, u/ and the low vowels /a, ɑ/ are acquired first, while the low-mid vowel /ɛ/ is mastered last. In other Dutch studies the acquisition of syllable structures was investigated (Fikkert, 1994; Levelt et al., 2000; Van den Berg et al., 2017), as well as word-initial consonant clusters (Jongstra, 2003), and place features and vowel height (Levelt, 1994). Van den Berg et al. (2017), Fikkert (1994), and Levelt et al. (2000) concluded that simple syllable types (CV, V, and CVC) appear simultaneously and before complex syllable types. In most of the children examined, onset clusters emerged before final clusters, while the order of acquisition of complex clusters was found to be variable (Jongstra, 2003; Van den Berg et al., 2017). All studies mentioned were based on spontaneous speech samples, apart from the study by Jongstra (2003), who used a picture-naming task.

Priester et al. (2011) reviewed the British-English and Dutch literature on normative data for speech sound development and found a universal trend for the two languages. In both, all vowels are mastered at three years of age and most single consonants are present around the age of four, except for /s/ and /r/. A difference between English and Dutch was found in the age of acquisition of consonant clusters. In English, most consonant clusters were mastered by the age of five (Dodd et al., 2003), whereas in Dutch most clusters were not acquired until the age of six, with the development possibly even continuing up to the age of 10 (Stes, 1977). Priester et al. (2011) suggest that these differences may be caused by language differences, Stes' data (1977) being outdated, and/or differences in the analysis methods used. Of note, Dodd et al.'s (2003) was a broad description of the development of consonant clusters, while that of Stes' (1977) was based on a detailed analysis. However, Smit (1993) showed that, although all initial consonant clusters are produced as clusters in typically developing English-speaking children by the age of 5;0 years, there may continue to be segmental errors within these clusters. Also other studies report that in English the development of consonant clusters still continues after 5;0 years of age (McLeod et al., 2001).

Thus, although multiple studies are available on the typical speech sound development of Dutch-speaking children, no recent studies have focused on both the phonetic and phonological aspects of this process in a sufficiently large sample that includes a sufficiently wide age range. All Dutch studies on the acquisition of vowels and syllable structures were conducted in small groups of children (n = 12 to n = 45) comprising young children only, with ages ranging between 6 months and 3;4 years (Fikkert, 1994; Jongstra, 2003; Levelt, 1994; Levelt et al., 2000; Van den Berg et al., 2017). The Stes (1977) and Priester and Goorhuis-Brouwer (2013) studies did have large samples, but both only reported on phonetic development, with the latter study being restricted to consonants. Furthermore, having been collected in the late 1970s, the findings Stes reports are most likely at least partly outdated. Also, even though Beers (1995) did describe both phonetic and phonological features, she did so on the basis of observations obtained in 90 children. Moreover, there is no research on the percentage of consonants correct (PCC) in Dutch, notably the most well-known and well-established measures used in clinical practice that is frequently cited in research literature (Fabiano-Smith, 2019; Masso et al., 2018). Accordingly, there is a clear need for norms of speech sound development for the Dutch language that are clinically-sensitive to differentiate children with delayed or disordered speech development from typically developing children (Dodd et al., 2003), where delayed speech manifests itself in error patterns that are typical of a younger chronological age and disordered speech by error patterns that are atypical of any age group in a normative sample (Dodd, 2011).

Methods of speech elicitation for the assessment of speech

There are different methods to elicit speech for assessment purposes. The studies on typical Dutch speech acquisition mentioned above used two such methods: conversational or spontaneous speech and single word naming (using a picture-naming or word-imitation task). The advantages of both techniques have been described extensively (Masterson et al., 2005; Wolk and Meisler, 1998; Morrison and Shriberg, 1992), with both methods having been shown to be useful for clinical assessments (Masterson et al., 2005; Wolk and Meisler, 1998). Conversational or spontaneous speech has the advantage of providing phonetic contexts while allowing the child's abilities to be tested in real-life, natural communication. On the other hand, spontaneous speech introduces undesired variability due to individual differences in the propensity and motivation to talk, such that the child might not perform at maximum level and, for instance, avoid problematic target sounds or sounds that are not yet firmly embedded in its phonological system. In addition, analysing spontaneous speech is time consuming. A word-naming task can thus be a more efficient way to elicit and analyse speech in children, with the target words covering all aspects of Dutch speech sound production.

The current study

With this cross-sectional study we aim to give a detailed description of the speech sound development of Dutch-speaking, typically developing children and provide normative data for use in clinical practice to differentiate children with speech sound disorders (SSDs) from children showing typical development. To ensure efficiency in our data collection and analysis, we opted for a picture-naming task to elicit speech, of which the audio recordings were evaluated, scoring the following parameters: PCC and PVC, consonant, vowel, and syllable-structure inventories, degrees of complexity (phonemic feature hierarchy), and phonological processes.

Methods

Research design

A cross-sectional design was used to identify trends of speech sound development.

Recruitment of participants

This study analyses the speech samples of the picture-naming task collected within the framework of our group's normative study of the Computer Articulation Instrument (CAI); see Van Haaften et al. (2019a) and Maassen et al. (2019) for information on the data-collection method and sample characteristics. The children were aged between 2;0 and 6;11 years and drawn from 47 nurseries and 71 elementary schools located in four different regions of the Netherlands. The nurseries and schools were sent a letter explaining the purpose of the study and inviting them to participate. All parents of the children in the participating nurseries and schools were handed an information letter. After the signed parental consent form had been received, the child was included in the study. The 4- to 7-year-old children were recruited between January 2008 and December 2014, and the 2- to 4 year-olds from March 2011 to April 2015.

Participants

Of the total of 1,524 children participating in the CAI normative study, 1,503 children completed the picture-naming task. We opted for the age range of 2;0 and 6;11 because during this period speech sound development is expected to be completed. The minimum age of 2;0 years was chosen because at that age a child's vocabulary and attention span is sufficient for a picture-naming task. Stratifying for age, 14 groups were created with a range of 4 months for children aged 2;0-5;11 years and a range of 6 months for those aged 6;0-6;11 years. As is recommended for the assessment of speech language development (Andersson, 2005), all age groups contained more than 100 children, except for the youngest age group (*n* = 72) and the group of 4;0-4;3-year-olds (*n* = 99).

The criteria for inclusion were: no hearing loss and Dutch being the spoken language at the nursery or primary school. The parents and teachers of eligible children were asked to complete a questionnaire about the children's development. Another language than Dutch (e.g. Turkish, Arabic, or German) was spoken at home in 3.9% (*n* = 59) of the participants. To ensure the normative sample was representative of the Dutch population, we also included children with a history of speech and language difficulties (*n* = 32, 2.1%). The sample was representative of the general Dutch population in terms of gender, geographic region, degree of urbanization, and parental socio-economic status (Van Haaften et al., 2019a). Table 3 summarizes the characteristics of the sample.

Table 3. Age and gender for the 14 age groups of the study population.

Age group (years;months)	Mean age (years;months)	Girls (*n*)	Boys (*n*)	Total (*n*)
2;0-2;3	2;1	42	30	72
2;4-2;7	2;5	46	55	101
2;8-2;11	2;10	55	46	101
3;0-3;3	3;1	51	51	102
3;4-3;7	3;6	46	61	107
3;8-3;11	3;9	45	56	101
4;0-4;3	4;2	45	54	99
4;4-4;7	4;5	53	58	111
4;8-4;11	4;10	57	55	112
5;0-5;3	5;2	53	64	117
5;4-5;7	5;5	57	71	128
5;8-5;11	5;10	52	64	116
6;0-6;5	6;2	48	69	117
6;6-6;11	6;9	62	57	119
Total		712	791	1503

Ethical considerations

The research ethics committee of the Radboud University Nijmegen Medical Centre judged that our study did not fall within the remit of the Dutch Medical Research Involving Human Subjects Act (WMO; file number: CMO 2016-2985). Therefore, the study was allowed to be carried out without approval by an accredited research ethics committee. Informed consent was obtained from all parents or legal guardians.

Materials

The speech samples recorded during the performance of the picture-naming task in the CAI study (Maassen et al., 2019) were used. The psychometric properties of this task have overall been found to be sufficient to good (Van Haaften et al., 2019a). The interrater reliability was sufficient to good, with percentages for point-to-point agreement above 95% for all measures. The construct validity of the CAI was demonstrated by the correlation of the outcomes of the CAI with age. Monotonous increases with age were found for all parameters of picture naming, such as the PCC and the PVC, and the percentages of cluster reductions and correctly produced syllable structures. Together, these results indicate that the picture-naming task of the CAI is a reliable and valid test to assess speech in typically developing Dutch children.

Our picture-naming task comprises 60 words incorporating the full repertoire of vowels, consonants, consonant clusters, and syllable structures of the Dutch language. The target words vary from simple to more complex in terms of the number of syllables, comprising 40 one-syllable words, 13 two-syllable words, 6 three-syllable words, and 1 word with four syllables (see Appendix A). The task thus assesses all Dutch phonemes in all possible syllable and word positions, except for /g/ because in Dutch this consonant only occurs in loanwords. All phonemes occur at least twice in different positions in different contexts (see Appendix B). Words were presented in a fixed order. For the 4- to 7-year-olds the complexity of words varied, while for the 2- to 4-year-olds the CVC words were presented first, followed by the words with more complex syllable structures.

Both seated in front of a computer screen, the speech language pathologist (SLP) asks the child to name the (colour) pictures that appear consecutively on the screen aloud. A pre-recorded audio prompt provided a semantic cue when the child was unable to name the picture spontaneously. When the cue did not elicit the target word, the target word was spoken by the computer, which the child then had to repeat out loud.

Procedure

The children were tested individually in a quiet room in their own nursery or primary school. The administer and child were seated side by side at a table on which a laptop computer was placed in a position comfortable for both. They both wore headsets or, if preferred, a speaker and microphone were used. All utterances were audio recorded and stored in the CAI software program.

The task was administered by 14 SLPs in the younger age groups (2-4-year olds) and in the older children (4-7-year olds) by 110 third- or fourth (final)-year SLP students working in pairs. All were trained in the administration of the CAI by the first two authors, having received precise instructions and training in the scoring procedure (phonetic transcription). Scoring was done by the same SLP(s) that had administered the test under supervision of the first two authors.

Data analysis: phonetic transcription

Each utterance of each audio recording was transcribed phonetically using the Logical International Phonetics Programs software (LIPP) (Oller and Delgado, 2000), which allows for the transcription of IPA via the traditional keyboard, along with user-designed analysis based on featural characterizations of segments. The assessors transcribed all speech recordings based on the correct target transcriptions by 'editing in' the child's production errors. They used a broad phonetic transcription in which phonetic variation (e.g. a lisp) was not represented, whereas sound distortions that resulted in a change of feature (place, manner, voice) were. The transcriptions were used to investigate:

- *PCC and PVC.* All consonants and all vowels were considered when calculating PCC and PVC, where PCC is the percentage of correctly produced consonants divided by the total number of target consonants. In this study, both common and uncommon clinical consonant distortions were scored as correct, similar to the calculation of the Percentage of Consonants Correct–Revised (PCC-R), as described by Shriberg et al. (1997), since investigating systematic distortions was not the aim of our analysis. Consistent speech sound production with or without a consistent distortion reflects both correct phonemic selection and correct phonetic production (albeit the distortion). A phonemically irrelevant consistent distortion can be diagnostically isolated from the correct phoneme selection and articulatory realization processes; the production of distorted phonemes in different contexts signifies mastery of gestures at the phonemic and articulatory level albeit the distortion itself. PVC was calculated by dividing the vowels pronounced correctly by the total number of target vowels elicited with the picture-naming task.
- *Phonetic inventory.* Applying the 75% frequency criterion, we deemed speech sounds (vowels and single-syllable initial and final consonants) to have been acquired when 75% of the children of an age group produced the targeted speech sound correctly, while a speech sound was considered to be produced correctly when a child produced the target sound ≥ 75% of the cases correctly. Like in the study of Beers (1995), this percentage was based on at least two attempts of a target sound, except for /ʒ/ in syllable-initial position as this sound only occurred once in the item list (see Appendix B for the frequency distributions of the phonological features of the picture-naming task). The mean percentages of correct productions per speech sound (vowels and single-syllable initial consonants) were calculated.
- *Degrees of complexity.* Having studied the acquisition of contrastive features in syllable-initial position in typically developing children, Beers (1995) classified the degrees of complexity for the Dutch language (see Table 2). We used her classification system (or phonemic feature hierarchy) for the present study and performed relational analyses comparing the child's productions with the target form. A specific degree of complexity was classified as age-appropriate when the syllable-initial consonants of that complexity were, on average,

correctly produced ≥ 75% of the cases by at least 75% of the children in an age group.

- *Syllable structure inventory.* A syllable structure was considered to be produced correctly when a child produced the syllable structure ≥ 75% of the cases correctly, irrespective of whether the syllable was produced correctly at the segmental level. Comparable with Gangji et al. (2015) and Clausen and Fox-Boyer (2017), we considered a syllable structure to be present in the inventory of an age group when 75% of the children produced the syllable structure correctly. Our task comprised the following syllable structures: V, CV, CVC, CCV, CVCC, CCVC, CCVCC, and CCCVC.
- *Phonological processes.* In accordance with Dodd et al. (2003), and several others (Kirk and Vigeland, 2015, Clausen and Fox-Boyer, 2017, Hua and Dodd, 2000), we classified a phonological process as age-appropriate when it fulfilled the 10% criterion, i.e. when it occurred at least 10% in at least 10% of the children within an age group. We charted both 'normal' phonological processes as described by Beers (1995) and unusual processes.

Statistical analyses

The analyses of PCC-R and PVC, phonetic inventory, degrees of complexity, syllable-structure inventory, and phonological processes consisted of a description of the data per age group.

To compare the effect of age on PCC-R and PVC and to test the hypothesis that there is a difference between PCC-R and PVC for the 14 age groups, a two-way mixed ANOVA was conducted with the percentage of correct productions as the dependent variable, type of measure as the within-subject factor with two levels (PCC-R and PVC), and age group as the between-subject factor with 14 levels (14 age groups).

Results

PCC-R and PVC

The mean scores and standard deviations of each age group for PCC-R and PVC are shown in Table 4. The mean number of both types of percentage correct scores increased with age. The results of the two-way mixed ANOVA showed there was a significant main effect of type of measure; the difference between PCC-R and PVC was significant, $F(1, 1489) = 779.54$, $p < .001$, effect size or partial $\eta^2 = .34$, with PVC being systematically higher than PCC-R. There was also a significant main effect of age group on the percentage of correct productions ($F(1, 13) = 94.83$, $p < .001$, effect size or partial $\eta^2 = .45$). In addition, there was a significant interaction between 'type of measure' and 'age group' ($F(13, 1489) = 34.89$, $p < .001$, effect size or partial $\eta^2 = .23$). Descriptive statistics demonstrated that the difference between

PCC-R and PVC was larger for the younger age groups than it was for the older age groups.

Table 4. Percentage of consonants correct-revised and percentage of vowels correct by age group.

Age group (years;month)	***n***	**PCC-R**	***SD***	**PVC**	***SD***
2;0-2;3	72	76.3	12.8	87.5	9.71
2;4-2;7	101	80.9	12.8	89.2	8.10
2;8-2;11	101	89.0	7.38	93.3	4.96
3;0-3;3	102	91.5	6.05	95.1	4.15
3;4-3;7	107	91.7	5.71	95.3	3.83
3;8-3;11	101	92.6	5.48	96.5	3.49
4;0-4;3	99	94.5	5.25	96.8	4.13
4;4-4;7	111	96.0	3.18	97.7	2.87
4;8-4;11	112	96.2	2.85	98.0	2.24
5;0-5;3	117	95.7	3.91	97.7	3.09
5;4-5;7	128	96.3	5.19	97.6	5.52
5;8-5;11	116	97.3	3.05	98.5	2.41
6;0-6;5	117	97.1	3.01	98.4	2.33
6;6-6;11	119	97.6	2.19	98.6	1.78

Note. PCC-R = Percentage of consonants correct-revised; PVC = Percentage of vowels correct

Phonetic inventory

Table 5 summarizes the phonetic inventory of each age group. All vowels were acquired before the age of 3;4 years. All short vowels and most of the long vowels (except /e/), and the diphthongs (except /ɑu/) were acquired at age 2;7 years. By the age of 3;7 years, 75% of the children were able to produce all the syllable-initial consonants ≥ 75% of the cases correctly, except for the voiced fricatives /v/ and /z/ and the liquid /r/. All final consonants were acquired before the age of 4;4 years.

Table 5. Phonetic inventory (≥75% of the children produce the sound correctly).

Age group	*n*	Consonants										Vowels			
		Syllable initial					Syllable final								
		Plosives	Fricatives	Nasals	Liquids	Glides	Plosives	Fricatives	Nasals	Liquids	Glides	Short	Long	Reduced	Diphthongs
2;0-2;3	72	/b, t/		/m, n/				/f, s/	/m/	/l/		/ɪ, ɛ, ɔ, ʉ, ɑ/	/y, u, o, a/		/ʉy, ɛɪ/
2;4-2;7	101		/s/			/j/	/p/				/w/		/i, ø/		
2;8-2;11	101	/p, d, k/	/f, ʃ, h/			/w/	/t, k/	/ʃ/	/n/					/ə/	/ɑu/
3;0-3;3	102		/ʒ, x/										/e/		
3;4-3;7	107				/l/			/x/							
3;8-3;11	101									/r/					
4;0-4;3	99								/ŋ/						
4;4-4;7	111		/v/		/r/										
4;8-4;11	112														
5;0-5;3	117														
5;4-5;7	128		/z/												
5;8-5;11	116														
6;0-6;5	117														
6;6-6;11	119														

Table 6. Percentages of children per age group who acquired the degrees of complexity.

Degrees of Complexity	Segments	Age groups													
		2;0-2;3	2;4-2;7	2;8-2;11	3;0-3;3	3;4-3;7	3;8-3;11	4;0-4;3	4;4-4;7	4;8-4;11	5;0-5;3	5;4-5;7	5;8-5;11	6;0-6;5	6;6-6;11
Degree 1	/p/, /t/, /m/, /j/, /n/	83.3	86.1	99.0	99.0	99.1	100	100	100	100	100	100	100	100	100
Degree 2	/k/	50.0	64.9	82.0	87.1	87.6	94.0	94.9	96.4	100	97.4	97.7	99.1	98.3	99.2
Degree 3	/s/, /x/, /h/	62.0	80.0	93.1	94.1	96.3	99.0	96.0	100	99.1	98.3	96.9	99.1	100	100
Degree 4	/b/, /f/, /w/	63.9	74.0	83.2	91.1	93.5	97.0	97.0	99.1	99.1	100	98.4	100	100	100
Degree 5	/l/, /r/	20.0	32.7	48.0	60.0	70.1	79.0	88.9	86.5	94.6	92.3	95.3	96.6	99.1	99.2

Note. Grey cells indicate that a degree is acquired in an age group; the syllable-initial consonants of a degree were produced ≥ 75% correct on average by at least 75% of the children

Degrees of complexity

Table 6 shows the phonemic feature hierarchy in terms of the percentages of the occurrence of the various degrees of complexity across the age groups. The results indicate that the syllable-initial consonants /p/, /t/, /m/, /j/ and /n/ of degree 1 were produced correctly at the age of 2;0 years. The children aged 2;8 years were able to produce the dorsal consonant /k/ correctly. At the age of 2;4 years, the continuants /s/, /x/ and /h/ had been acquired, and at age 2;8 years the consonants /b/, /f/ and /w/, with those of degree 5 being acquired at 3;8 years of age. This order of acquisition confirmed that the older children in our sample used more phonological contrasts than the younger children, thereby corroborating Beers' complexity model.

Syllable-structure inventory

The results of the syllable-structure inventory are shown in Table 7. All two-year-old children had acquired the simple syllable structures CVC, CV, and V, and the more complex structures with an initial or final consonant cluster of two consonants by all 3-year-olds. Children in the 4;4-4;7 age group had acquired the syllable structure with an initial consonant cluster of three consonants (CCCVC), while the CCVCC structure was not acquired until after the age of 6;11.

Table 7. Syllable structure inventory (>75% of the children produce the syllable structure correctly).

Age group (years;month)	Correctly produced syllable structures (75% criterion)
2;0-2;3	CV, CVC
2;4-2;7	
2;8-2;11	V
3;0-3;3	CCV, CCVC
3;4-3;7	
3;8-3;11	CVCC
4;0-4;3	
4;4-4;7	CCCVC
4;8-4;11	
5;0-5;3	
5;4-5;7	
5;8-5;11	
6;0-6;5	
6;6-6;11	
>7;0	CCVCC

Table 8. Percentages of children per age group who use the phonological processes at least 10%.

Phonological processes	Age groups													
	2;0-2;3	**2;4-2;7**	**2;8-2;11**	**3;0-3;3**	**3;4-3;7**	**3;8-3;11**	**4;0-4;3**	**4;4-4;7**	**4;8-4;11**	**5;0-5;3**	**5;4-5;7**	**5;8-5;11**	**6;0-6;5**	**6;6-6;11**
Simplification processes														
Fronting	47.9	34.0	37.6	19.8	24.3	10.9	7.1	7.2	2.7	7.7	0.8	2.6	1.7	0.0
Stopping of fricatives	35.2	13.9	9.9	4.0	1.9	4.0	2.0	0.0	0.0	0.9	0.0	0.0	0.0	0.0
Voicing	6.9	3.0	1.0	0.0	0.9	0.0	0.0	0.0	0.0	0.9	0.0	0.0	0.0	0.0
Devoicing	45.8	32.0	18.8	8.9	10.3	14.9	11.1	4.5	8.0	6.0	2.3	3.4	3.4	0.8
Gliding	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Clusred 2to1 ini	90.0	75.5	51.0	37.0	29.2	19.8	11.1	8.1	4.5	9.4	7.0	2.6	0.9	0.0
Clusred 3to1 ini	60.9	38.2	24.2	9.2	12.4	9.1	1.0	0.9	2.7	2.6	2.4	2.6	1.7	0.8
Clusred 3to2 ini	57.8	59.6	61.1	38.8	40.0	32.3	26.8	13.6	19.6	17.9	11.0	9.5	11.2	14.3
Clusred 2to1 final	94.1	86.7	78.6	73.0	70.8	71.3	51.5	52.3	52.7	53.8	41.4	38.8	39.3	44.5
Unusual processes														
Backing	4.2	6.9	1.0	1.0	0.9	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Unusual stopping	16.9	9.0	2.0	1.0	0.0	1.0	0.0	0.0	0.9	0.0	0.0	0.0	0.0	0.0
Nasalisation	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Denasalisation	14.1	14.1	6.9	6.9	0.9	2.0	2.0	0.9	3.6	1.7	1.6	0.9	1.7	0.0
Hsation	2.8	3.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Dentalisation	19.4	11.9	13.9	5.9	3.7	3.0	2.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Lateralisation	1.4	0.0	2.0	2.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

Note. Grey cells indicate the process is present in the particular age group, that is, reaches the criterion of at least 10% occurrence in at least 10% of the participants; Clusred = cluster reduction.

Phonological processes

The phonological processes that were observed in our normative sample are presented in Table 8. Most phonological processes are resolved after 4;3 years, except initial cluster reduction from 3 to 2 consonants, e.g. [stɪk] for [strɪk] 'bow' and final cluster reduction from 2 to 1 consonant, as in [kɑs] for [kɑst] 'closet'. Backing (e.g. [kɔŋ] for [tɔŋ] 'tongue'), nasalisation (e.g. [nɪp] for [wɪp] 'seesaw'), voicing (e.g. [zɔk] for [sɔk] 'sock'), gliding (e.g. [bjuk] for [bruk] 'pants'), h-sation (consonants are replaced by /h/, e.g. [hɛɪn] for [trɛɪn] 'train') and lateralisation (e.g. [lɑs] for [jɑs] 'coat') did not occur in the normative sample.

Discussion

This cross-sectional study provides in-depth information on the typical speech sound development of Dutch-speaking children aged between 2;0 and 6;11 years in terms of PCC-R and PVC, the age of acquisition of consonants and vowels, while describing age-specific syllabic structure inventories, degrees of complexity (phonemic feature hierarchy), and phonological processes.

PCC-R and PVC

Consonant accuracy (PCC-R) and vowel accuracy (PVC) significantly increased with age, demonstrating a gradual progress in the children's ability to speak the Dutch language adequately. Between the ages of 2;0 and 2;3 years, the children in our sample produced consonants with a 76.4% accuracy, while the PCC-R of the children aged 6;6 to 6;11 was 97.6%. PVC scores increased from 87.5% in the youngest to 98.6% in the oldest age group. These results are broadly comparable with the PCC and PVC findings of studies evaluating other languages (Clausen and Fox-Boyer; 2017, Gangji et al., 2015; Grech and Dodd, 2008; MacLeod et al., 2011; Maphalala et al., 2014), although the comparison is not conclusive because some of the other studies used PCC instead of PCC-R. When calculating PCC-R, both common and uncommon clinical consonant distortions are scored as correct (Shriberg et al., 1997), which results in higher scores. We found no studies that used both measures.

The PVC scores were significantly higher than the PCC-R scores, which is also typical for other languages (PVC versus PCC) (Clausen and Fox-Boyer, 2017; Dodd et al., 2003; Pascoe et al., 2018). This was expected since the phenomenon is explained by the phonetic difference between vowels and consonants, where the production of the latter sounds, and especially consonant clusters, requires more precise speech motor skills than does the production of vowels. Furthermore, even though speakers may show variation in the speech production of a specific vowel, the acoustic output of that vowel is still recognized as the same vowel (Johnson et al., 1993). As a result, the judgment of vowels is less strict than that of consonants (Howard and Heselwood, 2012).

Phonetic inventory

The phonetic inventories supported the PCC-R and PVC findings in that, as expected, the older children were able to produce more vowels and consonants correctly than their younger counterparts. All the children aged 3;4 years had acquired a complete vowel inventory. Similar results were found for the English language (Dodd et al., 2003). The consonant inventory was almost complete at age 3;7 years for the syllable-initial consonants, except for the voiced fricatives /v/ and /z/, and the liquid /r/. All syllable-final consonants were acquired before the age of 4;4 years, which is comparable with the results Stes (1977) and Priester and Goorhuis-Brouwer (2013) reported and the findings for other languages. For example, the consonant /r/ is one of the latest acquired consonants in English-speaking children (Dodd et al., 2003) and in children speaking Swahili (Gangji et al., 2015).

Like in most languages (McLeod and Crowe, 2018), nasals, plosives, and glides in syllable-initial position were acquired earlier than syllable-initial liquids and some fricatives. In syllable-final position, plosives and glides were acquired before fricatives, liquids, and nasals. All short vowels had been acquired at the age of 2;3 years, earlier than most long vowels, the reduced vowel /ə/, and the diphthong /ɑu/.

The order of acquisition in which consonants were learned is broadly comparable with what Beers (1995) described, provided that in her study all syllable-initial consonants were acquired before the age of 3;0 years. Curiously, she does not mention the age of acquisition of the consonants /v/ and /z/ . We found that, in syllable-initial position, these two consonants were not acquired until 4;3 years of age (4;4 and 5;4 years, respectively). The difference in the age of acquisition Beers and we recorded may be due to the different methods of speech elicitation that were used. In her 1995 study, Beers analyzed spontaneous speech samples, which, as alluded to above, carries the risk that children avoid phonetic contexts that they have (more) difficulty with, 'choosing' the consonants that they can produce more easily and accurately. As the picture-naming task we used includes all Dutch phonemes, the children in our sample were made to produce a wider range of consonants, which inevitably elicits less accurate utterances. Note that the acquisition criterion is based on the proportion of correct productions, not on the total number of productions. This avoidance of difficult phonemes in spontaneous speech may then also be one of the explanations why Beers does not report on the production of /v/ and /z/. Alternatively or additionally, dialect variation may have played a role. In the Western part of the Netherlands the voiced consonants /v/ and /z/ are often pronounced as the voiceless consonants /f/ and /s/ and the children in the study of Beers (1995) all lived in the Central Western part of the Netherlands, where voiced fricatives tend to be devoiced. The children we tested resided in all four regions of our country, making our sample more representative of the general Dutch population in terms of geographic range.

Degrees of complexity

As to the distinctive features in typical Dutch speech sound development, our results pertaining to the degrees of complexity broadly confirmed the order of acquisition Beers (1995) had observed, with the exception of the 'dorsal' contrast, which in our study was acquired after the 'continuant' contrast. We noted that all contrasts were produced correctly at 3;8 years of age, whereas Beers (1995) concluded that most were mastered at the younger age of 2;9 years. Again, this disparity in the age of acquisition may be due to Beers' use of spontaneous speech rather than a naming task, with the children in her study possibly selecting the consonants in contexts that they were most comfortable with, while we confronted the children in our sample with a fixed set of words in varying contexts.

Syllable structure inventory

All syllable structures were acquired at the age of 4;7 years, except for the CCVCC sequence, which had not yet been acquired at 6;11 years of age. The simple structures, such as CV, CVC, and V were established first, followed by the syllables with an initial or final consonant cluster of two consonants (CCV, CCVC, CVCC), with those with an initial consonant cluster of three consonants (CCCVC) being acquired last. Syllable structures with initial clusters were established before those with final clusters, which closely resembles the order of acquisition reported in previous studies on the acquisition of Dutch (Van den Berg et al., 2017, Fikkert, 1994, Levelt et al., 2000) and other languages (Gangji et al., 2015, Mahura and Pascoe, 2016).

Phonological processes

As expected, we observed more phonological simplification processes in the children in the younger age groups. By the age of 4;4 years, all simplification processes had disappeared, except for the initial cluster reduction from three to two consonants (14.3%) and the final cluster reduction from two to one consonant (44.5%). These results are consistent with Dodd et al. (2003), who reported that in English-speaking children most phonological processes were resolved by 4;0 years and comparable with the findings in other languages (Clausen and Fox-Boyer, 2017; Pascoe et al., 2018). In our study, of all phonological processes, cluster reduction was present the longest, which, again, is in line with other studies in other languages (Aalto et al., 2019; Pascoe et al., 2018).

Besides simplification processes, we studied the use of unusual phonological processes, systematic speech errors that do not usually occur during typical development and are considered to indicate deviant development. Most of the unusual processes Beers (1995) had noted in her sample of typically developing children (i.e. backing, nasalization, H-sation, and lateralization) did not occur in our sample. We did, however, observe stopping of non-fricatives, denasalization, and dentalization in a small number of children in the youngest age groups (up to the age of 3;0 years).

Surprisingly, we found no evidence of 'gliding'. Beers (1995) described this substitution process as one of the most frequently occurring phonological processes in typically developing Dutch-speaking children, which is commonly used until the age of 4;0 years, similar to trends found in other languages like British English and South-African English (Dodd et al., 2003; Pascoe et al., 2018). Gliding occurs when the liquids /l/ and /r/ are replaced by the glides /j/ or /w/. In our data, the /l/ and /r/ are two of the latest consonants acquired, that is, not until the ages of 3;7 and 4;7, respectively. The glides /j/ and /w/ are acquired at a far younger age, i.e. at age 2;7 and 2;11 years, respectively. Possibly, the children in our study omitted these consonants more than they substituted them.

Limitations

In order to be able to compare narrow age ranges (14 age groups), we needed as large a sample as possible (*n* = 1,503), which is why we opted for a cross-sectional design. For most sounds, a monotonous increase in accuracy with age was found, confirming the reliability and validity of accuracy as an indicator of speech development, with only minor discontinuities of just a few percentage points occurring for most sounds. We chose to define the age of acquisition as the first age category at which 75% of the children produced a sound correctly 75% of the time. For two sounds, the /x/ and the /r/, these discontinuities led to uncertainty in determining the age of acquisition. For example, applying the 75% criterion consistently, the syllable initial consonant /x/ was found to have been acquired at age 3;0-3;3, but not in the 3;4-3;7 age group, and then again in the children aged 3;8-3;11 years. With the /r/ sound, the score of the 5;0-5;3-year-olds posed a problem, being substantially below 75%, whereas two younger age-groups scored well above this threshold. We hence chose to take the youngest age category in which the 75% criterion was reached as our reference for the classification of typical development in such cases, thereby taking into account the possible variability in speech production during a transitional period as Sosa (2015) suggested.

Young children with typically developing speech show sometimes distortions of sounds (Shriberg et al., 1997) that reflect an imprecise production of targeted sounds (e.g. dentalization or lateralization of the /s/, or labialization of the /r/) but with a correct phoneme selection. However, in words or in context, it cannot be distinguished whether distortions are of a phonetic or a phonological origin (Namasivayam et al., 2020). Despite providing a detailed description of speech sound development, we did not record systematic distortions (e.g. lisps). The distortion (e.g. the lisp) itself cannot be diagnosed with the CAI. However, with respect to all other aspects of speech sound development a child with a lisp can be compared to the norms. Our norms are suitable for these children, but not for diagnosing the distortion per se. In ongoing and planned research of the CAI software, we will focus on the development of rules to support the analysis of sound-by-sound contextual speech error patterns in word naming and conversational or spontaneous speech.

A final limitation we need to mention is that all results were based on analyses at the syllable level, which, among other restrictions, implies that weak syllable deletion was not considered. Possible effects of word length – expressed as the number of syllables – could therefore not be assessed. Since previous studies did report word-length effects, finding that children's speech production was less accurate for long words than it was for short words (Gangji et al., 2015, Maphalala et al., 2014, Vance et al., 2005), we will be adding word length and word structure as features for analysis to the next version of the CAI.

Clinical implications

No previous studies reported PCC-R and PVC for typically developing Dutch-speaking children despite the fact that these measures are widely used to support the diagnosis of SSDs (McLeod and Crowe, 2018), where PCC-R is most relevant to determine the severity of involvement (Shriberg et al., 1997).

Providing normative data obtained in 1,503 typically developing Dutch-speaking children, our inventory may be of use to SLPs who work with children suspected of an SSD. The norm scores were derived from the items of the picture-naming task of the CAI (Maassen et al., 2019), whose psychometric properties were verified, with our earlier studies revealing sufficient interrater reliability, test-retest reliability, and construct validity (Van Haaften et al., 2019a), and supported known-group validity (Van Haaften et al., 2019b). The CAI has since been made available for use in Dutch clinical practice. Describing typical speech sound development in terms of PCC-R and PVC, consonant, vowel, and syllabic structure inventories, degrees of complexity (phonemic feature hierarchy), and phonological processes, our assessment provides Dutch SLPs with a baseline against which the speech of children can be compared to determine the presence of an SSD. Based on the normative data on typically occurring phonological processes, clinicians can determine whether a child's speech development is comparable to that of age peers or whether it is delayed or impaired. The picture-naming task of the CAI is a practical and efficient means to gain detailed information about a child's production of speech sounds with the norm scores aiding the decision whether a child is in need of speech language therapy services.

References

Aalto, E., Saaristo-Helin, K. & Stolt, S. (2019). Phonological development of Finnish speaking children at 3;6 and associations to previous and simultaneous lexical ability. *Clinical Linguistics & Phonetics*, *34*(7), 617-633.

Andersson, L. (2005). Determining the adequacy of tests of children's language. *Communication Disorders Quarterly*, *26*, 207-225.

Archer, J., Champion, T., Tyrone, M. E. & Walters, S. (2017). Phonological Development of Monolingual Haitian Creole–Speaking Preschool Children. *Communication Disorders Quarterly*, *39*(3), 426-437.

Beers, M. (1995). *The Phonology of Normally Developing and Language-Impaired Children.* University of Amsterdam.

Clausen, M. C. & Fox-Boyer, A. (2017). Phonological development of Danish-speaking children: A normative cross-sectional study. *Clinical Linguistics & Phonetics*, *31*, 440-458.

Collins, B. & Mees, I. (2003). *The phonetics of English and Dutch.* Leiden, Brill.

Dodd, B., Holm, A., Hua, Z. & Crosbie, S. (2003). Phonological development: a normative study of British English speaking children. *Clinical Linguistics & Phonetics*, *17*, 617-643.

Dodd, B. (2011). Differentiating speech delay from disorder: does it matter? *Topics in Language Disorders*, *31*, 96-111

Fabiano-Smith, L. (2019). Standardized tests and the diagnosis of speech sound disorders. *Perspectives of the ASHA Special Interest Groups*, *4*, 58-66.

Fernhout, E., Permentier, L., van der Sijs, N., Cornips, L. & Nortier, J. (2011). Tweetalig, meer talig. Over mensen die meer talen gebruiken. *Taalpeil*, 1-8.

Fikkert, P. (1994). *On the acquisition of prosodic structure.* University of Leiden.

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 & communication disorders*, *50*, 151-164.

Grech, H. & Dodd, B. (2008). Phonological acquisition in Malta: A bilingual language learning context. *International Journal of Bilingualism*, *12*, 155-171.

Howard, S. & Heselwood, B. (2012). The contribution of phonetics to the study of vowel development and disorders. *In:* Ball, M., Damico, J., Ball, M. & Gibbon, F. (eds.) *Handbook of Vowels and Vowel Disorders.* Abingdon: Routledge.

Hua, Z. & Dodd, B. (2000). The phonological acquisition of Putonghua (modern standard Chinese). *Journal of child language*, *27*, 3-42.

Ingram, D. & Ingram, K. D. (2001). A whole-word approach to phonological analysis and intervention. *Language Speech and Hearing Services in Schools*, *32*, 271-283.

Johnson, K., Ladefoged, P. & Lindau, M. (1993). Individual differences in vowel production. *The Journal of the Acoustical Society of America*, *94*, 701-714.

Jongstra, W. (2003). *Variation in reduction strategies of Dutch word-initial consonant clusters.* University of Toronto.

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*, 14-29.

Levelt, C. (1994). *On the acquisition of place.* University of Leiden.

Levelt, C., Schiller, N. & Levelt, W. (2000). The acquisition of syllable types. *Language acquisition*, *8*, 237-264.

Maassen, B. & Terband, H. (2015). Process-oriented diagnosis of childhood and adult apraxia of speech (CAS and AOS). *In:* Redford, M. A. (ed.) *The Handbook of Speech Production.* Wiley.

Maassen, B., van Haaften, L., Diepeveen, S., Terband, H., van den Engel-Hoek, L., Veenker, T. & de Swart, B. (2019). *Computer Articulation Instrument*, Amsterdam, Boom Uitgevers.

MacLeod, A. A. N., Sutton, A., Trudeau, N. & Thordardottir, E. (2011). The acquisition of consonants in Québécois French: A cross-sectional study of pre-school aged children. *International Journal of Speech Language Pathology*, *13*, 93-109.

Mahura, O. O. & Pascoe, M. (2016). The acquisition of Setswana segmental phonology in children aged 3.0–6.0 years: A cross-sectional study. *International Journal of Speech Language Pathology*, *18*, 533-549.

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 & Phonetics*, *28*, 176-194.

Masso, S., McLeod, S. & Baker, E. (2018). Tutorial: assessment and analysis of polysyllables in young children. *Language, Speech, and Hearing Services in Schools*, *49*, 42-58.

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*, 229-241.

McLeod, S., van Doorn, J. & Reed, V. A. (2001). Normal acquisition of consonant clusters. *American Journal of Speech Language Pathology*, *10*, 99-110.

McLeod, S. & Crowe, K. (2018). Children’s Consonant Acquisition in 27 Languages: A Cross-Linguistic Review. *American Journal of Speech Language Pathology*, *27*, 1546-1571.

Mennen, I., Levelt, C. & Gerrits, E. (2006). Acquisition of Dutch phonology: an overview. *In:* Scobbie, J. M., Mennen, I. & Watson, J. (eds.) *QMUC Speech Science Research Centre Working Paper WP10.*

Morrison, J. A. & Shriberg, L. D. (1992). Articulation testing versus conversational speech sampling. *Journal of Speech, Language, and Hearing Research*, *35*, 259-273.

Namasivayam, A. K., Coleman, D., O'dwyer, A. & van Lieshout, P. (2020). Speech Sound Disorders in Children: An Articulatory Phonology Perspective. *Frontiers in Psychology*, *10*, 1-22.

Oller, D. & Delgado, R. (2000). Logical International Phonetics Program (Version 2.02). Miami, Florida: Intelligent Hearing Systems.

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 & Phonetics*, *32*, 1145-1161.

Phoon, H. S., Abdullah, A. C., Lee, L. W. & Murugaiah, P. (2014). Consonant acquisition in the Malay language: A cross-sectional study of preschool aged malay children. *Clinical Linguistics & Phonetics*, *28*, 329-345.

Priester, G. H. & Goorhuis-Brouwer, S. M. (2013). Measuring speech sound development: An item response model approach. *International Journal of Pediatric Otorhinolaryngology*, *77*, 1469-1473.

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*, 592-596.

Rys, K., van der Meulen, M., Hinskens, F., van der Gucht, F., de Caluwe, J., Heeringa, W. & van der Peet, M. (2017). Staat van het Nederlands. Nederlands Taalunie, Meertens Instituut and Universiteit Gent.

Shriberg, L. D., Austin, D., Lewis, B. A., McSweeny, J. L. & Wilson, D. L. (1997). The percentage of consonants correct (pcc) metric: extensions and reliability data. *Journal of Speech, Language, and Hearing Research*, *40*, 708-722.

Smit, A.B. (1993). Phonological error distributions in the Iowa-Nebraska Articulation Norms Project: Word-initial consonant clusters. *Journal of Speech and Hearing Research, 36*(5), 931-947.

Sosa, A. V. (2015). Intraword variability in typical speech development. *American Journal of Speech Language Pathology*, *24*, 24-35.

Stes, R. (1977). Normatieve studie over articulatieontwikkeling bij Nederlands-sprekende kinderen tussen 3-8 jaar (Normative study on articulation development of Dutch speaking children between 3-8 years of age). *Tijdschrift voor Logopedie en Audiologie*, *7*, 140-149.

Tresoldi, M., Barillari, M. R., Ambrogi, F., Sai, E., Barillari, U., Tozzi, E., Scarponi, L. & Schindler, A. (2018). Normative and validation data of an articulation test for Italian-speaking children. *International journal of pediatric otorhinolaryngology*, *110*, 81-86.

Van den Berg, R., van Severen, L., Molemans, I. & Gillis, S. (2017). Syllable type development in toddlers acquiring Dutch. *Dutch Journal of Applied Linguistics*, *6*, 119-146.

Van Haaften, L., Diepeveen, S., van den Engel-Hoek, L., Jonker, M., de Swart, B. & Maassen, B. (2019a). The psychometric evaluation of a speech production test battery for children: the reliability and validity of the computer articulation instrument. *Journal of Speech, Language and Hearing Research*, *62*, 2141-2170.

Van Haaften, L., Diepeveen, S., Terband, H., Vermeij, B., van den Engel-Hoek, L., de Swart, B. & Maassen, B. (2019b). Profiling speech sound disorders for clinical validation of the computer articulation instrument. *American Journal of Speech Language Pathology*, *28*, 844-856.

Vance, M., Stackhouse, J. & Wells, B. (2005). Speech production skills in children aged 3–7 years. *International Journal of Language & Communication Disorders*, *40*, 29-48.

Winitz, H. (1969). *Articulatory acquisition and behaviour*, New York, Appleton-Century-drofts.

Wolk, L. & Meisler, A. W. (1998). Phonological assessment: A systematic comparison of conversation and picture naming. *Journal of Communication Disorders*, *31*, 291-313.

Appendix A

The words elicited in the picture-naming task of the Computer Articulation Instrument (CAI)

No.	Item (English translation)	IPA transcription	No.	Item (English translation)	IPA transcription
1	auto *(car)*	/ɑu-to/	31	strik *(bow)*	/strɪk/
2	bal *(ball)*	/bɑl/	32	snoepje *(candy)*	/snup/
3	bloem *(flower)*	/blum/	33	trein *(train)*	/trɛin/
4	fiets *(bicycle)*	/fits/	34	vis *(fish)*	/vɪs/
5	stuur *(steering wheel)*	/styr/	35	water *(water)*	/wa-tər/
6	wiel *(wheel)*	/wil/	36	bus *(bus)*	/bʉs/
7	flesje *(bottle)*	/flɛʃ-ʃə/	37	wip *(seesaw)*	/wɪp/
8	fluit *(flute)*	/flʉyt/	38	zeep *(soap)*	/zep/
9	gieter *(watering can)*	/xi-tər/	39	zon *(sun)*	/zɔn/
10	nat *(wet)*	/nɑt/	40	klok *(clock)*	/klɔk/
11	haan *(rooster)*	/han/	41	lepel *(spoon)*	/le-pəl/
12	kip *(chicken)*	/kɪp/	42	mes *(knife)*	/mɛs/
13	huis *(house)*	/hʉys/	43	pop *(doll)*	/pɔp/
14	deur *(door)*	/dør/	44	ring *(ring)*	/rɪŋ/
15	raam *(window)*	/ram/	45	spin *(spider)*	/spɪn/
16	meisje *(girl)*	/mɛiʃ-ʃə/	46	televisie *(television)*	/te-lə-vi-si/
17	broek *(pants)*	/bruk/	47	knoop *(button)*	/knop/
18	jongen *(boy)*	/jɔŋ-ŋən/	48	man *(man)*	/mɑn/
19	jas *(coat)*	/jɑs/	49	lamp *(lamp)*	/lɑmp/
20	springtouw *(jump rope)*	/sprɪŋ-tɑuw/	50	dak *(roof)*	/dɑk/
21	jurk *(dress)*	/jʉr-rək/	51	gordijn *(curtain)*	/xɔr-dɛin/
22	sleutel *(key)*	/slø-təl/	52	giraf *(giraffe)*	/ʒi-rɑf/
23	schaar *(scissors)*	/sxar/	53	vrachtwagen *(truck)*	/vrɑxt-wa-xən/
24	sok *(sock)*	/sɔk/	54	kleurpotlood *(crayon)*	/klør-pɔt-lot/
25	speld *(pin)*	/spɛlt/	55	olifant *(elephant)*	/o-li-fɑnt/
26	neus *(nose)*	/nøs/	56	kapstok *(coat rack)*	/kɑp-stɔk/
27	tong *(tongue)*	/tɔŋ/	57	vliegtuig *(airplane)*	/vlix-tʉyx/
28	kast *(closet)*	/kɑst/	58	viltstift *(felt-tip pen)*	/vɪlt-stɪft/
29	stoel *(chair)*	/stul/	59	paraplu *(umbrella)*	/pa-ra-ply/
30	strijkijzer *(iron)*	/strɛik-ɛi-zər/	60	telefoon *(telephone)*	/te-lə-fon/

Appendix B

Frequency distributions of the phonological features in the picture-naming task

Class	Feature	Number of syllable-initial features
Consonants	p	4
	b	2
	t	9
	d	3
	k	3
	g	-
	ŋ	-
	m	3
	n	2
	l	6
	r	4
	f	6
	v	3
	s	5
	z	3
	ʃ	2
	ʒ	1
	j	3
	x	3
	h	2
	ʋ	4
Vowels	i	8
	y	2
	e	4
	ø	4
	a	7
	o	5
	u	4
	ɪ	9
	ɛ	3
	ɑ	11
	ʉ	2
	ə	12
	ɔ	9
Diphthongs	ɛi	5
	ɑu	2
	ʉy	3
Syllable structures	V	3
	CV	17
	VC	-
	CVC	40
	CCV	3
	CVCC	6
	CCVC	15
	CCVCC	3
	CCCVC	3
Initial consonant clusters	/vl-/, /vr-/, /fl-/, /bl-/, /br-/, /pl-/, /tr-/, /kl-/, /kn-/, /sn-/, /sp-/, /st-/, /sx-/, /sl-/, /spr-/, /str-/	
Final consonant clusters	/-ft/, /-xt/, /-lt/, /-mp/, /-nt/, /-rk/, /-ts/, /-st/	

3

FOUR

CHAPTER 4

MAXIMUM REPETITION RATE

IN A LARGE CROSS-SECTIONAL SAMPLE OF TYPICALLY DEVELOPING DUTCH-SPEAKING CHILDREN

Leenke van Haaften and Sanne Diepeveen, Hayo Terband, Bert de Swart,
Lenie van den Engel-Hoek and Ben Maassen

REVISED VERSION SUBMITTED FOR INTERNATIONAL JOURNAL OF SPEECH-LANGUAGE PATHOLOGY

Abstract

Purpose: The current study aims to provide normative data for the maximum repetition rate (MRR) development of Dutch-speaking children based on a large cross-sectional study using a standardised protocol.

Method: A group of 1014 typically developing children aged 3;0 to 6;11 years performed the MRR task of the Computer Articulation Instrument (CAI). The number of syllables per second was calculated for mono-, bi-, and trisyllabic sequences (MRR-pa, MRR-ta, MRR-ka, MRR-pata, MRR-taka, MRR-pataka). A two-way mixed ANOVA was conducted to compare the effects of age and gender on MRR scores in different MRR sequences.

Result: The data analysis showed that overall MRR scores were affected by age group, gender and MRR sequence. For all MRR sequences the MRR increased significantly with age. MRR-pa was the fastest sequence, followed by respectively MRR-ta, MRR-pata, MRR-taka, MRR-ka and MRR-pataka. Overall MRR scores were higher for boys than for girls, for all MRR sequences.

Conclusion: This study presents normative data of MRR of Dutch-speaking children aged 3;0 to 6;11 years. These norms might be useful in clinical practice to differentiate children with speech sound disorders from typically developing children. More research on this topic is necessary. It is also suggested to collect normative data for other individual languages, using the same protocol.

Introduction

Maximum repetition rate (MRR), or *diadochokinesis*, involves alternating motion rate tasks comprising speech like syllables (Kent, 2015). MRR is one of the most commonly used oral-motor assessments in clinical practice (Icht & Ben-David, 2014; Williams & Stackhouse, 2000). It is suggested as an important part of a test battery to differentiate between various speech disorders (Diepeveen, Van Haaften, Terband, De Swart, & Maassen, 2019; Maassen & Terband, 2015; Terband, Maassen, & Maas, 2019). However, there is also still a debate about the clinical value of the MRR. A higher-faster-farther approach might not be a good assessment because in speech speed is not a necessary skill (Ziegler et al., 2019). Although this is the case, MRR can play a role in diagnosing underlying articulomotor planning and programming problems (Maassen & Terband, 2015; Rvachew et al., 2005; Van Haaften, Diepeveen, Terband, et al.. 2019). MRR is therefore often used in the assessment of children with a suspicion of a motor speech disorder (MSD) and/or childhood apraxia of speech (CAS) (Murray, McCabe, Heard, & Ballard, 2015; Thoonen, Maassen, Gabreels, & Schreuder, 1999), and it has been used in the characterization of speech language phenotypes (e.g., Peter et al., 2017; Peter, Matsushita, & Raskind, 2012; Turner et al., 2015). To be able to interpret the results of the MRR adequately, it must be part of a set of speech tasks. By comparing the results of the MRR task with the results of other tasks (i.e. picture naming, nonword repetition) a complete speech profile can be obtained. The results of the MRR should not be used solely to diagnose children with speech sound disorders, because many children with SSD show similar behavioural symptoms in speech. The traditional way of diagnosing children with SSD might not be sufficient, because the different levels involved in speech influence each other (Namasivayam et al., 2020). The underlying processes involved in speech production are lemma access, word form selection, phonological encoding, speech motor planning and programming, and speech motor execution (Terband, Maassen, & Maas, 2019). Insight into the deficits that might be the underlying causes of an SSD, requires an extensive analysis of a child's performance on a range of speech tasks that reflect different underlying processes. A study of our research group (Van Haaften, Diepeveen, Van den Engel-Hoek et al., 2019) showed the distinctive function of four different speech tasks of a new speech production test battery for children: the Computer Articulation Instrument (CAI). The CAI contains the tasks picture naming, nonword imitation, word and nonword repetition and MRR. Factor analyses were conducted based on the assumption that clusters of selected parameters would reflect different aspects of speech production, either within or across tasks. Factor analyses revealed five meaningful factors: all picture-naming parameters (PN), the segmental parameters of nonword imitation (NWI-Seg), the syllabic structure parameters of nonword imitation (NWI-Syll), (non)word repetition consistency (PWV), and all MRR parameters. Each task reflects different aspects of speech production. Furthermore, the construct validity was underlined by the weak correlations between CAI factor scores, indicating the independent contribution of each factor to the speech profile. In another study with 41 children (age 3;0 to 6;4; 26 boys and 15 girls) with SSD data were collected from the four tasks of the CAI. The children were categorised in two groups, moderate or a severe SSD indicated by their speech language pathologist (SLP). Results indicated a significant difference between the two groups for picture naming, nonword imitation (segmental and syllable structure) and the bisyllabic and trisyllabic MRR factor (Van Haaften, Diepeveen, Terband et al., 2019). The findings of these two studies suggest that the MRR should be part of the diagnostic process. Normative data of MRR is essential to differentiate children with delayed or disordered speech development from typically developing children. The availability of these data is important for SLPs to make clinical decisions.

Several studies have investigated MRR in typically developing children. The overall conclusion, across languages, is that MRR increases with age. Contrasting results were found in studies investigating gender differences and differences between specific MRR sequences. Some studies found differences between boys and girls (Modolo, Berretin-Felix, Genaro, & Brasolotto, 2011) or between MRR sequences (Blech, 2010; Prathanee, Thanaviratananich, & Pongjanyakul, 2003), while other studies found no differences between gender (Fletcher, 1972; Icht & Ben-David, 2015; Wong, Allegro, Tirado, Chadha, & Campisi, 2011; Zamani, Rezai, & Garmatani, 2017) or MRR sequence (Rvachew, Ohberg, & Savage, 2006; Thoonen, Maassen, Wit, Gabreels, & Schreuder, 1996). However, considerable methodological differences exist between the studies, with different methods of data collection and different scoring methods of MRR. Several studies used a time-by-count procedure (the time needed to repeat a certain number of syllables) (Blech, 2010; Fletcher, 1972; Prathanee et al., 2003; Rvachew et al., 2006; Thoonen et al., 1999; Thoonen et al., 1996; Yaruss & Logan, 2002; Zamani et al., 2017), while in other studies a procedure of count-by-time was used (the number of syllables repeated in a certain amount of time) (Henry, 1990; Icht & Ben-David, 2015; Juste et al., 2012; Modolo et al., 2011; Robbins & Klee, 1987). Because of these methodological differences, the normative data is difficult to compare. To reduce these differences, a standardised protocol is proposed in a study by Diepeveen et al. (2019). In this protocol, it is suggested that MRR should not be assessed in children under the age of 3 years. The maximum age up to seven years has been chosen, because previous research has shown that speech sound development continues up to seven years (Priester and Goorhuis-Brouwer, 2013). Monosyllabic sequences and bi- and trisyllabic sequences should be described as separate outcome measures and if children cannot produce the monosyllabic sequences, the bi- and trisyllabic sequences should not be administered. Nonsense syllabic sequences are used instead of real words as MRR is supposed to measure motor speech abilities rather than linguistic skills (Williams & Stackhouse, 2000). The measurement procedure follows the time-by-count principle. The data indicates that children do not have to be encouraged to perform series of at least ten syllables, but that series of five syllables is sufficient for a reliable and valid calculation of the MRR (Diepeveen et al., 2019). After exclusion of the first and last syllable, the mean rate is then based on the duration of at least three syllables.

Most of the MRR studies in typically developing children are based on a small number of children and relatively limited age ranges (Blech, 2010; Prathanee et al., 2003; Rvachew et al., 2006; Thoonen et al., 1999; Thoonen et al., 1996; Wong et al., 2011; Yaruss & Logan, 2002). As typically developing children show progress in speech motor skills as they grow older, normative data is required for consecutive age groups. Therefore, the aim of the present study is to provide normative data for the MRR development of Dutch-speaking children aged 3;0 to 6;11 years based on a large cross-sectional study using the standardised protocol by Diepeveen et al. (2019). Differences between age groups, gender and MRR sequences are described.

Method

Participants

The 1014 participants of this study participated in a large normative study in the context of the development of a new speech production test battery in Dutch: the Computer Articulation Instrument (CAI; Maassen et al., 2019; Van Haaften, Diepeveen, Van den Engel-Hoek et al., 2019). The CAI consists of four tasks: (1) picture naming, (2) nonword imitation, (3) word and nonword repetition, and (4) maximum repetition rate (MRR) task. The data of the MRR task was used for the current study. Between January 2008 and April 2015, typically developing Dutch-speaking children aged between 2;0 and 7;0 were recruited via nurseries (n = 47) and mainstream primary schools (n = 71) in the Netherlands. Inclusion criteria were no hearing loss and Dutch being the spoken language at the nursery or primary school. The sample was representative for gender, geographic region and degree of urbanisation (Van Haaften, Diepeveen, Van den Engel-Hoek et al., 2019). The parents or caregivers were asked to fill out a questionnaire containing questions about hearing problems, speech and language development, developmental problems and whether the child is seen by an SLP. Children were excluded if they had developmental problems that could influence the speech performance. See Maassen et al. (2019) and Van Haaften, Diepeveen, Van den Engel-Hoek et al. (2019) for detailed information on sample characteristics and data collection. As Diepeveen et al. (2019) concluded that the MRR protocol of the CAI is applicable for children of 3 years and older, this study only used the data of children aged between 3;0 and 7;0, divided in 11 age groups. Table 1 shows the number of subjects per MRR sequence per age group and gender.

Table 1. Sample composition: numbers of children per age group, broken down by gender.

Age group (years;months)	Total number of children	M_{age}	Gender (*n*)	
			Boys	Girls
3;0-3;3	68	3;01	32	36
3;4-3;7	65	3;05	34	31
3;8-3;11	86	3;08	46	40
4;0-4;3	77	4;01	42	35
4;4-4;7	90	4;05	48	42
4;8-4;11	93	4;08	43	50
5;0-5;3	103	5;01	54	49
5;4-5;7	111	5;05	61	50
5;8-5;11	104	5;08	55	49
6;0-6;5	108	6;02	63	45
6;6-6;11	109	6;07	53	56
Grand total	1014		531	483
% sample	100		52.4	47.6

Ethical considerations

The research ethics committee of the Radboud University Nijmegen Medical Centre stated that this study does not fall within the remit of the Medical Research Involving Human Subjects Act (WMO). Therefore, this study can be carried out (in the Netherlands) without an approval by an accredited research ethics committee. The study was conducted according to the ethical principles and guidelines in the Netherlands. For example, informed consent was obtained from all parents or caregivers.

Procedure

In the CAI project 14 SLPs administrated the test for the younger children (2 to 4 years of age) and 110 SLP students (working in pairs) assessed the older children (4 to 7 years of age). All assessors were trained in the administration of the MRR task by the first two authors. The assessment took place at the child's nursery or primary school in a quiet room. The CAI was administered using a computer laptop and the acoustic signal (minimum of 44.1 Hz; 16 bits) was automatically stored on the computer's hard disk. The child and SLP or SLP student were seated side by side in front of the computer. Both wore a headset, or a speaker and microphone were used. Testing took approximately 30 minutes for all the tasks of the CAI. The administration of the MRR task took about five to ten minutes per child.

MRR administration

For the administration of the MRR task the CAI uses the protocol described by Diepeveen et al. (2019). This protocol was developed based on previous studies in the Dutch language (Thoonen et al., 1999; Thoonen et al., 1996; Wit, Maassen, Gabreels, & Thoonen, 1993). Instructions were given by the CAI computer program to maximise standardisation. During the task children are required to reproduce pre-recorded sequences on one single breath: first three monosyllabic sequences (/papa../, /tata../ and /kaka../), followed by one trisyllabic sequence (/pataka.../) and finally two bisyllabic sequences (/pata../ and /taka../). It was not possible to change the order of sequences; the computer program was fixed.

First, the children were asked to repeat a short sequence of three syllables (e.g. /papapa/) in a normal speaking rate after an audio model. Second, children were asked to repeat a longer sequence of six syllables in a normal rate (e.g. /papapapapapa/). The third instruction included imitation of a sequence of 12 syllables at a faster speech rate after an audio example. Finally, the children were asked to produce the syllable sequences as fast as possible, without an audio model. The CAI allows a maximum of three attempts per sequence.

MRR analysis

Six SLP students of HAN University of Applied Sciences and three SLPs analysed the mono-, tri- and bi-syllabic sequences according to the analysis protocol for calculating the MRR proposed by Diepeveen et al. (2019). They were trained by one of the first authors (SD) and practiced with one sample before analysing the other samples. Since the program stores all tasks and all trials of a child in one recording, the recordings were spliced into fragments per trial manually with Praat software, version 6.0.21 (Boersma & Weenink, 2016). First the administrator determined if the sequence was pronounced correctly. The sequence was correct when the syllables were pronounced fluently in succession and had no articulation errors, allowing for dialect variances. The test administrator analysed the attempts the child has produced upon the last two instructions, calculated the syllables per second and recorded this in the database. The audio-recordings, each containing just one attempt of one sequence, were analysed with the help of a customised Praat-script (developed by one of the authors; HT). The script detected and marked syllable onsets by localising the noise burst of the voiceless plosives. The first and the last syllable were excluded because speakers often produce the first syllable with a longer duration and higher intensity (Thoonen et al., 1996) and the last syllable is also often lengthened (Ackermann, Hertrich, & Hehr, 1995). Before extracting the number of syllables, syllable durations and MRR score, the marked syllable onsets were depicted in the waveform and inspected visually and any errors in the number of syllables indicated by the script were corrected manually. Figure 1 gives an example of one of the sequences with the markers. Only sequences with a remaining minimum of three syllables, after exclusion of the first and last syllable, were included in the analysis. In 30% of the cases, the script could not detect syllable onsets correctly. These samples were analysed manually to determine the number of syllables and the duration of the sequence; administrators used both visual examination of the waveform and playback of the audio recording. In a pilot study for our MRR-protocol, we studied the reliability (n = 126) between the computer script and the manually analysed recordings. The intraclass correlation coefficients (ICCs) were sufficient to good: /pa/ = .79; /ta/ = .90; /ka/ = .85; /pataka/ = .74; /pata/ = .79; /taka/ = .76. MRR score was calculated by dividing the number of syllables of the sequence by the duration of the sequence (syll/s). Eventually, number of syllables, duration time, and MRR score were merged in SPSS, version 24 for Windows (SPSS Inc., Chicago, IL, USA). The fastest correctly produced series of syllables, based on the number of syllables, is used for analysis.

Not all children completed all MRR sequences for reasons of shyness or inattentiveness. Furthermore, in some cases the audio files were damaged due to technical problems or background noise that prevented recognising the individual syllables. In this case, the recordings were excluded from the sample. Table 2 shows the number of children from whom an analysable MRR sequence was collected.

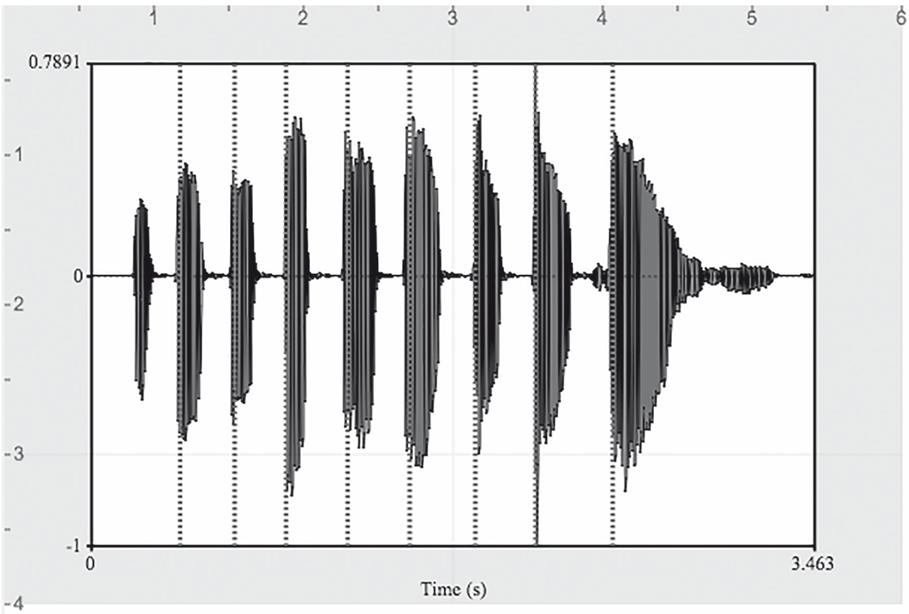

Figure 1. Example of the analysis with the Praat-script of one of the MRR sequences.

Reliability

Interrater and test-retest reliability of the MRR scores (syll/s) were examined and described by Van Haaften, Diepeveen, Van den Engel-Hoek et al. (2019). In this study, typically developing children aged between 2;0 and 7;0 were included. To measure interrater reliability the audio recordings of 103 children were randomly selected and scored by 33 raters. Their MRR scores were compared with those of one independent rater. A total of 107 children were randomly selected for the test-retest reliability study; these children were examined twice within three months by the same administrator. Two raters scored the audio recording of the initial test and retest, with the same rater scoring the tests of the same child. Interrater reliability, calculated with interclass correlation coefficient (ICC), was good for the monosyllabic sequences /pa/ (ICC 0.81) and /ka/ (ICC 0.83) and sufficient for /ta/ (ICC 0.77). The interrater reliability for the bisyllabic and trisyllabic items was insufficient, with ICCs ranging from 0.41 to 0.62. Especially the younger children (i.e., the 2- to 3-year-olds) had difficulties performing the bisyllabic and trisyllabic items, whereas a large number of children were not able to perform the task at all. The data of children who failed to perform the task were not included in the reliability study; had we included whether the attempts were successful or not, the ICC might have been higher. Test-retest reliability was sufficient for /pa/ (ICC 0.70) and insufficient for the other sequences, with ICCs ranging from 0.18 to 0.60. Reasons for these low scores could be the rapid development of the younger children during the interval between test and retest or a test-retest training effect. Based on these results, and the results of the study of Diepeveen et al. (2019), the younger children aged between 2;0 and 3;0 were not included in the current study. Further details and interpretations of the reliability study are discussed in Van Haaften, Diepeveen, Van den Engel-Hoek et al. (2019).

Statistical Analysis

To compare the effects of age and gender on MRR scores in different MRR sequences, and to test the hypotheses that there is a difference between the six MRR sequences and between boys and girls for the 11 age groups, a two-way mixed ANOVA was conducted. MRR score (syll/s) was the dependent variable, MRR sequence was the within-subject factor with six levels (MRR-pa, MRR-ta, MRR-ka, MRR-pataka, MRR-pata, MRR-taka), and there were two between-subject factors: age group (11 age groups) and gender (2 levels: boys and girls). Mauchly's test of Sphericity was conducted to test the hypothesis that the variances of differences between conditions are equal. Bonferroni correction was applied for post hoc comparisons. Statistical analyses were performed using SPSS version 20 for Windows (SPSS Inc., Chicago, IL, USA).

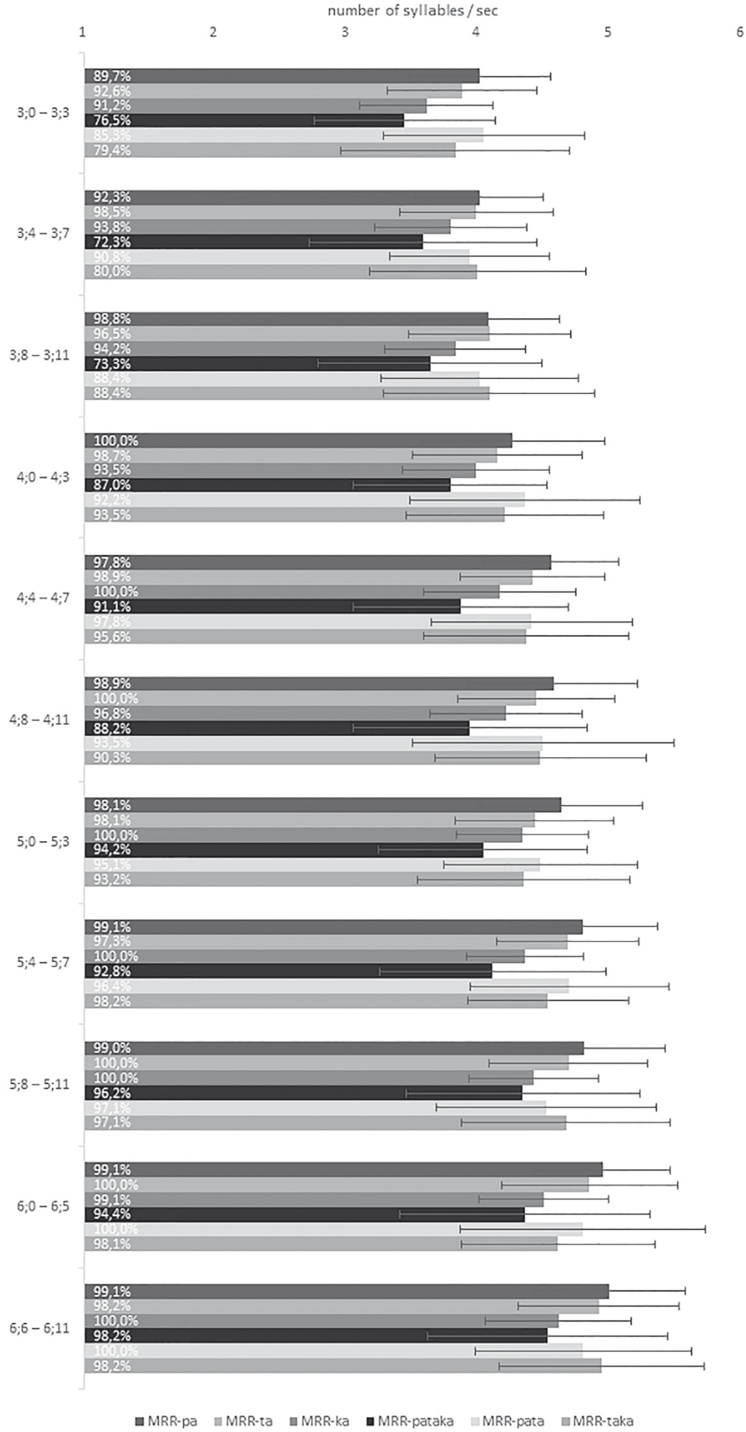

Figure 2. Mean number of syllables / second per age group and per sequence. The percentages of children able to perform the task (in relation to the total number of children of the respective age group) are shown at the beginning of the bars.

Results

The results of the mean number of syllables / second per age group and per sequence are presented in Figure 2. The percentage of children (in relation to the total number of children of the respective age group) who could perform the sequence correctly (fluently in succession; no articulation errors, allowing for dialect variances) is shown at the beginning of the bars.

The mean and standard deviations of each MRR sequence are depicted by age group and gender in Table 2, showing data of children who could perform all the six sequences correctly. Mauchly's test indicated that the assumption of sphericity was violated [$\chi 2$ (14) = 521.6, $p < .001$], therefore degrees of freedom were corrected using Huynh-Feldt estimates of sphericity ($\varepsilon = .85$).

The two way mixed ANOVA revealed a significant effect of the within-subject factor 'MRR sequence' ($F(4.24, 3382.89) = 100.16$, $p < .001$, effect size or partial $\eta^2 = .112$), which means that the MRR scores were significantly different for the MRR sequences. Post-hoc analyses showed that the difference between mean MRR scores was significant for most of the pairwise comparisons, but was not significant between MRR-ta and the bi-sylabic sequences MRR-pata ($p = 1.000$) and MRR-taka ($p = 1.000$), nor between MRR-pata and MRR-taka ($p = 1.000$). The fastest sequence is MRR-pa ($M = 4.64$, $SD = 0.64$) and the slowest sequence is MRR-pataka ($M = 4.07$, $SD = 0.90$), see Table 2. 4

The effect of between-subject factor 'age group' was also significant ($F(10, 798) = 29.96$, $p < .001$, effect size or partial $\eta^2 = .273$). The number of syllables per second increased with age for all MRR sequences. As shown in Table II, MRR sequences increased on average with 1.02 syllables per second from the youngest to the oldest age group.

The statistical analysis also yielded a significant effect of the between-subject factor 'gender' on overall MRR scores ($F(1, 798) = 9.49$, $p = .002$, effect size or partial $\eta^2 = .012$). As shown in Table II, MRR scores were higher for boys than for girls for all MRR sequences.

No significant interaction was found between 'MRR sequences' and 'age group' ($F(42.39, 3382.89) = 1.181$, $p = .196$, effect size or partial $\eta^2 = .015$), 'MRR sequences' and 'gender' ($F(4.24, 3382.89) = 2.172$, $p = .066$, effect size or partial $\eta^2 = .003$), 'age group' and 'gender' ($F(10, 798) = .876$, $p = .555$, effect size or partial $\eta^2 = .011$), or 'MRR sequences' and 'age group' and 'gender' ($F(42.39, 3382.89) = 1.069$, $p = .351$, effect size or partial $\eta^2 = .013$).

Table 2. Descriptive statistics (means and standard deviations) of the MRR score (syll/s) per age group and gender, broken down by MRR sequence.

Gender	Age group		MRR sequence					
			MRR-pa	MRR-ta	MRR-ka	MRR-pataka	MRR-pata	MRR-taka
Total	3;0 – 3;3	*n*	37	37	37	37	37	37
		M	3.95	3.91	3.66	3.40	4.01	3.81
		SD	0.59	0.56	0.46	0.55	0.88	0.78
	3;4 – 3;7	*n*	38	38	38	38	38	38
		M	4.06	4.06	3.76	3.54	3.99	4.08
		SD	0.50	0.51	0.57	0.83	0.60	0.82
	3;8 – 3;11	*n*	51	51	51	51	51	51
		M	4.15	4.11	3.84	3.74	4.03	4.07
		SD	0.52	0.67	0.53	0.87	0.79	0.83
	4;0 – 4;3	*n*	60	60	60	60	60	60
		M	4.27	4.17	4.00	3.82	4.35	4.25
		SD	0.57	0.61	0.54	0.73	0.90	0.78
	4;4 – 4;7	*n*	77	77	77	77	77	77
		M	4.59	4.40	4.14	3.88	4.41	4.38
		SD	0.51	0.57	0.54	0.82	0.76	0.74
	4;8 – 4;11	*n*	77	77	77	77	77	77
		M	4.55	4.42	4.20	3.93	4.49	4.47
		SD	0.67	0.62	0.56	0.90	0.97	0.83
	5;0 – 5;3	*n*	87	87	87	87	87	87
		M	4.64	4.40	4.33	4.04	4.49	4.36
		SD	0.54	0.59	0.48	0.79	0.70	0.84
	5;4 – 5;7	*n*	97	97	97	97	97	97
		M	4.82	4.69	4.37	4.14	4.68	4.53
		SD	0.55	0.54	0.46	0.83	0.72	0.57
	5;8 – 5;11	*n*	94	94	94	94	94	94
		M	4.83	4.70	4.45	4.35	4.55	4.70
		SD	0.62	0.62	0.47	0.89	0.84	0.80
	6;0 – 6;5	*n*	99	99	99	99	99	99
		M	4.96	4.87	4.48	4.37	4.86	4.64
		SD	0.51	0.66	0.49	0.96	0.91	0.72
	6;6 – 6;11	*n*	103	103	103	103	103	103
		M	5.03	4.92	4.63	4.51	4.80	4.96
		SD	0.56	0.59	0.56	0.86	0.83	0.78
	Total	*n*	820	820	820	820	820	820
		M	4.64	4.52	4.26	4.07	4.51	4.48
		SD	0.64	0.67	0.58	0.90	0.86	0.81
Boys	3;0 – 3;3	*n*	18	18	18	18	18	18
		M	3.95	3.86	3.63	3.28	4.14	3.57
		SD	0.56	0.62	0.52	0.68	1.06	0.78
	3;4 – 3;7	*n*	21	21	21	21	21	21
		M	4.24	4.18	3.87	3.58	4.23	4.24
		SD	0.48	0.47	0.66	0.64	0.53	0.84

	3;8 – 3;11	*n*	28	28	28	28	28	28
		M	4.27	4.22	3.90	3.90	4.14	4.21
		SD	0.45	0.76	0.53	1.00	0.82	0.93
	4;0 – 4;3	*n*	33	33	33	33	33	33
		M	4.36	4.31	4.03	4.00	4.52	4.19
		SD	0.51	0.66	0.60	0.73	0.96	0.89
	4;4 – 4;7	*n*	38	38	38	38	38	38
		M	4.64	4.39	4.29	3.83	4.45	4.35
		SD	0.49	0.59	0.57	0.77	0.92	0.73
	4;8 – 4;11	*n*	37	37	37	37	37	37
		M	4.51	4.51	4.18	3.94	4.50	4.46
		SD	0.75	0.58	0.64	1.03	1.03	0.95
	5;0 – 5;3	*n*	44	44	44	44	44	44
		M	4.68	4.49	4.34	4.04	4.65	4.44
		SD	0.59	0.71	0.47	0.77	0.71	0.97
	5;4 – 5;7	*n*	56	56	56	56	56	56
		M	4.80	4.68	4.30	4.26	4.66	4.48
		SD	0.55	0.57	0.47	0.94	0.74	0.57
	5;8 – 5;11	*n*	52	52	52	52	52	52
		M	4.90	4.76	4.46	4.39	4.55	4.69
		SD	0.72	0.62	0.53	0.90	0.80	0.84
	6;0 – 6;5	*n*	57	57	57	57	57	57
		M	4.94	4.96	4.55	4.43	4.92	4.71
		SD	0.50	0.72	0.5	1.11	0.95	0.80
	6;6 – 6;11	*n*	51	51	51	51	51	51
		M	5.21	4.98	4.62	4.53	4.98	5.02
		SD	0.63	0.59	0.59	0.86	0.92	0.83
	Total	*n*	435	435	435	435	435	435
		M	4.70	4.59	4.29	4.13	4.60	4.49
		SD	0.66	0.70	0.60	0.94	0.89	0.87
Girls	3;0 – 3;3	*n*	19	19	19	19	19	19
		M	3.95	3.97	3.69	3.51	3.89	4.03
		SD	0.63	0.49	0.40	0.38	0.69	0.72
	3;4 – 3;7	*n*	17	17	17	17	17	17
		M	3.84	3.91	3.61	3.49	3.69	3.88
		SD	0.44	0.54	0.41	1.04	0.55	0.77
	3;8 – 3;11	*n*	23	23	23	23	23	23
		M	4.02	3.98	3.75	3.54	3.90	3.89
		SD	0.57	0.54	0.53	0.65	0.75	0.67
	4;0 – 4;3	*n*	27	27	27	27	27	27
		M	4.17	3.97	3.97	3.61	4.15	4.32
		SD	0.63	0.51	0.46	0.68	0.81	0.62
	4;4 – 4;7	*n*	39	39	39	39	39	39
		M	4.54	4.41	4.00	3.92	4.36	4.42
		SD	0.52	0.56	0.47	0.88	0.57	0.76
	4;8 – 4;11	*n*	40	40	40	40	40	40
		M	4.59	4.34	4.22	3.92	4.48	4.48
		SD	0.60	0.65	0.49	0.79	0.92	0.71

5;0 – 5;3	*n*	43	43	43	43	43	43
	M	4.60	4.30	4.31	4.04	4.33	4.28
	SD	0.48	0.44	0.49	0.83	0.65	0.68
5;4 – 5;7	*n*	41	41	41	41	41	41
	M	4.85	4.69	4.47	3.98	4.72	4.59
	SD	0.54	0.51	0.43	0.63	0.70	0.56
5;8 – 5;11	*n*	42	42	42	42	42	42
	M	4.74	4.61	4.45	4.29	4.54	4.71
	SD	0.46	0.61	0.39	0.88	0.89	0.76
6;0 – 6;5	*n*	42	42	42	42	42	42
	M	4.99	4.74	4.38	4.30	4.79	4.54
	SD	0.52	0.57	0.43	0.71	0.86	0.61
6;6 – 6;11	*n*	52	52	52	52	52	52
	M	4.86	4.86	4.64	4.50	4.63	4.91
	SD	0.43	0.60	0.54	0.87	0.69	0.72
Total	*n*	385	385	385	385	385	385
	M	4.58	4.44	4.23	4.02	4.42	4.46
	SD	0.62	0.63	0.55	0.84	0.80	0.74

Note. n = number of children from whom an MRR sequence was analysed; *M* = mean of the MRR score (syll/s); *SD* = standard deviation of the mean MRR score (syll/s); MRR-pa = number of syllables per second of sequence /pa/; MRR-ta = number of syllables per second of sequence /ta/; MRR-ka = number of syllables per second of sequence /ka/; MRR-pataka = number of syllables per second of sequence /pataka/; MRR-pata = number of syllables per second of sequence /pata/; MRR-taka = number of syllables per second of sequence /taka/.

Discussion

This study presents normative data of MRR from a large population of Dutch-speaking children aged 3;0 to 6;11 years. Tight ranges of age groups were used to be able to examine the relationship between age and MRR score. A cross-sectional study was performed, using a standardised protocol (Diepeveen et al., 2019). This protocol was used for both the administration of the MRR task and the analysis of the MRR scores. Effects of age, MRR sequence and gender were investigated.

Effect of age on MRR scores

For all MRR sequences the number of syllables per second increased significantly and monotonously with age. No interaction was found between MRR sequence and age group. The MRR score of all sequences was about 1 syllable per second faster for the oldest age group when compared with the youngest age groups. These results are in accordance with the findings in previous studies (Henry, 1990; Icht & Ben-David, 2015; Juste et al., 2012; Modolo et al., 2011; Prathanee et al., 2003; Robbins & Klee, 1987; Zamani et al., 2017). Thus, MRR score increases with age, which is likely to be caused by maturation of the speech motor system (Kent, Kent, & Rosenbek, 1987). Our study included children from 3;0 to 6;11 years of age. Fletcher (1972) found an increase of MRR score in a study with 48 children between the ages of 6;0 and 13;0 years. Wong et al. (2011) demonstrated that MRR score still increases up to the age of 18 years. Between 18 and 60 years of age, Knuijt, Kalf, Van Engelen, Geurts, and de Swart (2019) found stable MRR scores, with a decrease in maximum number of syllables per second from 60 years of age. To conclude, the increase in MRR score seen in the current study in children aged 3 to 7 years is in line with the results of other studies in older children and with studies in adults.

Effect of MRR sequences on MRR scores

The present results show that at the group level typically developing children produce the monosyllabic sequence MRR-ta slower than MRR-pa, and MRR-ka was slower than MRR-pa and MRR-ta. This is in agreement to similar studies with children (Kent et al., 1987; Prathanee et al., 2003; Robbins & Klee, 1987; Rvachew et al., 2006; Thoonen et al., 1996) and adults (Knuijt et al., 2019; Padovani, Gielow, & Behlau, 2009). The production of velar sounds takes longer than the production of alveolar and lip sounds. This might be due to the involvement of physiological factors. The production of /ka/ requires movement of the tongue dorsum, which has a larger mass than the tongue tip, required for pronouncing /ta/; larger inertia of the larger mass, might be (part of) the explanation. The difference in speed between MRR-pa and MRR-ta, with MRR-ta being slower, could be explained by an earlier neurological maturation of jaw and lip movements as compared to tongue tip movements. Lip and jaw movements stabilise earlier in speech motor control development as compared to tongue movement (Terband, Maassen, Van Lieshout, & Nijland, 2011; Terband, Van Brenk, Van Lieshout, Nijland, & Maassen, 2009).

Taken all MRR sequences into account, our results show that MRR-pataka is the slowest sequence, which is probably due to the fact that the motor program of trisyllabic sequences is more complex than mono- or bisyllabic sequences (Wright et al., 2009). Furthermore, it can also be due to physiological aspects as described above. However, contradictory results are described in previous studies. In the studies of Rvachew et al. (2006) and Thoonen et al. (1996) the monosyllabic sequences were slower than the trisyllabic sequences, whereas several other studies found that in their population the MRR-pataka was slower than the monosyllabic sequences (Blech, 2010; Modolo et al., 2011; Wong et al., 2011). Differences in these outcomes are probably due to the use of different protocols. In addition to other studies, our study also investigated the MRR rate of bisyllabic sequences. The mean MRR rate of both bisyllabic sequences was similar to MRR-ta, and thus faster than the production of the monosyllabic sequence MRR-ka. Also, no previous studies have described normative data of MRR scores based on such a large representative sample as in our study. To summarise, the data of our study shows influences from *physiological factors*; larger movement inertia of the tongue body as compared to the tongue tip (i.e. MRR-ta > MRR-ka); from *neurological maturation*; jaw and lips movements stabilise earlier than tongue tip and tongue body movements (i.e. MRR-pa > MRR-ta and MRR-ka):, and *sequence complexity*; sequencing is more complex when more different units must be produced (i.e. MRR monosyllabic sequences > MRR bisyllabic sequences > MRR trisyllabic sequences). How these three factors (physiological factors, neurological maturation and sequence complexity) interact will have to be investigated further.

Gender differences

For all MRR sequences, overall rates were higher for boys than for girls. Prathanee et al. (2003) also found significant higher MRR scores for boys than for girls for /pə/, /tə/, /kə/, and /pə-tə/. Modolo et al. (2011) described older children and found for the 8-year-old children that boys performed faster on /pa/ and girls performed faster on /ta/ and /ka/. For the 9-year-old children these results were different; girls were overall faster than boys. At the age of 10 years girls were still faster than boys, except for the sequences /pataka/. However, other studies (Fletcher, 1972; Henry, 1990; Icht & Ben-David, 2015; Robbins & Klee, 1987; Wong et al., 2011; Zamani et al., 2017) found no differences between the performance of boys and girls in similar age ranges as our study. Our findings suggest that at the level of motor speech tasks, less taxing on linguistic skills, boys outperform girls. This is in contrast with studies that found boys showing a slower maturation of the speech motor development (Smith & Zelaznik, 2004), and in contrast with studies concluding that phonological accuracy measures of girls are better than that of boys (Dodd, Holm, Hua, & Crosbie, 2003).However, the results of this study should be interpreted with care; the sample is large, yet the effect size is small (Pek & Flora, 2018). Further research is needed.

Clinical implications and future perspectives

Despite of the ongoing debate on the clinical value of MRR, it has been suggested to have an important function in the assessment of children with MSD, and especially in children with CAS (Murray et al., 2015). Children with MSD show difficulties on MRR tasks when compared to typically developing children, more specifically with the speed(ing up) (Henry, 1990; Thoonen et al., 1996; Wit et al., 1993) and with the sequencing of different speech sounds (Henry, 1990; Thoonen et al., 1996). The studies of Thoonen (1999; 1996) indicate that monosyllabic MRR sequences differentiate children with spastic dysarthria from children with CAS and typically developing children. In addition, MRR can contribute to a first step in differential diagnosis between different types of speech sound disorders (SSD), and especially between different types of MSD. MRR offers insight into possible underlying motor execution impairments (Terband et al., 2019), and is thereby a potential added value in describing a complete speech profile. With only tasks like picture naming and nonword imitation it is not possible to distuingish a speech motor execution impairment from problems in lemma acces, word form selection, and phonological encoding (Van Haaften, Diepeveen, Terband, et al.. 2019).

In this protocol, articulation errors were not included in the analysis. As a result, there are missing values in the norm dataset. However, we consider the remaining data as sufficient to draw conclusions. Studies are currently being conducted to collect MRR data from children with SSD. With the normative data presented in this study and MRR data from children with SSD, clinicians will be able to distinguish typically developing children from children with SSD.

The present study is the largest available study using a standardised administration procedure for the age range 3;0 to 6;11 years. However, the test-retest of the norm group shows a low score for the bi- and tri-syllabic sequences. This is related to a test-retest effect; children were significantly faster on the second test moment because they know what they are expected (Diepeveen, et al. 2019). The normative data of our study is based on a large and representative sample of only Dutch-speaking children. Therefore, the clinical usability of our data in other languages must be discussed. Icht and Ben-David (2014) demonstrated that MRR score is influenced by language differences. They found significant differences in adults in MRR scores between English, Portuguese, Farsi and Greek-speaking persons, with the mean MRR in the Portuguese and Greek sample being faster than the mean MRR in the English sample and the mean MRR in Farsi being slower than in English. Prathanee et al. (2003) found differences in speech rate on an MRR task between English-speaking and Thai-speaking children. They therefore suggest using the norm data of English with English-speaking children and the Thai norms for children who speak Thai. They suggest that the shorter height, and coinciding smaller lung volume, of Thai children when compared to Western children, influences the slower MRR score of Thai children. However, we hypothesise that this explanation is not plausible, since lung volume is related mainly to length of sequence (Pennington et al., 2006) and not to speed of the articulation. Furthermore, Diepeveen et al. (2019) showed that length of sequence is independent of rate. The described language differences can be a possible explanation for the differences found between the results of the present study and other studies, besides differences in sample size and sample representativeness. For example, in the English language the voiceless stops (/p, t, k/) are aspirated in syllable initial position, whereas in Dutch these stops are not aspirated. These findings suggest that reference norms cannot be generalised across languages. In addition, in the past different protocols were used for measuring MRR score (time-by-count or count-by-time measures), making it even more difficult to compare normative data between languages (Diepeveen et al., 2019). We suggest to use this protocol for MRR studies in children for further studies in other languages.

Acknowledgements

The authors did not receive any grant funding for this study, but this research was partially supported by Het Revalidatiefonds (The Fund for Rehabilitation). The authors would like to thank all children, parents, SLP students, and SLPs for their participation in this study.

Declaration of interest

No potential conflict of interest was reported by the authors.

References

Ackermann, H., Hertrich, I., & Hehr, T. (1995). Oral diadochokinesis in neurological dysarthrias. *Folia Phoniatrica Et Logopaedica, 47*(1), 15-23.

Blech, A. (2010). *Untersuchung zur Kernsymptomatik bei Kindern mit einer kindlichen Sprechapraxie im Alter von 4-7 Jahren.* Rheinisch-Westfälischen Technischen Hochschule Aachen, Aachen.

Boersma, P., & Weenink, D. (2016). Praat: doing phonetics by computer (Version 6.0.21).

Diepeveen, S., Van Haaften, L., Terband, H., De Swart, B., & Maassen, B. (2019). A standardized protocol for maximum repetition rate assessment in children. *Folia Phoniatrica Et Logopaedica*, *71*, 238-250.

Dodd, B., Holm, A., Hua, Z., & Crosbie, S. (2003). Phonological development: a normative study of British English speaking children. *Clinical Linguistics & Phonetics, 17*(8), 617-643.

Fletcher, S. G. (1972). Time-by-count measurement of diadochokinetic syllable rate. *Journal of Speech and Hearing Research, 15*, 763-770.

Henry, C. E. (1990). The development of oral diadochokinesia and non-linguistic rhythmic skills in normal and speech-disordered young children. *Clinical Linguistics & Phonetics, 4*(2), 121-137.

Icht, M., & Ben-David, B. M. (2014). Oral-diadochokinesis rates across languages: English and Hebrew norms. *Journal of Communication Disorders, 48*, 27-37.

Icht, M., & Ben-David, B. M. (2015). Oral-diadochokinetic rates for Hebrew-speaking school-age children: Real words vs. non-words repetition. *Clinical Linguistics & Phonetics, 29*(2), 102-114.

Juste, F. S., Rondon, S., Sassi, F. C., Ritto, A. P., Colalto, C. A., & Furquim de Andrade, C. R. (2012). Acoustic analyses of diadochokinesis in fluent and stuttering children. *Clinics, 67*(5), 409-414.

Kent, R. D. (2015). Nonspeech oral movements and oral motor disorders: A narrative review. *American Journal of Speech-Language Pathology, 24*(4), 763-789.

Kent, R. D., Kent, J. F., & Rosenbek, J. C. (1987). Maximum performance tests of speech production. *Journal of Speech and Hearing Disorders, 52*(4), 367.

Knuijt, S., Kalf, J., Van Engelen, B., Geurts, A., & de Swart, B. (2019). Reference values of maximum performance tests of speech production. *International Journal of Speech-Language Pathology*, *21*(1), 56-64.

Maassen, B., & Terband, H. (2015). Process-oriented diagnosis of childhood and adult apraxia of speech (CAS and AOS). In M. A. Redford (Ed.), *The handbook of speech production* (Vol. First Edition, pp. 331-350). Hoboken, NJ: Wiley.

Maassen, B., Van Haaften, L., Diepeveen, S., Terband, H., Van den Engel-Hoek, L., Veenker, T., & De Swart, B. (2019). *Computer Articulation Instrument*. Amsterdam: Boom Uitgevers.

Modolo, D. J., Berretin-Felix, G., Genaro, K. F., & Brasolotto, A. G. (2011). Oral and vocal fold diadochokinesis in children. *Folia Phoniatrica Et Logopaedica, 63*(1), 1-8.

Murray, E., McCabe, P., Heard, R., & Ballard, K. J. (2015). Differential diagnosis of children with suspected childhood apraxia of speech. *Journal of Speech, Language, and Hearing Research, 58*(1), 43-60.

Namasivayam, A. K., Coleman, D., O'Dwyer, A., & van Lieshout, P. (2020). Speech Sound Disorders in Children: An Articulatory Phonology Perspective. *Frontiers in Psychology*, *10*, 1-22.

Padovani, M., Gielow, I., & Behlau, M. (2009). Phonarticulatory diadochokinesis in young and elderly individuals. *Arq Neuropsiquiatr, 67*(1), 58-61.

Pek, J., & Flora, D. B. (2018). Reporting effect sizes in original psychological research: A discussion and tutorial. *Psychological Methods*, *23*(2), 208.

Pennington, L., Smallman, C., Farrier, F. (2006). Intensive dysarthria therapy for older children with cerebral palsy: findings from six cases. *Child Language Teaching and Therapy, 22*(3), 255-273.

Peter, B., Lancaster, H., Vose, C., Fares, A., Schrauwen, I., & Huentelman, M. (2017). Two unrelated children with overlapping 6q25. 3 deletions, motor speech disorders, and language delays. *American Journal of Medical Genetics Part A, 173*(10), 2659-2669.

Peter, B., Matsushita, M., & Raskind, W. H. (2012). Motor sequencing deficit as an endophenotype of speech sound disorder: A genome-wide linkage analysis in a multigenerational family. *Psychiatric genetics, 22*(5), 226.

Prathanee, B., Thanaviratananich, S., & Pongjanyakul, A. (2003). Oral diadochokinetic rates for normal Thai children. *International Journal of Language & Communication Disorders, 38*(4), 417-428.

Priester, G. H. & Goorhuis-Brouwer, S. M. (2013). Measuring speech sound development: An item response model approach. *International Journal of Pediatric Otorhinolaryngology, 77*, 1469-1473.

Robbins, J., & Klee, T. (1987). Clinical assessment of oropharyngeal motor development in young children. *Journal of Speech and Hearing Disorders, 52*(3), 271-277.

Rvachew, S., Hodge, M., & Ohberg, A. (2005). Obtaining and interpreting maximum performance tasks from children: A tutorial. *Journal of Speech Language Pathology and Audiology*, *29*(4), 146.

Rvachew, S., Ohberg, A., & Savage, R. (2006). Young children's responses to maximum performance tasks: preliminary data and recommendations. *Journal of Speech Language Pathology and Audiology, 30*(1), 6-13.

Smith, A., & Zelaznik, H. N. (2004). Development of functional synergies for speech motor coordination in childhood and adolescence. *Developmental Psychobiology, 45*(1), 22-33.

Terband, H., Maassen, B., & Maas, E. (2019). A psycholinguistic framework for diagnosis and treatment planning of developmental speech disorders. *Folia Phoniatrica Et Logopaedica, 71*, 216-227.

Terband, H., Maassen, B., Van Lieshout, P., & Nijland, L. (2011). Stability and composition of functional synergies for speech movements in children with developmental speech disorders. *Journal of Communication Disorders, 44*(1), 59-74.

Terband, H., Van Brenk, F., Van Lieshout, P., Nijland, L., & Maassen, B. (2009). Stability and composition of functional synergies for speech movements in children and adults. *Interspeech 2009*, 788-791.

Thoonen, G., Maassen, B., Gabreels, F., & Schreuder, R. (1999). Validity of maximum performance tasks to diagnose motor speech disorders in children. *Clinical Linguistics & Phonetics, 13*(1), 1-23.

Thoonen, G., Maassen, B., Wit, J., Gabreels, F., & Schreuder, R. (1996). The integrated use of maximum performance tasks in differential diagnostic evaluations among children with motor speech disorders. *Clinical Linguistics & Phonetics, 10*(4), 311-336.

Turner, S. J., Mayes, A. K., Verhoeven, A., Mandelstam, S. A., Morgan, A. T., & Scheffer, I. E. (2015). GRIN2A: an aptly named gene for speech dysfunction. *Neurology, 84*(6), 586-593.

Van Haaften, L., Diepeveen, S., Van den Engel-Hoek, L., Jonker, M., de Swart, B., & Maassen, B. (2019). The psychometric evaluation of a speech production test battery for children: the reliability and validity of the computer articulation instrument. *Journal of Speech, Language and Hearing Research, 62*, 2141-2170.

Van Haaften, L., Diepeveen, S., Terband, H., Vermeij, B., van den Engel-Hoek, L., de Swart, B., & Maassen, B. (2019). Profiling speech sound disorders for clinical validation of the computer articulation instrument. *American Journal of Speech-Language Pathology*, *28*(2S), 844–856.

Williams, P., & Stackhouse, J. (2000). Rate, accuracy and consistency: diadochokinetic performance of young, normally developing children. *Clinical Linguistics & Phonetics, 14*(4), 267-293.

Wit, J., Maassen, B., Gabreels, F., & Thoonen, G. (1993). Maximum performance tests in children with developmental spastic dysarthria. *Journal of Speech, Language, and Hearing Research, 36*(3), 452-459.

Wong, A. W., Allegro, J., Tirado, Y., Chadha, N., & Campisi, P. (2011). Objective measurement of motor speech characteristics in the healthy pediatric population. *Int J Pediatr Otorhinolaryngol, 75*(12), 1604-1611.

Wright, D. L., Robin, D. A., Rhee, J., Vaculin, A., Jacks, A., Guenther, F. H., & Fox, P. T. (2009). Using the self-select paradigm to delineate the nature of speech motor programming. *Journal of Speech, Language, and Hearing Research, 52*(3), 755-765.

Yaruss, J. S., & Logan, K. J. (2002). Evaluating rate, accuracy, and fluency of young children's diadochokinetic productions: a preliminary investigation. *Journal of Fluency Disorders, 27*(1), 65-86.

Zamani, P., Rezai, H., & Garmatani, N. T. (2017). Meaningful words and non-words repetitive articulatory rate (oral diadochokinesis) in Persian speaking children. *Journal of psycholinguistic research, 46*(4), 897-904.

Ziegler, W., Schölderle, T., Brendel, B., Amsellem, J., & Staiger, A. (2019). Higher-faster-farther: maximum performance tests in the assessment of neurogenic speech impairment. *Folia Phoniatrica et Logopaedica, 71*, 261-274.

4

TWO

CLINICAL IMPLEMENTATION OF THE CAI PART 2

FIVE

CHAPTER 5

PROFILING SPEECH SOUND DISORDERS

FOR CLINICAL VALIDATION OF THE COMPUTER ARTICULATION INSTRUMENT

Leenke van Haaften, Sanne Diepeveen, Hayo Terband, Bernadette Vermeij, Lenie van den Engel-Hoek, Bert de Swart and Ben Maassen

PUBLISHED IN
AMERICAN JOURNAL OF SPEECH LANGUAGE PATHOLOGY, 2019; 28(2S); 844-856

Abstract

Purpose: The current article presents data from 2 studies on clinical groups of children referred for speech assessment. The aims of these studies are to validate the Computer Articulation Instrument (CAI) with the known-group validation method and to determine the differential diagnostic power of the resulting speech profiles.

Method: Study 1 examined known-group validity by comparing the scores of 93 children diagnosed with speech-language difficulties on the picture naming (PN) task of the CAI with intelligibility judgments given by speech-language pathologists. In Study 2, the speech profiles of 41 children diagnosed with speech sound disorders (SSDs), consisting of 4–6 factor scores extracted from the 4 tasks of the CAI, namely, PN, nonword imitation (NWI), word and nonword repetition, and maximum repetition rate (MRR), were validated against clinical judgments of severity of the SSD given by speech-language pathologists.

Results: In Study 1, a repeated-measures analysis of variance revealed a significant effect of intelligibility level on the PN performance of the CAI and there were highly significant correlations between intelligibility and PN performance in the expected direction. Neither intelligibility level nor PN performance was related to nonverbal intelligence and language scores. The analysis of variance and a series of *t* tests in Study 2 revealed significant differences between the moderate and severe groups for the CAI factors based on PN and NWI and the bisyllabic and trisyllabic sequences of MRR, but not for the factor word and nonword proportion of whole-word variability based on word and nonword repetition, and the monosyllabic sequences of MRR. These results suggest that, especially, the tasks PN, NWI, and the bisyllabic and trisyllabic sequences of MRR are most sensitive for diagnosing SSDs.

Conclusions: The findings of these 2 studies support the known-group validity of the CAI. Together with the results of a previous study of our group on reliability and validity (van Haaften et al., 2019), we can conclude that the CAI is a reliable and valid tool for assessment of children with SSDs.

Introduction

Children with speech production problems are one of the four subtypes that can be distinguished in children with a specific language impairment (Van Weerdenburg, Verhoeven, & Van Balkom, 2006). They show a specific profile as compared to the other subtypes of children with language impairments: difficulties with lexical–semantic abilities, with auditory conceptualization, or with verbal sequential memory (Van Weerdenburg et al., 2006). Recently, Bishop et al. (2017) proposed to use the term *developmental language disorder* (DLD) when a language disorder was not associated with a known biomedical etiology. They state that DLD is a heterogeneous category that encompasses a wide range of problems, including expressive phonological problems. Phonological problems in preschoolers that are not accompanied by other language problems do not meet the criteria for DLD. Therefore, Bishop et al. propose to use the more general term *speech sound disorder* (SSD) for such cases. SSD is an umbrella term that includes expressive phonological problems and problems with speech production that have motor or physical origins or involve misarticulations such as a lisp, where a sound is produced in a distorted way without losing the contrast with other sounds. Children with SSDs are one of the most common clinical populations for speech-language pathologists (SLPs; Mullen & Schooling, 2010); the reported prevalence is highly variable, ranging from 2.3% to 24.6% (Eadie et al., 2015; Law, Boyle, Harris, Harkness, & Nye, 2000). They form a heterogeneous group, showing variability in severity, etiology, proximal causes, speech error characteristics, and response to treatment (Dodd, 2011).

There are several widely recognized classification systems for SSDs featuring a variety of approaches, namely, etiology, descriptive linguistics, and psycholinguistic and psychomotor processing (Waring & Knight, 2013). In current practice, symptom patterns form the basis of diagnostic classification (Dodd, 1995b, 2014). The Speech Disorders Classification System described by Shriberg et al. (2017) divides SSDs into three classes, based on etiology: speech delay, speech errors, and motor speech disorder (MSD; including dysarthria, childhood apraxia of speech [CAS], and MSD–not otherwise specified). Examples of symptoms of MSD include slow speech rate, distorted substitutions of speech sounds, increased difficulty with multisyllabic words, and prosodic errors. Yet, there is no validated list of diagnostic patterns for differential diagnosis of SSDs. For example, one of the speech symptoms that is described for different types of SSDs is inconsistency of speech errors. From a phonological point of view, high inconsistency of speech errors could indicate an unstable phonological system, also called a *phonological planning deficit* (Dodd, 1995a; Macrae, Tyler, & Lewis, 2014), or unstable lexical representations (Sosa & Stoel-Gammon, 2012). However, inconsistency is also a characteristic of CAS (Davis, Jakielski, & Marquardt, 1998; Forrest, 2003; Iuzzini-

Seigel, Hogan, & Green, 2017). In the latter case, inconsistency is explained by an unstable motor system (articulomotor planning and programming). Thus, the same symptom can refer to different underlying deficits, and the same deficit can result in different symptoms, leading to a wide variety of symptoms within subtypes and much symptomatic overlap between subtypes of SSDs. Therefore, in clinical practice, a reorientation from behavioural diagnostics to process-oriented diagnostics is required in order to reveal the proximal causes of SSDs (Terband & Maassen, 2012).

Psycholinguistic and psychomotor models give a conceptual basis to analyze speech disorders and form the basis for a process-oriented diagnostic classification system based on the identification of the breakdown in the chain of sequential and parallel speech processes (Baker, Croot, McLeod, & Paul, 2001). Rather than categorization of SSDs based on single symptoms or sets of symptoms, process-oriented diagnostics primarily focus on speech profiles comprising clustered symptoms that can be interpreted in terms of the underlying speech production processes. An example of a psycholinguistic processing model is the model described by Levelt (1989), in which "conceptualizing a preverbal message," either from memory or from perception, is the first process in speaking. The next process is formulating a word or sentence, driven by two steps of lexicalization: selecting a lemma, containing meaning and grammatical information, and the corresponding lexeme or word form, which forms the input for the next stage of phonological encoding. *Phonological encoding* entails specifying the sequence of speech sounds together with their syllabic and prosodic structure. These syllables are the basic units of articulomotor planning and programming. The final process of actually performing the articulatory movements is *execution*, resulting in an acoustic speech signal (Maassen & Terband, 2015). Levelt, Roelofs, and Meyer (1999) validated this processing model with normal speech production data, and Nijland (2003) further elaborated on the planning, execution, and monitoring stages of the model and applied it to analyses of SSDs. By conducting different speech experiments in children with CAS, Nijland could conclude that both phonetic planning and motor programming are deviant in children with CAS. Levelt's model is relevant for analyzing SSDs because of the stages lexeme retrieval, phonological encoding, and self-monitoring, which are the processes underlying consistent and inconsistent phonological disorder (PD). MSDs, of which CAS and dysarthria are the main diagnostic categories, can be described by means of the motor planning, programming, and execution processes. However, the main objective of a process-oriented approach is not to categorize but to give a complete characterization of the speech profile, such that underlying processing deficits can be identified. Insight into the deficits that might be the underlying causes of the child's difficulty requires an extensive analysis of a child's performance on a range of speech tasks

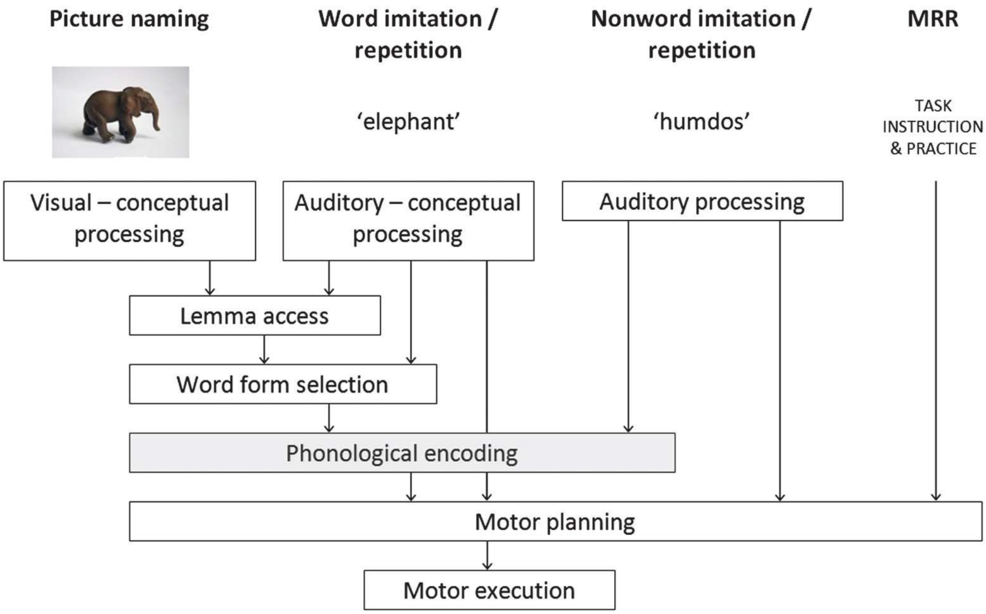

Figure 1. The speech production processes assessed in the four tasks of the Computer Articulation Instrument (Maassen & Terband, 2015; Figure 15.2). MRR = maximum repetition rate.

that reflect different levels of processing. Based on these premises, the Computer Articulation Instrument (CAI) was developed (Maassen et al., 2019). The CAI consists of a battery of speech production tasks and is based on a series of studies of Dutch children with developmental and acquired SSDs (Nijland, Maassen, & van der Meulen, 2003; Nijland, Maassen, van der Meulen, Gabreëls, et al., 2003; Nijland, Terband, & Maassen, 2015; Thoonen, Maassen, Gabreels, & Schreuder, 1999; Thoonen, Maassen, Gabreëls, & Schreuder, 1994). The CAI has a modular structure and provides an interactive administration and scoring of four speech tasks. The tasks comprise (a) picture naming (PN), (b) nonword imitation (NWI), (c) word and nonword repetition (WR and NWR), and (d) maximum repetition rate (MRR), thereby covering phonological and speech motor skills. 5

As demonstrated in Figure 1, PN taps into the whole chain of speech processes, from preverbal visual–conceptual processing to lemma access, word-form selection, phonological encoding, motor planning, and articulation (motor execution; Maassen & Terband, 2015). During NWI, a child is asked to reproduce nonwords (or nonsense words). In contrast to PN, a child cannot revert to its lexicon during this task, and thus the child either needs to analyze the phonological structure of the nonword directly, addressing the phonological decoding and encoding system, or follows the auditory-to-motor-planning pathway. In WR and NWR, a child is asked to repeat a word or nonword five times. This task aims to assess variability in speech production, which occurs when a child uses multiple productions of the same word or nonword. MRR is a pure motor task (articulomotor planning and programming) and does not require any knowledge of words, syllables, or phonemes. The evaluation of speech production in the CAI is based on phonetic transcriptions and acoustic measurements. Both the tasks and speech analyses are computer implemented (van Haaften et al., 2019). Rather than focusing on single diagnostic markers, two types of analyses are conducted within the CAI: (a) objective and quantitative assessment of symptoms and (b) contrasting severity of symptoms across tasks. The outcome of this assessment battery is a speech performance profile that can be interpreted as characteristics of breakdown in underlying processes. Normative data from 1,524 children in the age range of 2;0–6;11 (years;months) have been collected, such that performance on the CAI as a whole, as well as the profile of performances on the different tasks, can be quantified in percentile scores, which allows for interpretation in terms of strengths and weaknesses (Maassen et al., 2019).

In a previous study of our research group, we assessed the psychometric properties of the CAI, including reliability and construct validity (van Haaften et al., 2019). Overall, sufficient to good values were found for interrater reliability, but intraclass correlation coefficients on test–retest reliability were low, probably due to better performance at retest reflecting a test–retest learning effect in addition to normal development. The study also described two aspects of construct validity. The first aspect, criterion validity, was confirmed by clear and significant age trends in CAI parameters in a large sample of typically developing children aged between 2 and 7 years. The second aspect of construct validity, structural validity, was assessed by factor analysis and correlations. Factor analyses on a total number of 20 parameters revealed five meaningful factors: PN; segmental quality of NWI (NWI-Seg); quality of syllabic structure of NWI (NWI-Syll); word and nonword proportion of whole-word variability (PWV), based on WR and NWR; and MRR. Weak correlations were found between CAI factor scores, indicating the independent contribution of each factor to the speech profile.

Further steps are needed in the validation process of the CAI. The ultimate goal is to assess the strengths of the five CAI factors in identifying the breakdown of speech processes in children with SSDs (process-oriented diagnostics), which will be described in future articles. The more immediate step, determining known-group validity, is presented in the current study. Known-group validity is a third aspect of construct validity and refers to the degree to which a measure is sensitive to differentiate between subgroups that are hypothesized to have different scores (Portney & Watkins, 2009). To assess this aspect of construct validity of the CAI, this article presents data from two studies on clinical groups of children with speech language impairments and SSDs. The aim of Study 1 is to determine known-group validity by comparing the scores of children with speech language impairments, as diagnosed on the basis of language and intelligence tests, on one task of the CAI (PN) with intelligibility judgments given by SLPs. Study 2 aims to determine the diagnostic power of all four tasks of the CAI by comparing the five CAI factors: PN, NWI-Seg, NWI-Syll, PWV, and MRR (see also Table 4) with a severity judgment of the speech difficulties (mild, moderate, and severe) of children with SSDs.

Study 1

The first study was designed to validate the scores on the PN task of the CAI with intelligibility judgments (good, moderate, poor) in children diagnosed with speech language impairments. For this study, the parameter "percentage of consonants correct" of the PN task is used (PN-PCC), and nonverbal intelligence and language tests are used for the speech language impairment diagnosis.

Method

Ethics, Consent, and Permissions

The research ethics committee of the Radboud University Nijmegen Medical Centre stated that this study does not fall within the remit of the Medical Research Involving Human Subjects Act (Wet medisch-wetenschappelijk onderzoek met mensen; file number: CMO 2016-2985). Therefore, this study can be carried out (in the Netherlands) without an approval by an accredited research ethics committee. Informed consent was obtained from all parents or guardians.

Participants

Ninety-three children aged between 3;0 and 4;0 participated in this study (see Table 1). The sample consisted of 73 boys and 20 girls, representative for the gender distribution in children with speech language impairments. All children attended one of the intervention centers for preschoolers with speech language impairments at the Nederlandse Stichting voor het Dove en Slechthorende Kind, a specialized diagnostic and intervention center for children with hearing loss or speech language impairments. Before admission to the center, these children had been referred to an audiology center (AC) by their family doctor or health care physician on the basis of suspected speech language impairment. At the AC, nonverbal intelligence is assessed by a psychologist, receptive and expressive language tests are administered by an SLP, and hearing status is evaluated by audiometry. Children meet the criteria for referral to a speech language impairment intervention center when they have difficulties in language production and/or language comprehension and/or when their speech is highly unintelligible. Admission takes place if they have a score of at least 1.5 SDs below the mean on at least one standardized, norm-referenced language test. Children with hearing loss of 25 dB or more were excluded for this study.

Nonverbal intelligence and language skills were assessed within a period ranging from 3 months before until 3 months after the start of the intervention. If language scores were missing or were older than 3 months at the start of the intervention, language performance was assessed by the SLP of the intervention center within 3 months after the intervention started.

Table 1. Number of children per age category and completed tests.

Age category	N	Boys	Girls	NVIQ	QPPVT	RLQ	SWQ	PN-PCC-Q
36–39 months	29	23	6	26	25	23	22	29
40–43 months	35	28	7	32	33	21	22	35
44–47 months	29	22	7	26	28	19	17	29
Total	93	73	20	84	86	63	61	93
% Missing values				9.7%	7.5%	32.3%	34.4%	0%

Note. NVIQ = nonverbal intelligence quotient; QPPVT = Peabody Picture Vocabulary Test, vocabulary quotient; RLQ = receptive language quotient; SWQ = sentence and word production quotient; PN-PCC-Q = Computer Articulation Instrument's picture naming percentage consonants correct quotient.

Materials and Procedure

Nonverbal intelligence was assessed with the Snijders- Oomen Nonverbal Intelligence Test 2½-7–Revised (Snijders, Tellegen, Winkel, & Laros, 2003), yielding a nonverbal intelligence quotient (NVIQ). Vocabulary was tested with the Dutch version of the Peabody Picture Vocabulary Test–III (Schlichting, 2005), yielding a vocabulary quotient (QPPVT). The Schlichting Test for Language Comprehension and Language Production (Schlichting & Spelberg, 2010a, 2010b) was used to measure receptive (receptive language quotient: RLQ) and expressive (sentence and word production quotient: SWQ) language skills. These norm-based standard scores or *Q* scores ($M = 100$, $SD = 15$) of each test were used for the analyses.

In addition to the measures for nonverbal intelligence and language, the CAI was administered to all the children (Maassen et al., 2019). For this study, the PN task of the CAI was used. The task was administered by SLPs of the speech language impairment early intervention group, specifically trained in the administration of the CAI. PN contains 60 words, covering the full inventory of vowels, consonants, clusters, and syllable structures of the Dutch language. For this study, the parameter PN-PCC was used for analyses. Individual PN-PCC scores were transformed into *z* scores by subtracting the mean of the normative group and dividing by the standard deviation of the study group; this was done for three age groups (36–39, 40–43, and 44–47 months) separately. The reason for dividing by the standard deviation of the study group rather than the standard deviation of the norm group was that the former was approximately three times as large as the latter (18.9 compared to 6.3). Applying the broader confidence intervals of the study group yields the more conservative estimates. *z* Scores were transformed into *Q* scores (formula: $Q = 100 + 15*z$) to make them comparable to the cognitive and language scores NVIQ, QPPVT, RLQ, and SWQ.

For each child, the SLP rated the intelligibility on a three-level scale: good, moderate, or poor. The same method is used in the study of Lohmander, Lundeborg, and Persson (2016). Twenty-two children were rated with a "good" intelligibility, 46 were rated as with a "moderate" intelligibility, and 25 children were rated with a "poor" intelligibility.

Statistical Analyses

To test the hypothesis that there is a difference in mean *Q* scores of the nonverbal intelligence test, language tests, and CAI for the three intelligibility levels, a one-way repeated-measures analysis of variance (ANOVA) was conducted with *Q* score as a dependent variable, test instrument as a within-subject factor (five levels: NVIQ, QPPVT, RLQ, SWQ, and PN-PCC quotient [PN-PCC-Q]), and intelligibility level as a between-subjects factor (three levels: good, moderate, and poor). Mauchly's test of sphericity was conducted to test the hypothesis that the variances of differences between conditions are equal. Bonferroni correction was applied for post hoc comparisons. A series of ANOVAs was performed to evaluate differences between *Q* scores for the three levels of intelligibility. Levene's test of equality of error variances was conducted to test the homogeneity of variance assumption. Bonferroni correction was applied for post hoc comparisons. Correlations between *Q* scores and intelligibility levels were calculated with Spearman rank correlation coefficients, and correlations between the *Q* scores of the different tests were calculated with Pearson rank correlation coefficients. Missing values were replaced by the mean per age group (i.e., mean imputation method). All statistical analyses were performed using SPSS Version 20 for Windows (SPSS Inc.).

Results

Mean *Q* scores and standard deviations of all tests for the three intelligibility levels are shown in Table 2. Comparing the profiles of *Q* scores across tests, it was found that, in the levels of moderate and poor intelligibility, on average, children achieved the highest scores on the nonverbal intelligence test, followed by the vocabulary test, the receptive language test, and the expressive language tests. The lowest *Q* scores were obtained for PN-PCC-Q. In contrast, children with a "good" intelligibility also showed the highest scores for the nonverbal intelligence, but in this group, PN-PCC-Q was higher than the language *Q* scores, which were approximately equal. Thus, of all *Q* scores, PN-PCC-Q shows the largest decrease between groups from good to poor intelligibility.

Table 2. Mean Q scores for the nonverbal intelligence, language, and speech tests.

Intelligibility score	N	NVIQ		QPPVT		RLQ		SWQ		PN-PCC-Q	
		M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Good	22	102.8	11.6	84.7	18.8	78.6	11.3	78.2	10.4	92.5	5.99
Moderate	46	99.9	11.2	89.9	16.6	80.9	13.7	74.0	9.50	73.4	11.3
Poor	25	100.9	11.4	90.9	18.2	82.2	14.0	71.3	9.33	62.5	14.1

Note. NVIQ = nonverbal intelligence quotient; QPPVT = Peabody Picture Vocabulary Test, vocabulary quotient; RLQ = receptive language quotient; SWQ = sentence and word production quotient; PN-PCC-Q = Computer Articulation Instrument's picture naming percentage consonants correct quotient.

A one-way repeated-measures ANOVA was conducted with the *Q* scores of the five test instruments as repeated measures and intelligibility level as a between-subjects variable. Mauchly's test indicated that the assumption of sphericity had been violated, $\chi^2(9) = 58.9$, $p < .001$; therefore, degrees of freedom were corrected using Huynh–Feldt estimates of sphericity ($\varepsilon = .78$). The results show that the within-subject factor "test instrument" was significant, $F(3.10, 278.96) = 79.78$, $p < .001$, effect size or partial $\eta^2 = .47$, which means that the scores on the test instruments were significantly affected by intelligibility level. The between-subjects factor "intelligibility level" was marginally significant, $F(2, 90) = 3.09$, $p = .051$, effect size or partial $\eta^2 = .064$. Post hoc analyses showed that the difference of mean *Q* scores was not significant between "good" and "moderate" levels ($p = .217$), nor between "moderate" and "poor" levels ($p = .556$), but was significant between "good" and "poor" levels ($M = 6.78$, $SE = 2.47$, $p = .022$). In addition, there was a significant interaction between intelligibility levels and "test instrument," $F(6.20, 278.96) = 10.00$, $p < .001$, effect size or partial $\eta^2 = .18$. To further examine this interaction, a series of ANOVAs was conducted to test the differences between the three intelligibility levels for the Q scores of each test instrument separately. There was no significant difference between intelligibility levels for NVIQ, $F(2, 90) = 0.47$, $p = .626$; QPPVT, $F(2, 90) = 0.87$, $p = .421$; RLQ, $F(2, 90) = 0.43$, $p = .650$; or SWQ, $F(2, 90) = 3.07$, $p = .051$. For the latter, marginally significant factor SWQ, post hoc analyses revealed a significant mean difference between "good" and "poor" levels ($p = .047$) and no significant mean differences between "good" and "moderate" levels ($p = .276$) or "moderate" and "poor" levels ($p = .795$). For PN-PCC-Q, the Levene's test for equality of variances was significant, indicating that the requirement of homogeneity of variance was violated. Therefore, the Welch *F* ratio was calculated, showing that the difference in mean PN-PCC-Q between intelligibility levels was significant, $F(2, 51.28) = 69.48$, $p \leq .001$.

Table 3 shows correlations between intelligibility and *Q* scores. A strong, significant correlation was found between PN-PCC-Q and intelligibility (Spearman $r(93) = .69$, $p < .001$), which is in the expected direction: PN-PCC-Q decreases when the intelligibility level decreases. No other *Q* scores, not even the expressive language score SWQ, correlated significantly with intelligibility or with PN-PCC-Q.

There were weak, significant correlations between the outcome of the nonverbal intelligence test and language tests and moderate correlations among the language tests, with correlations between RLQ and SWQ and between QPPVT and RLQ being moderate and the correlation between QPPVT and SWQ being weak. No significant correlations were found between PN-PCC-Q and the *Q* scores of the nonverbal intelligence test and language tests. Inspection of the scatter plots did not reveal any outliers.

Table 3. Spearman and Pearson rank correlations between intelligibility levels and *Q* scores and between *Q* scores (*N* = 93).

Intelligibility level and Q scores		Intelligibility level	NVIQ	QPPVT	RLQ	SWQ	PN-PCC-Q
Intelligibility level	Spearman *r*	1	.027	-.14	-.11	.20	.69**
NVIQ	Pearson *r*	—	1	.36**	.31**	.35**	.10
QPPVT	Pearson *r*		—	1	.52**	.36**	-.22
RLQ	Pearson *r*			—	1	.48**	-.15
SWQ	Pearson *r*				—	1	.21
PN-PCC-Q	Pearson *r*					—	1

Note. NVIQ = nonverbal intelligence quotient; QPPVT = Peabody Picture Vocabulary Test, vocabulary quotient; RLQ = receptive language quotient; SWQ = sentence and word production quotient; PN-PCC-Q = Computer Articulation Instrument's picture naming percentage consonants correct quotient.

*Correlation of factor scores is significant at the .05 level (two-tailed). **Correlation of factor scores is significant at the .01 level (two-tailed).

Study 2

The second study aims to determine the diagnostic power of all four tasks of the CAI. For this, the relation between the five CAI factors (PN, NWI-Seg, NWI-Syll, word and nonword PWV, and MRR) and clinical judgments of severity of the speech disorder by the SLPs is investigated.

Method

Ethics, Consent, and Permissions

The ethics approval for Study 1 also applied to Study 2.

Participants

The participants in Study 2 were 41 children with an age range from 3;0 to 6;4, with 26 boys and 15 girls. For this study, children with SSDs were recruited from several institutions: 19 children from primary health care services, one child from an AC, and 21 children from a special school for children with language and hearing impairments. All parents or caregivers were given an information letter. After obtaining the signed parental consent form, the child was included in the study.

The parents or caregivers of all 41 children were asked to provide information about the children's hearing status. They were asked whether the child had a history of hearing problems, if hearing problems had been recorded during the regular governmental (neonatal) hearing screening, and, if available, if they could provide us with hearing acuity data (pure-tone thresholds). Thirty children passed a bilateral hearing screening at 20 dB. Parents or caregivers of the other 11 children reported no history of hearing problems and no hearing problems recorded during the regular governmental (neonatal) hearing screening.

Prior to the procedures of this study, a speech diagnosis was reported by the SLP of the child, based on clinical observation and a standard speech-language protocol, including standardized language tests. Speech was observed with different instruments. Until now, for the Dutch language, no standardized and normalized speech assessment is available. All children were diagnosed with SSDs, most of them (*n* = 36) with a PD, two children with CAS, and three children with an unknown diagnosis because no details were available about the children's speech apart from the fact that their SSD was severe. Differential diagnosis was part of the clinical reasoning process of the SLP and was done based on diagnostic criteria described in studies such as Forrest (2003) and Shriberg and Kwiatkowski (1994).

Materials and Procedure

For this study, all participants were tested on their speech skills with the CAI. All four tasks (PN, NWI, WR and NWR, and MRR) were administered. Both the administration of the tests and the analyses of the speech are computer implemented. Table 4 shows the parameters used to assess task performance; a detailed description of the CAI and these parameters, as well as a description of the normative data set, is presented in Maassen et al. (2019) and van Haaften et al. (2019); for all parameters, percentile scores can be determined. A factor analysis on all 20 parameters of the normative data, obtained from a total number of 1,524 children, yielded five factors: (a) PN, (b) NWI-Seg, (c) NWI-Syll, (d) PWV of words and nonwords, and (e) MRR (van Haaften et al., 2019). For this study, factor scores were calculated based on the factor weights obtained from this factor analysis. Because there were many missing values in the MRR task (see below), separate factor scores were calculated on only the monosyllabic MRR sequences (/papa../, /tata../, /kaka../; yielding factor MRRMono) and the bisyllabic (/pata../, /taka../) and trisyllabic (/pataka../) sequences, yielding factor MRR-BiTri.

Prior to the administration of the CAI, severity of the SSDs was judged by the child's SLP (*N* = 11) on a severity scale with three categories—mild, moderate, and severe— following the categories proposed by Dodd (1995c). An SLP rated the severity of an SSD as *mild* when a child is mostly intelligible in spontaneous speech but errors are obvious and distracting from content. The severity was rated *moderate* when single words are often intelligible in context but connected speech is often difficult to understand, particularly out of context. The category *severe* was rated when most utterances are unintelligible on the first meeting. Also, the persistence of the speech disorder and the consequences on communication abilities were taken into account when rating severity. The category “moderate” was scored for 14 children, and 27 children were scaled as “severe.” None of the children was scaled as having a “mild” speech disorder. Therefore, the statistical analyses of this study are based on two severity categories: moderate and severe. Table 5 shows the distribution of the participants in the three severity categories by speech diagnosis.

The tasks of the CAI were administered by (candidate) SLPs specifically trained in the administration of the CAI.

Table 4. Computer Articulation Instrument parameters per speech task and extracted factors.

Task	Factor	Parameter	
PN	PN	PCCI	Percentage of consonants correct in syllable-initial position
		PVC	Percentage of vowels correct
		Level 5	Percentage of correct consonants /l/ and /R/
		RedClus	Percentage of reduction of initial consonant clusters from two consonants to one
		CCVC	Percentage of correct syllable structure CCVC (C = consonant, V = vowel)
NWI	NWI-Seg	PCCI	Percentage of consonants correct in syllable-initial position
		PVC	Percentage of vowels correct
		Level 4	Percentage of correct consonants /b/, /f/, and /ʋ/
		Level 5	Percentage of correct consonants /l/ and /R/
		CVC	Percentage of correct syllable structure CVC
	NWI-Syll	RedClus	Percentage of reduction of initial consonant clusters from two consonants to one
		CCVC	Percentage of correct syllable structure CCVC
WR	PWV	PWV Word	Proportion of whole-word variability: word repetition
NWR	PWV	PWV Nonword	Proportion of whole-word variability: nonword repetition
MRR	MRR-Mono	MRR-pa	Number of syllables per second of sequence /pa/
		MRR-ta	Number of syllables per second of sequence /ta/
		MRR-ka	Number of syllables per second of sequence /ka/
	MRR-BiTri	MRR-pataka	Number of syllables per second of sequence /pataka/
		MRR-pata	Number of syllables per second of sequence /pata/
		MRR-taka	Number of syllables per second of sequence /taka/

Note. PN = picture naming; NWI = nonword imitation; WR = word repetition; NWR = nonword repetition; MRR = maximum repetition rate; PN = factor score of all parameters of picture naming; NWI-Seg = factor score of the segmental parameters of nonword imitation; NWI-Syll = factor score of the syllable structure parameters of nonword imitation; PWV = factor score of the two PWV parameters of word and nonword repetition; MRR-Mono = factor score of the monosyllabic items of maximum repetition rate parameters; MRR-BiTri = factor score of the bisyllabic and trisyllabic items of maximum repetition rate parameters.

Table 5. Speech diagnosis by severity categories.

Severity category	Speech disorder			
	PD	**CAS**	**Unknown**	**Total**
Mild	0	0	0	0
Moderate	13	1	0	14
Severe	23	1	3	27
Total	36	2	3	41

Note. PD = phonological disorder; CAS = childhood apraxia of speech.

Statistical Analyses

The factor PWV had two missing values, and these were replaced by the overall PWV mean (M = -1.20; i.e., mean imputation method). Much more missing data were observed for the MRR tasks, due to speech–motor difficulties and/or shyness or inattentiveness of the child; also, a few recordings could not be analyzed due to the low acoustic quality. Of the total number of 41 children, only 23 produced at least two monosyllabic sequences correctly (44% missing), and only nine of these 23 (amounting to 78% missing data) produced at least two of the bisyllabic or trisyllabic sequences. Because of this large number of missing values, no imputation was applied, but a separate analysis was conducted instead on the group of 23 children. The 14 children who were not able to produce the bisyllabic or trisyllabic sequences were assigned the lowest z score, such that failure to produce these sequences was marked as poor performance. One-way repeated-measures ANOVAs were conducted to test the hypothesis that there is a difference in CAI factors for the two severity categories, comprising two levels: “moderate” and “severe.” Because of the missing data in factors MRR-Mono and MRR-BiTri, the first analysis was conducted on the four remaining factors: PN, NWI-Seg, NWI-Syll, and PWV. Subsequently, a one-way repeated-measures ANOVA was conducted with six CAI factors, including MRR-Mono and MRR-BiTri. Mauchly’s tests of sphericity were conducted to test the hypothesis that the variances of differences between conditions are equal. Next, if in the ANOVA either severity level or the interaction between severity level and CAI factor was significant, a series of independent t tests was conducted to evaluate the difference in factor scores between the moderate and severe groups for each of the four or six CAI factors separately. Levene’s test of equality of error variances was conducted to test the homogeneity of variance assumption. Correlations between CAI factors and severity categories were calculated by Spearman rank correlation coefficients (r), and correlations between the CAI factors were assessed by calculating Pearson rank correlation coefficients (r). All statistical analyses were performed using SPSS Version 20 for Windows (SPSS Inc.).

Results

Table 6 shows that, on average, children with a speech disorder of moderate severity have higher factor scores on PN, NWI-Seg, NWI-Syll, and PWV, than children with a severe speech disorder. For the children with a severe speech disorder, mean factor scores ranged from -1.13 to -1.72; and for the children with moderate severity, between -0.18 and -1.07. Thus, all mean scores were below the population average.

First, a one-way repeated-measures ANOVA with the four CAI factors PN, NWI-Seg, NWI-Syll, and PWV was conducted. Mauchly's test indicated that the assumption of sphericity had been violated, $\chi^2(5) = 15.13$, $p = .010$; therefore, degrees of freedom were corrected using Huynh–Feldt estimates of sphericity ($\varepsilon = .91$). The results show that the within-subject factor "CAI factors" was significant, $F(2.74, 106.96) = 18.29$, $p < .001$, effect size or partial $\eta^2 = .32$, indicating that the factor scores of the CAI were significantly affected by the severity of the speech disorder. The between-subjects factor "severity category" was also significant, $F(1, 39) = 11.98$, $p = .001$, effect size or partial $\eta^2 = .24$; there was a significant difference in factor scores between the children with moderate and severe speech disorders. There was also a significant interaction between CAI factors and severity categories, $F(2.74, 106.96) = 3.70$, $p = .017$, effect size or partial $\eta^2 = .087$. To further examine this interaction, a series of independent *t* tests was conducted to test the differences between the two severity categories for each CAI factor separately. Significantly lower factor scores for the severe versus moderate groups were found for PN, $t(39) = 3.62$, $p = .001$; NWI-Seg, $t(39) = 3.21$, $p = .003$; and NWISyll, $t(39) = 3.67$, $p = .001$. No significant difference was found between the mean factor scores of the moderate and severe groups for the CAI factor PWV, $t(39) = 1.11$, $p = .27$.

The second one-way repeated-measures ANOVA was conducted with all six CAI factors, including MRRMono and MRR-BiTri, on 23 children with complete data on these factors (see Table 7). A one-way repeated-measures ANOVA was conducted with these CAI factors: PN, NWI-Seg, NWI-Syll, PWV, MRR-Mono, and MRR-BiTri. Mauchly's test indicated that the assumption of sphericity had been violated, $\chi^2(14) = 32.99$, $p = .003$; therefore, degrees of freedom were corrected using Huynh-Feldt estimates of sphericity ($\varepsilon = .87$). Like the analysis with four factors, the results show that the six factor scores of the CAI were significantly affected by the severity level of the speech disorder; the within-subject factor "CAI factors" was significant, $F(4.3, 90.9) = 6.40$, $p < .001$, effect size or partial $\eta^2 = .23$. The between-subjects factor "severity category" was also significant, $F(1, 21) = 4.60$, $p = .04$, effect size or partial $\eta^2 = .18$, as well as the interaction between CAI factors and severity categories, $F(4.3, 90.9) = 4.17$, $p = .003$, effect size or partial $\eta^2 = .17$. To further examine this interaction, independent *t* tests were conducted to test the differences between the two severity categories for all six factors. For NWI-Syll, $t(21) = 2.61$, $p = .016$, and MRR-BiTri, $t(0.0) = 2.35$, $p = .043$, the differences between the mean factor scores of the moderate and severe groups reached significance. No significant difference was found between the severity groups for PWV. For PN and NWI-Seg, the differences were only marginally significant in this second analysis, most likely due to less power as compared to the first analysis. It is remarkable that there is no difference between the moderate and severe groups for MRR-Mono, but there is a large significant difference for MRR-BiTri. We will come back to this issue in the general discussion.

Table 8 shows correlations between severity category and CAI factors. Moderate, significant correlations were found between severity category and PN, NWI-Seg, and NWI-Syll. Children with a severe disorder had lower CAI factor scores. The factor scores of PN, NWI-Seg, and NWI-Syll showed strong correlations; the correlations with PWV and MRR-BiTri were weak to moderate. No significant correlations were found between MRR-Mono and any other CAI factor.

Table 6. Means and standard deviations of the factor scores of four Computer Articulation Instrument factors per severity category.

Severity category		**PN**	**NWI-Seg**	**NWI-Syll**	**PWV**
Moderate	*N*	14	14	14	14
	M	-1.07	-0.88	-0.18	-1.03
	SD	0.52	0.83	0.75	0.81
Severe	*N*	27	27	27	27
	M	-1.72	-1.69	-1.13	-1.29
	SD	0.56	0.73	0.80	0.67
Total	*N*	41	41	41	41
	M	-1.45	-1.42	-0.81	-1.20
	SD	0.62	0.85	0.89	0.72

Note. PN = factor score of all parameters of picture naming; NWI-Seg = factor score of the segmental parameters of nonword imitation; NWI-Syll = factor score of the syllable structure parameters of nonword imitation; PWV = factor score of the two PWV parameters of word and nonword repetition.

Table 7. Means and standard deviations of the factor scores of six Computer Articulation Instrument factors per severity category.

Severity category		PN	NWI-Seg	NWI-Syll	PWV	MRR-Mono	MRR-BiTri
Moderate	*N*	10	10	10	10	10	10
	M	-1.05	-0.80	-0.29	-0.97	-1.12	-1.15
	SD	0.46	0.69	0.80	0.85	0.89	1.56
Severe	*N*	13	13	13	12	13	13
	M	-1.52	-1.45	-1.14	-1.28	-0.60	-2.31
	SD	0.67	0.84	0.75	0.75	0.85	0.06
Total	*N*	23	23	23	22	23	23
	M	-1.31	-1.17	-0.77	-1.14	-0.82	-1.81
	SD	0.62	0.83	0.87	0.80	0.89	1.16

Note. PN = factor score of all parameters of picture naming; NWI-Seg = factor score of the segmental parameters of nonword imitation; NWI-Syll = factor score of the syllable structure parameters of nonword imitation; PWV = factor score of the two PWV parameters of word and nonword repetition; MRR-Mono = factor score of the monosyllabic items of maximum repetition rate parameters; MRR-BiTri = factor score of the bisyllabic and trisyllabic items of maximum repetition rate parameters.

Table 8. Spearman rank correlations and Pearson correlations between severity category and Computer Articulation Instrument factors and between Computer Articulation Instrument factors.

Severity category and CAI factors		Severity category	PN	NWI-Seg	NWI-Syll	PWV	MRR-Mono	MRR-BiTri
	N	**41**	**41**	**41**	**41**	**41**	**23**	**23**
Severity category	Spearman *r*	1	-.53**	-.50**	-.50**	-.19	.28	-.32
PN	Pearson *r*	—	1	.80**	.81**	.39*	-.09	.41*
NWI-Seg	Pearson *r*		—	1	.68**	.60**	.12	.53*
NWI-Syll	Pearson *r*			—	1	.51**	-.03	.44*
PWV	Pearson *r*				—	1	-.07	.21
MRR-Mono	Pearson *r*					—	1	-.02
MRR-BiTri	Pearson *r*						—	1

Note. NVIQ = nonverbal intelligence quotient; QPPVT = Peabody Picture Vocabulary Test, vocabulary quotient; RLQ = CAI = computer articulation instrument; PN = factor score of all parameters of picture naming; NWI-Seg = factor score of the segmental parameters of nonword imitation; NWI-Syll = factor score of the syllable structure parameters of nonword imitation; PWV = factor score of the two PWV parameters of word and nonword repetition; MRR-Mono = factor score of the monosyllabic items of maximum repetition rate parameters; MRR-BiTri = factor score of the bisyllabic and trisyllabic items of maximum repetition rate parameters.

*Correlation of factor scores is significant at the .05 level (two-tailed). **Correlation of factor scores is significant at the .01 level (two-tailed).

Discussion

The CAI is a computer-based assessment for speech production with a range of speech tasks that reflect different levels of processing (phonological and speech motor skills), and it provides normative data based on a sample of 1,524 children in the age range of 2;0–6;11. A previous study on psychometric characteristics of the CAI revealed sufficient interrater reliability, test–retest reliability, and construct validity (van Haaften et al., 2019). In this current article, we report known-group validity, based on the outcome of two studies in children with speech language impairment and SSDs.

The known-group validity of the CAI was supported by the results of Study 1. These results confirm the hypothesis that PN-PCC-Q is significantly affected by intelligibility level. There was a significant difference between the intelligibility levels with respect to the PCC parameter of the PN task of the CAI, and there was a highly significant correlation between the intelligibility levels and PN-PCC-Q in the expected direction. Correlations between PCC and intelligibility measures were also found in previous studies (Lagerberg et al., 2015; McLeod, Harrison, & McCormack, 2012; Neumann, Rietz, & Stenneken, 2017). In the study of McLeod et al. (2012), significant correlations were found between PCC (measured with the Phonology subtest of the Diagnostic Evaluation of Articulation and Phonology) and the outcome of the Intelligibility in Context Scale. Unfortunately, the Intelligibility in Context Scale could not be administered in our study, because the children in Study 1 fell out of its age range (too young). Therefore, the intelligibility was scored by the SLPs on a scale with three levels: good, moderate, and poor. In Study 1 and Study 2, subjective judgments of SLPs with ordinal scales were used. Due to this subjectivity, no optimal objective measurements were collected, which is a limitation of this study. No reliability measures are reported for these scales. However, it is a common way to judge children's speech, and they are used in several other studies (Gordon-Brannan & Hodson, 2000; Lohmander et al., 2016). Further validation studies are needed to corroborate the diagnostic value of the CAI. This study with "expert judgment" is the first step in this validation process. Different studies describe that experienced listeners tend to give higher intelligibility ratings than inexperienced listeners (Doyle, Swift, & Haaf, 1989; Landa et al., 2014). In the current study, the ratings were assigned by SLPs who are experienced listeners. As a consequence, the rating "poor intelligibility" must be considered as an indication of a serious speech difficulty. It emphasizes the validity of the strongly related parameter PN-PCC-Q. The results of our study showed a quite stable pattern of nonverbal intelligence and language scores in the children with a speech language impairment across intelligibility levels. Intelligibility level shows no or only a very weak, nonsignificant correlation with the outcomes on the nonverbal intelligence and language tests; similarly, no or a very weak, nonsignificant correlation was found between PN-PCC-Q and the outcomes on the nonverbal intelligence and language tests. The results of these correlations show that the PCC of PN of the CAI measures a distinct aspect of the language domain. This corresponds to the subtypes described by Van Weerdenburg et al. (2006), in which children with an SSD are one of the four distinct subtypes.

Study 2 supports the diagnostic power of the CAI factors in a group of children with SSDs. All children, with either a moderate or severe SSD, showed scores below average on the CAI factors PN, NWI-Seg and NWI-Syll, PWV, MRR-Mono, and MRR-BiTri, with mean factor scores being between -0.77 and -1.81.

Comparison of four CAI factors (without MRR) revealed significant differences among these factors and between the two severity categories. The severity of the speech disorder is mainly expressed in the parameters of PN and NWI, as shown by the significant difference between the moderate and severe groups for the CAI factors PN, NWI-Seg, and NWI-Syll, whereas PWV is stable across the two groups. These results suggest that especially PN and NWI are the most sensitive tasks to diagnose SSDs. This is in line with other authors who stated that NWI, in which articulatory competence is tested separately from lexical knowledge, is an important part of an assessment battery for children with SSDs (Vance, Stackhouse, & Wells, 2005). Other authors have also suggested to not only use PN in a speech assessment but also include an NWI task to gain better insight in the speech production of a child (Geronikou & Rees, 2016; Hodges, Baker, Munro, & McGregor, 2017). NWI is also associated with phonological short-term memory (Gathercole, 2006). Poor performance on NWI can be influenced by difficulties with phonological short-term memory and not just speech production difficulties. Krishnan et al. (2017) suggest that NWI skills have a unique role in the process of remembering and reproducing novel words. They found that NWI abilities were associated with oromotor praxis, reading fluency, and audiovisual sequence reproduction accuracy. The finding that PWV is relatively stable across severity groups might be related to the multiple origins of inconsistency. As elaborated in the introduction, inconsistency could indicate unstable lexical representations, an unstable phonological system, or unstable motor planning as is typical for CAS.

When all six CAI factors were compared (including MRR), significant differences were found among the six factors and the two severity categories. Differences between the moderate and severe groups were found for PN, NWI-Seg, NWI-Syll, and MRR-BiTri. Remarkably, no difference between the moderate and severe groups was found for MRR-Mono, whereas there was a significant difference between the moderate and severe groups for MRR-BiTri. The severe group showed the lowest *z* score for MRR-BiTri (-2.31) when compared with the other CAI factors. These results imply that MRR-BiTri is an important factor in diagnosing SSDs, such as PN and NWI. MRR-BiTri is especially useful in differential diagnosis of SSDs with a motor origin (CAS and dysarthria), as mentioned in other studies (Rvachew, Hodge, & Ohberg, 2005; Thoonen, Maassen, Wit, Gabreëls, & Schreuder, 1996). The fact that PN, NWI, and MRR-BiTri of the CAI were the most affected in the severe speech disorder group underlines the importance of these tasks in diagnosing SSDs. No differences between the two severity groups were found for the factors PWV and MRR-Mono. They correlate less with the SLPs' judgments of severity than the other factors. Nevertheless, the mean factor scores are below average in the SSD groups as compared to typically developing children with the same age. This indicates that these tasks do contribute to the diagnostic differentiation between typical and atypical development. In studies on speech development, speech variability, as assessed with the WR and NWR tasks, has been found to be relatively high in young typically developing children (2- and 3-year-olds; Sosa, 2015), and such variability decreases with age (Holm, Crosble, & Dodd, 2007). In a previous study (van Haaften et al., 2019), we also found minor decreases of the PWV with age. Increased variability has also been associated with certain types of speech disorders, such as CAS (Davis et al., 1998; Dodd, 1995b; Forrest, 2003; Holm et al., 2007; Iuzzini-Seigel et al., 2017) and inconsistent PDs (Dodd, 1995b). In this study, PWV shows a mean below-average factor score and a moderate to strong correlation (.39–.60) to the PN and NWI factors, although the PWV scores for moderate and severe disorders do not differ. To get a better understanding of these complex relations, a scatter plot of PWV and NWI-Seg factor scores was made (see Figure 2). Regression lines show a small difference in PWV between moderate and severe disorders; interestingly, for both severity groups, the correlation with NWI-Seg is equally strong. This suggests that PWV can serve as a diagnostic marker for SSDs; validation studies with other speech and language diagnoses need to be conducted.

MRR performance of monosyllabic sequences shows no relation with the other task parameters, suggesting that MRR-Mono assesses an independent aspect of speech production. This is in accordance with such studies as the one by Staiger, Schölderle, Brendel, Bötzel, and Ziegler (2017), who concluded, from factor analyses of speech data from patients with neurological movement disorders as compared to control subjects, that speech tasks and oral motor tasks such as rapid syllable repetition measure separate traits. Krishnan et al. (2017) studied the correlation between NWI and other tasks. They also found no correlation between MRR-Mono and NWI, whereas an alternate MRR task (such as MRR-BiTri) correlated significantly with NWI. From the perspective of a process-oriented approach, Maassen and Terband (2015) argued that MRR, being a pure motor task that does not require any knowledge of words, syllables, or phonemes, can be used to assess speech motor skills. Still, like PWV, mean MRR-Mono factor scores are below the population average and thus, like PWV, might serve as a diagnostic marker for SSDs. However, in contrast to PWV, MRR-Mono does not correlate with severity. Further studies are needed to delineate the role of the purely repetitive (MRR-Mono) and sequential (MRR-BiTri) variants in SSDs.

This study yields strong indications that comparison of the performance on the different speech tasks of the CAI provides information on the underlying speech processing difficulties of children with SSDs. Interestingly, the children with SSDs show a distinct factor structure, which differs from that of the normative study. As mentioned in the introduction, in the normative study on 1,524 typically developing children, weak and very weak correlations between factor scores were found, from which it can be concluded that the CAI factors represent independent components of the speech production process. Aligned with psycholinguistic models, such as Levelt's model, the current study describes the speech profile of a group of children with SSDs by conducting different speech tasks covering all different speech processes (phonological and speech motor skills). A limitation of this study is the use of a heterogeneous group of children with SSDs, without analyzing the results of different subgroups. This is an important next step in process-oriented diagnostics. The crucial statistical remark to be made here is that factor analysis is based not on average skills but on variability in skills and especially covariance. It can be argued that, in a typical population, variability in skills is not caused by specific underlying factors but rather reflects random noise. In contrast, in an atypical population such as children with SSDs, underlying deficits can cause large covariance if task requirements show overlap; analyzing this structure of overlapping and nonoverlapping task performances is the first step in process-oriented diagnostics. Future investigations are needed to compare subgroups of children with different types of SSDs, such that more profiles of CAI factors can be determined to further reveal the proximal causes of SSDs.

Following the results of the study, the most important implication for clinical practice is to distinguish typical speech development from atypical speech development by the administration of different speech tasks, such as incorporated in the CAI. This allows for process-oriented diagnostics, which is important for targeted intervention in children with SSDs.

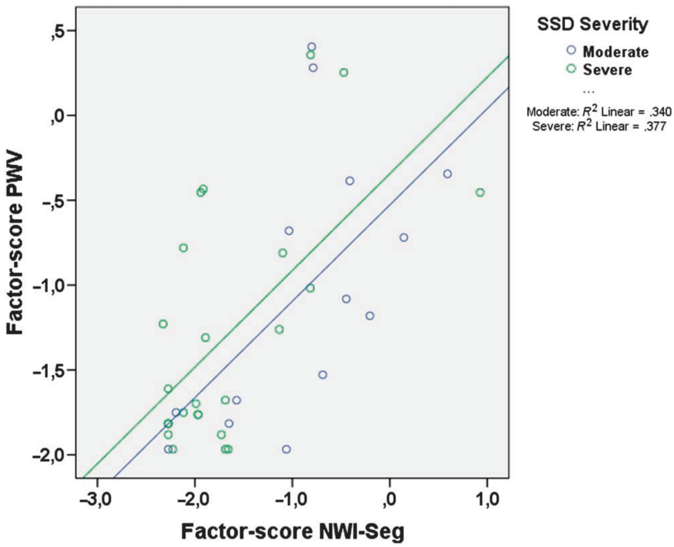

Figure 2. Scatter plot of the segmental quality of nonword imitation (NWI-Seg) and word and nonword proportion of whole-word variability (PWV factor scores), showing the correlations for both groups of children with moderate and severe speech sound disorders (SSDs). Although the difference in PWV between the two groups is small, the correlations with NWI-Seg are moderate to strong.

Acknowledgments

The authors did not receive any grant funding for this study, but this research was partially supported by Het Revalidatiefonds (The Fund for Rehabilitation). The authors wish to thank all the children who participated, their parents, and all the SLPs for their valuable help with the administration and scoring of the tests.

References

Baker, E., Croot, K.,McLeod, S., & Paul, R. (2001). Psycholinguistic models of speech development and their application to clinical practice. *Journal of Speech, Language, and Hearing Research, 44*(3), 685–702.

Bishop, D. V. M., Snowling, M. J., Thompson, P. A., et al. (2017). Phase 2 of CATALISE: A multinational and multidisciplinary Delphi consensus study of problems with language development: Terminology. *The Journal of Child Psychology and Psychiatry, 58*(10), 1068–1080.

Davis, B. L., Jakielski, K. J., & Marquardt, T. P. (1998). Developmental apraxia of speech: Determiners of differential diagnosis. *Clinical Linguistics & Phonetics, 12*(1), 25–45.

Dodd, B. (1995a). Children's acquisition of phonology. In B. Dodd (Ed.), *Differential diagnosis and treatment of children with speech disorder* (pp. 21–48). London, England: Whurr.

Dodd, B. (1995b). Children with speech disorder: Defining the problem. In B. Dodd (Ed.), *Differential diagnosis and treatment of children with speech disorder.* (pp. 1–19). London, England: Whurr.

Dodd, B. (1995c). Procedures for classification of subgroups of speech disorder. In B. Dodd (Ed.), *Differential diagnosis and treatment of children with speech disorder* (pp. 49–64). London, England: Whurr.

Dodd, B. (2011). Differentiating speech delay from disorder: Does it matter? *Topics in Language Disorders, 31*(2), 96–111.

Dodd, B. (2014). Differential diagnosis of pediatric speech sound disorder. *Current Developmental Disorders Reports, 1*(3), 189–196.

Doyle, P. C., Swift, E. R., & Haaf, R. G. (1989). Effects of listener sophistication on judgments of tracheoesophageal talker intelligibility. *Journal of Communication Disorders, 22*(2), 105–113.

Eadie, P., Morgan, A., Ukoumunne, O. C., Ttofari Eecen, K., Wake, M., & Reilly, S. (2015). Speech sound disorder at 4 years: Prevalence, comorbidities, and predictors in a community cohort of children. *Developmental Medicine & Child Neurology, 57*(6), 578–584.

Forrest, K. (2003). Diagnostic criteria of developmental apraxia of speech used by clinical speech-language pathologists. *American Journal of Speech-Language Pathology, 12*(3), 376–380.

Gathercole, S. E. (2006). Nonword repetition and word learning: The nature of the relationship. *Applied Psycholinguistics, 27*(4), 513.

Geronikou, E., & Rees, R. (2016). Psycholinguistic profiling reveals underlying impairments for Greek children with speech disorders. *Child Language Teaching and Therapy, 32*(1), 95–110.

Gordon-Brannan, M., & Hodson, B. (2000). Intelligibility/severity measurements of prekindergarten children's speech. *American Journal of Speech-Language Pathology, 9*(2), 141–150.
Hodges, R., Baker, E., Munro, N., & McGregor, K. K. (2017). Responses made by late talkers and typically developing toddlers during speech assessments. *International Journal of Speech-Language Pathology, 19*(6), 587–600.
Holm, A., Crosble, S., & Dodd, B. (2007). Differentiating normal variability from inconsistency in children's speech: Normative data. *International Journal of Language & Communication Disorders, 42*(4), 467–486.
Iuzzini-Seigel, J., Hogan, T. P., & Green, J. R. (2017). Speech inconsistency in children with childhood apraxia of speech, language impairment, and speech delay: Depends on the stimuli. *Journal of Speech, Language, and Hearing Research, 60*(5), 1194–1210.
Krishnan, S., Alcock, K. J., Carey, D., Bergström, L., Karmiloff-Smith, A., & Dick, F. (2017). Fractionating nonword repetition: The contributions of short-term memory and oromotor praxis are different. *PLOS ONE, 12*(7), 1–18.
Lagerberg, T. B., Hartelius, L., Johnels, J. Å., Ahlman, A.-K., Börjesson, A., & Persson, C. (2015). Swedish Test of Intelligibility for Children (STI-CH)—Validity and reliability of a computer-mediated single word intelligibility test for children. *Clinical Linguistics & Phonetics, 29*(3), 201–215.
Landa, S., Pennington, L., Miller, N., Robson, S., Thompson, V., & Steen, N. (2014). Association between objective measurement of the speech intelligibility of young people with dysarthria and listener ratings of ease of understanding. *International Journal of Speech-Language Pathology, 16*(4), 408–416.
Law, J., Boyle, J., Harris, F., Harkness, A., & Nye, C. (2000). Prevalence and natural history of primary speech and language delay: Findings from a systematic review of the literature. *International Journal of Language & Communication Disorders, 35*(2), 165–188.
Levelt, W. (1989). *Speaking: From intention to articulation.* Cambridge, MA: MIT Press.
Levelt, W., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioural and Brain Sciences, 22*(1), 1–75.
Lohmander, A., Lundeborg, I., & Persson, C. (2016). SVANTE—The Swedish Articulation and Nasality Test—Normative data and a minimum standard set for cross-linguistic comparison. *Clinical Linguistics & Phonetics, 31*, 137–154.
Maassen, B., & Terband, H. (2015). Process-oriented diagnosis of childhood and adult apraxia of speech (CAS and AOS). In M. A., Redford (Ed.), *The handbook of speech production* (1st ed., pp. 331–350). United Kingdom: Wiley.

Maassen, B., van Haaften, L., Diepeveen, S., Terband, H., van den Engel-Hoek, L., Veenker, T., & De Swart, B. (2019). *Computer articulation instrument.* Amsterdam, the Netherlands: Boom Uitgevers.

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*, 27–35.

McLeod, S., Harrison, L. J., & McCormack, J. (2012). The intelligibility in context scale: Validity and reliability of a subjective rating measure. *Journal of Speech, Language, and Hearing Research, 55*(2), 648–656.

Mullen, R., & Schooling, T. (2010). The national outcomes measurement system for pediatric speech-language pathology. *Language, Speech, and Hearing Services in Schools, 41*(1), 44–60.

Neumann, S., Rietz, C., & Stenneken, P. (2017). The German Intelligibility in Context Scale (ICS-G): Reliability and validity evidence. *International Journal of Language & Communication Disorders, 52*(5), 585–594.

Nijland, L. (2003). *Developmental apraxia of speech: Deficits in phonetic planning and motor programming.* Wageningen, the Netherlands: Ponson & Looijen.

Nijland, L., Maassen, B., & van der Meulen, S. (2003). Evidence of motor programming deficits in children diagnosed with DAS. *Journal of Speech, Language, and Hearing Research, 46*(2), 437–450.

Nijland, L., Maassen, B., van der Meulen, S., Gabreëls, F., Kraaimaat, F. W., & Schreuder, R. (2003). Planning of syllables in children with developmental apraxia of speech. *Clinical Linguistics & Phonetics, 17*(1), 1–24.

Nijland, L., Terband, H., & Maassen, B. (2015). Cognitive functions in childhood apraxia of speech. *Journal of Speech, Language, and Hearing Research, 58*(3), 550–565.

Portney, L., & Watkins, M. (2009). *Foundations of clinical research: Applications to practice.* Upper Saddle River, NJ: Prentice Hall.

Rvachew, S., Hodge, M., & Ohberg, A. (2005). Obtaining and interpreting maximum performance tasks from children: A tutorial. *Journal of Speech-Language Pathology and Audiology, 29*(4), 146–157.

Schlichting, L. (2005). *Peabody Picture Vocabulary Test–III Dutch version (PPVT-III-NL).* Amsterdam, the Netherlands: Pearson Assessment and Information B.V.

Schlichting, L., & Spelberg, H. L. (2010a). *Schlichting Test voor Taalbegrip* [Schlichting Test for Language Comprehension]. Houten, the Netherlands: Bohn Stafleu van Loghum.

Schlichting, L., & Spelberg, H. L. (2010b). *Schlichting Test voor Taalproductie-II* [Schlichting Test for Language Production]. Houten, the Netherlands: Bohn Stafleu van Loghum.

Shriberg, L. D., & Kwiatkowski, J. (1994). Developmental phonological disorders I: A clinical profile. *Journal of Speech and Hearing Research, 37*(5), 1100–1126.

Shriberg, L. D., Strand, E. A., Fourakis, M., et al. (2017). A diagnostic marker to discriminate childhood apraxia of speech from speech delay: I. Development and description of the pause marker. *Journal of Speech, Language, and Hearing Research, 60*(4), 1096–1117.

Snijders, J. T., Tellegen, P. J., Winkel, M., & Laros, J. A. (2003). *Snijders-Oomen Nonverbal Intelligence Test–Revised SON-R 5–17 jaar*. Lisse, the Netherlands: Swets & Zeitlinger.

Sosa, A. V. (2015). Intraword variability in typical speech development. *American Journal of Speech-Language Pathology, 24*(1), 24–35.

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.

Staiger, A., Schölderle, T., Brendel, B., Bötzel, K., & Ziegler, W. (2017). Oral motor abilities are task dependent: A factor analytic approach to performance rate. *Journal of Motor Behaviour, 49*(5), 482–493.

Terband, H., & Maassen, B. (2012). Childhood apraxia of speech: From symptom based diagnosis to process-oriented diagnosis. *Logopedie en Foniatrie, 84*, 229–234.

Thoonen, G., Maassen, B., Gabreëls, F., & Schreuder, R. (1994). Feature analysis of singleton consonant errors in developmental verbal dyspraxia (DVD). *Journal of Speech and Hearing Research, 37*(6), 1424–1440.

Thoonen, G., Maassen, B., Gabreels, F., & Schreuder, R. (1999). Validity of maximum performance tasks to diagnose motor speech disorders in children. *Clinical Linguistics & Phonetics, 13*(1), 1–23.

Thoonen, G., Maassen, B., Wit, J., Gabreëls, F., & Schreuder, R. (1996). The integrated use of maximum performance tasks in differential diagnostic evaluations among children with motor speech disorders. *Clinical Linguistics & Phonetics, 10*(4), 311–336.

Van Haaften, L., Diepeveen, S., Van den Engel-Hoek, L., Jonker, M., de Swart, B., & Maassen, B. (2019). The psychometric evaluation of a speech production test battery for children: The reliability and validity of the computer articulation instrument (CAI). *Journal of Speech, Language, and Hearing Research, 62*, 2141-2170.

Van Weerdenburg, M., Verhoeven, L., & Van Balkom, H. (2006). Towards a typology of specific language impairment. *The Journal of Child Psychology and Psychiatry, 47*(2), 176–189.

Vance, M., Stackhouse, J., & Wells, B. (2005). Speech-production skills in children aged 3–7 years. *International Journal of Language & Communication Disorders, 40*(1), 29–48.

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 & Communication Disorders, 48*(1), 25–40.

SIX

CHAPTER 6

EARLY SPEECH DEVELOPMENT

IN KOOLEN DE VRIES SYNDROME LIMITED BY ORAL PRAXIS AND HYPOTONIA

Angela T. Morgan, Leenke van Haaften, Karen van Hulst, Carol Edley, Cristina Mei, Tiong Yang Tan, David Amor, Simon E. Fisher and David A. Koolen

PUBLISHED IN
EUROPEAN JOURNAL OF HUMAN GENETICS, 2018; 26;75-84

Abstract

Communication disorder is common in Koolen de Vries syndrome (KdVS), yet its specific symptomatology has not been examined, limiting prognostic counselling and application of targeted therapies. Here we examine the communication phenotype associated with KdVS. Twenty-nine participants (12 males, 4 with KANSL1 variants, 25 with 17q21.31 microdeletion), aged 1.0–27.0 years were assessed for oral-motor, speech, language, literacy, and social functioning. Early history included hypotonia and feeding difficulties. Speech and language development was delayed and atypical from onset of first words (2; 5–3; 5 years of age on average). Speech was characterised by apraxia (100%) and dysarthria (93%), with stuttering in some (17%). Speech therapy and multi-modal communication (e.g., sign-language) was critical in preschool. Receptive and expressive language abilities were typically commensurate (79%), both being severely affected relative to peers. Children were sociable with a desire to communicate, although some (36%) had pragmatic impairments in domains, where higher-level language was required. A common phenotype was identified, including an overriding 'double hit' of oral hypotonia and apraxia in infancy and preschool, associated with severely delayed speech development. Remarkably however, speech prognosis was positive; apraxia resolved, and although dysarthria persisted, children were intelligible by mid-to-late childhood. In contrast, language and literacy deficits persisted, and pragmatic deficits were apparent. Children with KdVS require early, intensive, speech motor and language therapy, with targeted literacy and social language interventions as developmentally appropriate. Greater understanding of the linguistic phenotype may help unravel the relevance of KANSL1 to child speech and language development.

Introduction

Koolen de Vries syndrome (KdVS; MIM 610443) is a multi-system disorder caused by haploinsufficiency of *KANSL1*, either due to a 17q21.31 microdeletion or intragenic variant (Koolen, Kramer, & Neveling et al., 2012; Koolen, Sharp, & Hurst et al., 2008; Koolen, Vissers, & Pfundt et al., 2006; Zollino, Marangi, & Ponzi et al., 2015). The prevalence is estimated at 1 in 130,000 to 1 in 20,000 (Egger, Wingbermuhle, & Verhoeven et al., 2013; Koolen et al., 2008). Key phenotypic features include developmental delay, intellectual disability, hypotonia, facial dysmorphism; specifically upslanting palpebral fissures, epicanthal folds, a pear-shaped nose with bulbous nasal tip, and eversion of the lower lip (Egger et al., 2013; Koolen et al., 2012; Koolen, Pfundt, & Beunders et al., 2016; Tan, Aftimos, & Worgan et al., 2009). There is commonly central nervous system involvement of epilepsy (≈50% of individuals). There is commonly central nervous system involvement of epilepsy (≈50% of individuals) and brain anomalies on MRI (e.g., corpus callosum, hydrocephalus) (Koolen et al., 2016). Co-occurring medical features include recurrent joint sublaxation, urogenital, renal, and cardiac defects, and visual deficits, such as exotropia or strabismus (Koolen et al., 2016). Cleft palate and hearing loss (conductive or sensorineural) may also occur, but are less common (Koolen et al., 2016; Tan et al., 2009).

Communication deficits have also been observed as part of the complex profile seen in KdVS (Bernardo, Madia, & Santulli et al., 2016; Keen, Samango-Sprouse, Dubbs, & Zackai, 2017; Koolen et al., 2016). Based on a limited number of case reports, expressive communication is suggested to be severely impaired in the preschool years, characterised by a striking late onset of first words (as late as 3 years of age) and a need for therapy in both verbal and nonverbal domains (e.g., sign language, aided communication, such as computer touch screens) (Bernardo et al., 2016; Egger et al., 2013; Koolen & de Vries, 2010; Koolen et al., 2016). Anecdotally, expressive speech and language abilities are more severely impaired than receptive language abilities or more generally, motor skills (Koolen et al., 2016). However, one study found commensurate expressive and receptive language skills in two of three young adults examined, with only one of the three having better receptive language (Egger et al., 2013).

Information about social skills in individuals with KdVS is limited. In the three adults described by Egger et al., participants showed a relatively strong memory for social-contextual information, appropriate emotion perception, less social fear, more approaching behaviour, and a high level of frustration tolerance. The authors concluded social skills were a relative strength for children with KdVS as also seen in Angelman (15q11-q13) and Williams–Beuren (7q11.23) syndromes. Nonetheless, social skills encompass a broad range of areas beyond those examined in KdVS cases to date, including pragmatic language abilities of initiation, nonverbal

Table 1. Participant developmental history and medical characteristics.

Case	Age (y;m)	Variant	Origin	Sex	Hypotonia	Feeding impairment	Visual impairment	Seizures	Cognitive impairment	Speech therapy	Motor apraxia/delay	OT/PT
1	1;0	17q21 del	AU	F	Y	Y	N	NR	NR	Y	NR	N
2	2;0	17q21 del	US	M	Y	Y	Y	Y	NR	Y	Y	Y
3	2;5	17q21 del	US	F	Y	Y	Y	NR	Y	Y	Y	Y
4	2;5	17q21 del	US	F	Y	Y	Y	Y	NR	Y	Y	N
5	2;7	17q21 del	AU	F	NR	NR	N	NR	NR	Y	Y	N
6	2;9	17q21 del	AU	M	Y	NR	N	Y	NR	Y	NR	N
7	3;0	17q21 del	NZ	M	Y	Y	N	Y	NR	Y	Y	N
8	3;1	17q21 del	AU	M	Y	Y	[a]	Y	Y	Y	Y	Y
9	3;4	17q21 del	AU	F	NR	NR	N	Y	NR	Y	NR	Y
10	3;7	17q21 del	US	M	NR	Y	N	NR	NR	Y	NR	Y
11	3;8	17q21 del	US	F	Y	Y	Y	N	NR	Y	Y	Y
12	4;8	17q21 del	Brazil	M	Y	Y	N	Y	Y	Y	Y	Y
13	4;11	17q21 del	AU	F	Y	Y	N	Y	NR	Y	Y	Y
14	5;6	17q21 del	US	F	NR	NR	N	NR	NR	Y	NR	Y
15	5;8	17q21 del	US	F	Y	Y	N	Y	Y	Y	Y	Y
16	7;3	c.531_540del, p. (Gly179Leufs)	US	F	Y	Y	Y	Y	Y	Y	Y	Y
17	9;0	17q21 del	US	M	Y	NR	N	NR	Y	Y	Y	N
18	9;11	17q21 del	US	M	Y	Y	N	Y	Y	Y	Y	Y
19	10;6	c.1816C>T, p. (Arg606Ter)	US	F	Y	Y	Y	Y	N	Y	Y	Y

20	10;10	17q21 del	US	M	Y	Y	N	Y	N	Y	Y	Y
21	11;7	17q21 del	US	F	Y	Y	N	NR	Y	Y	Y	Y
22	12;3	17q21 del	AU	F	Y	Y	Y	Y	Y	N	Y	N
23	12;5	c.2699_2702dup, p.(Ser901Argfs)	Nederlands	M	Y	Y	Y	Y	Y	Y	Y	Y
24	12;8	17q21 del	AU	M	NR	NR	N	Y	Y	Y	NR	N
25	15;6	17q21 del	Nederlands	F	Y	Y	Y	N	Y	Y	Y	Y
26	16;11	c.1652+1G>A, p. (?)	Nederlands	M	Y	Y	Y	N	Y	Y	Y	NR
27	21;0	17q21 del	US	F	Y	Y	N	Y	Y	N	Y	Y
28	25;11	17q21 del	Nederlands	F	Y	Y	Y	Y	Y	Y	Y	Y
29	27;0	17q21 del	AU	F	Y	Y	Y	NR	Y	Y	NR	Y

Amino-acid positions are provided according to the NM_001193466.1 transcript and the NP_001180395.1 isoform.17q21 del refers to the recurrent ~600 kb deletion at 17q21.31, defined as chr17:g.(43582682_43868942)_(44110194_44479336)del (hg19)

AU Australia, *US* United States, *NR* not reported in collated health professional reports, *Y* feature present, *N* feature absent; Speech therapy from onset of first words to current age; *OT* occupational therapy, *PT* physiotherapy

[a]Yet to have vision tested

6

communication, social relations, interests, and context. An evaluation of pragmatic social language abilities has not yet been carried out in a cohort with KdVS, and as such it remains unknown whether features of autism spectrum disorder (ASD) are associated with the syndrome.

As elucidated here, current evidence for the communication phenotype associated with KdVS is based on case studies only. There has been no cohort study in this field, limiting understanding of homogeneity of phenotype and features most closely associated with *KANSL1*. Further, there has been no systematic examination of specific diagnoses or severity of involvement across speech (e.g., articulation, dysarthria, apraxia), language (e.g., expressive and receptive abilities) and literacy (e.g., reading and spelling profiles). The lack of a well-defined phenotype limits current prognostic counselling for speech and language outcomes in this syndrome, and prevents efficient application of targeted therapies to newly presenting affected children.

Here, we conduct the first prospective study of oral-motor, speech, language, literacy, and pragmatic social skills in a large cohort of unrelated children with KdVS, using standardised tests normed for typical behaviour, to precisely characterise the communication phenotype associated with this syndrome.

Methods

Inclusion criteria were a confirmed diagnosis of KdVS (chromosome 17q21.31 microdeletion or *KANSL1* variants) and aged ≥1;0 year (Table 1). Participants were ascertained via a parent support group website (*http://www.supportingkdvs.com*); a clinical-research website (*http://www.17q21.com/en/*) relating to KdVS and Victorian Clinical Genetics Services; a statewide clinical genetics service based in Melbourne, Australia. Ethics approval for the study was obtained from the Royal Children's Hospital, Melbourne, Human Research Ethics Committee (HREC 27053).

Twenty-nine participants (12 males), aged between 1.0 and 25 years 11 months took part in the study. The majority of participants (n = 25) had the common ≈600 kb deletion at 17q21.31 (Koolen & de Vries, 2010) and the remainder of the group (n = 4) had nonsense *KANSL1* variants (Table 1). Children were recruited internationally (14 US, 9AU, 4 Netherlands, 1 New Zealand, 1 Brazil). Local treating speech pathology clinicians completed a pre-determined protocol examining oral motor structure and function, speech, language and pragmatic social skills functioning as outlined below. Standardised tests were administered and scored relative to normative data, in line with the respective test manuals. The same tests were used where both Dutch and English versions were available. In the absence of the same standardised speech assessments in Brazil, this child's performance across speech, language, literacy and social skills domains was reported by his local treating speech pathologist with reference to local normative data.

The 17q21.31 deletions and phenotypic data were submitted to the Decipher database (*https://decipher.sanger.ac.uk/*) and *KANSL1* sequence variants and phenotypic data were submitted to the Clin Var Database (*https://www.ncbi.nlm.nih.gov/clinvar/*).

Developmental history and co-occurring health conditions

Data were collected on genotype (*KANSL1* variant or chromosome 17q21.31 microdeletion), development (e.g., intellectual quotient, first words, feeding history, motor milestones), co-occurring medical features (e.g., laryngomalacia, hypotonia, epilepsy, neurological MRI results, hearing, vision, cleft lip/palate, dysmorphic features, renal, cardiac, urogenital), presence of neurodevelopmental conditions (e.g., ASD, attention deficit hyperactivity disorder), and type and amount of therapeutic input (e.g., speech therapy, occupational therapy) (Table 1). Data was collected from relevant health practitioner reports (e.g., clinical geneticists, audiologists, optometrists, neurologists, craniofacial specialists, speech therapists). The denominators used to determine the proportion of affected cases reflect available data.

Oral-motor

Structural or functional impairments of the oral region were assessed with the Clinical Assessment of Oropharyngeal Motor Development in Young Children (Robbins & Klee, 1987) for children aged ≥2.0 years. This tool examines oral-facial structural integrity (e.g., symmetry, occlusion, size of facial features, height of palatal vault, dental alignment/gaps/decay) and oral motor function (e.g., seal of the lips, fasciculations/atrophy/furrowing of the tongue, ability to retract and protrude the tongue). This tool was administered in the local language (i.e., Dutch, English, or Portugese). Oral motor structure and function performance does not vary across linguistically diverse groups. The Schedule for Oral Motor Assessment (Reilly, Skuse, & Wolke, 1999) examined oral motor structure and function in one participant aged <2.0 years.

Speech measures

The standardised Goldman–Fristoe Test of Articulation (sounds-in-words subtest and stimulability probe) (GFTA- 2) (Goldman & Fristoe, 2000) was administered in English-speaking children with sufficient verbal production at the single world level. The Dutch-normed computer articulation instrument (CAI) was administered (Maassen et al., 2019) to Dutch children. Phonological process analysis was conducted on the GFTA-2 and CAI productions to differentially diagnose articulation (movement plan and motor production of the sound) ability and phonological performance (a child’s understanding of sound rules of their language) (Dodd

& Morgan, 2017). Where children had sufficient speech, a 5- min conversational sample was obtained, and analysed using pre-determined cross-linguistically valid protocols for childhood apraxia of speech (CAS) (Fedorenko, Morgan, & Murray et al., 2016) and dysarthria (Fedorenko et al., 2016; Morgan & Liégeois, 2010; Morgan, Liégeois, & Liederkerke et al., 2011; Morgan, Mei, & Da Costa et al., 2015).

Language measures

The Preschool Language Fundamentals-5 (Zimmerman, Steiner, & Pond, 2011) (PLS-5), the Clinical Evaluation of Language Fundaments (CELF)-Preschool 2 (Wiig, Secord, & Semel, 2006) or CELF-IV, (Semel, Wiig, & Secord, 2003) were most commonly used to assess language; depending on age of the participant and availability of the tool for the clinician. The PLS-5, is a test of receptive and expressive language for children aged 0–7 years 11 months. Standard scores were obtained for the auditory comprehension (receptive language) and expressive communication (expressive language) subscales. The CELF-IV (age range 5–21 years) and CELF-P2 (3–6 years) also provide standardised expressive and receptive language summary scores. Dutch participants were tested with the CELF—Dutch version (Kort, Schittekatte, & Compaan, 2008), or the Schlichting test for language comprehension and production (Schlichting & Lutje Spelberg, 2010) (n = 2). The Schlichting test also provides receptive and expressive language scores. All of the language tools described here examine similar domains and have a mean score of 100 (SD 15), with a score of 85–115 representing average range performance; with language severity as follows: mild (1–1.5SD below mean), moderate (1.5–2SD below mean) and severe (>2SD below mean). One adult was assessed with the Mt. Wilga High Level Language test (Christie, Clark, & Mortensen, 1986).

Literacy

The Wide Range Achievement Test—Fourth Edition (WRAT-4) word reading and spelling subtests were administered (Wilkinson & Robertson, 2006) to English-speaking children aged ≥ 5 years, with standard scores (mean = 100, SD = 15) and equivalent severity ratings as for language above (Wilkinson & Robertson, 2006).

Social skills—pragmatic language

The Children's Communication Checklist—Second Edition (CCC-2) Social Interaction Difference Index (age range 4–16 years) was used to examine verbal and non-verbal social communication skills (Bishop, 2003) in Dutch and English-speaking children. The participant's treating speech pathologist made a subjective clinical rating on social pragmatic abilities relative to peers (appropriate/within normal limits, mildly, moderately or severely affected), where participants did not fulfil the age range for the CCC-2 or where the tool was not available. Formal diagnoses of ASD were recorded.

Results

Developmental history and co-occurring health conditions

Hypotonia was a core deficit (*n* = 24/24; 100%) and related to early feeding difficulties (23/23, 100%), tracheomalacia or laryngomalacia (*n* = 11/11; 100%) and gastroesophageal reflux were regularly seen (*n* = 9/11; 81%) (see Table 1, Supplementary Table 1). Chewing difficulties (*n* = 20/23; 87%) and profuse anterior drooling were common (*n* = 20/ 22; 91%). Drooling resolved in preschool or early school years for most (*n* = 9/20; 45%).

A majority (*n* = 22/22; 100%) had generalised motor delay or disorder due to hypotonia and/or a motor programming (praxis) deficit. Occupational and/or physiotherapy was commonly required (*n* = 20/28; 71%). Motor deficits included difficulties managing buttons and zippers, writing, drawing, using scissors, riding a bike and toilet training.

Non-verbal cognitive impairment (score <85 on standardised tools in neuropsychological reports) was common where examined (*n* = 16/18; 89%). Yet, few children under the age of 5 years (3/13) had received cognitive examinations. Seizures were common (18/21, 86%) and epilepsy confirmed via electroencephalogram in two-thirds of these cases (10/18; 55%). Hearing impairment included mild and mild–moderate sensorineural deafness (ID 18, *n* = 1/29; 3%), and periodic conductive loss associated with otitis media (*n* = 4/29; 14%). All participants with visual impairments wore glasses (*n* = 12/28; 43%) and strabismus was the most common diagnosis. Features seen in only one participant were: hypothyroidism, congenital heart defect, which resolved by 2 years of age; a tethered spinal cord with sacral sinus; malignant melanoma of the forearm; and hepatic dysfunction alongside hypoglycemia and ketosis.

Academically, seven participants (7/18, 39%) attended mainstream schools. The remaining 11 (61%) attended special schools. This proportion is likely influenced by the geographical location of the family (i.e., whether the region supported mainstream schooling or separate special schools), not only the individual child's abilities.

First words were delayed in all but three children, who had appropriate onset of first words at 12 and 13 months, respectively. Most had first words between ages 2.5 and 3.5 years, with two individuals having onset delayed until 5 and 7 years, respectively. All but two (*n* = 26/28, 93%) had received regular speech therapy from the onset of first words until the time of this study, with increased intensity in the preschool period (typically once per week or fortnight, but as much as twice per week where it could be afforded).

Table 2. Speech, oral motor, language, literacy and social skills.

Case	Age	Variant	Dysarthria	Speech apraxia	Oral motor impairment	Expressive language impairment	Receptive language impairment	Reading impairment	Spelling impairment	Social skills impairment
1	1;0	17q21 del	NA	NA	Y	WNL	WNL	NA	NA	WNL
2	2;0	17q21 del	Y	Y	Y	Severe	Severe	NA	NA	NR
3	2;5	17q21 del	NA	Y	Y	Severe	Severe	NA	NA	NR
4	2;5	17q21 del	NA	Y	Y	Mild	WNL	NA	NA	Appropriate
5	2;7	17q21 del	NA	Y	Y	Mild	Mild	NA	NA	NR
6	2;9	17q21 del	NA	Y	Y	Severe	Severe	NA	NA	Appropriate
7	3;0	17q21 del	NA	Y	Y	Severe	Severe	NA	NA	WNL
8	3;1	17q21 del	Y	Y	Y	Severe	Severe	NA	NA	Mild
9	3;4	17q21 del	N	Y	Y	Moderate	Mild	NA	NA	Mild
10	3;7	17q21 del	NR	Y	Y	Severe	Severe	NA	NA	WNL
11	3;8	17q21 del	NR	Y	Y	Mild	WNL	NA	NA	Appropriate
12	4;8	17q21 del	Y	NR	NR	Severe	Severe	NA	NA	Appropriate
13	4;11	17q21 del	Y	Y	Y	Mild	WNL	NA	NA	Moderate[a]
14	5;6	17q21 del	NR	Y	Y	Severe	Severe	NA	NA	NR
15	5;8	17q21 del	NR	NR[b]	NR	Severe	Severe	Mild	Mild	NR
16	7;3	c.531_540del, p. (Gly179Leufs)	Y	NR	Y	Moderate	Moderate	WNL	Mild	Appropriate
17	9;0	17q21 del	Y	Y	Y	Severe	Severe	NA	NA	NR
18	9;11	17q21 del	NR	Y	Y	Severe	Severe	NA	NA	NR

19	10;6	c.1816C>T, p.(Arg606Ter)	NR	Y	Y	Severe	Severe	NA	NA	Moderate
20	10;10	17q21 del	Y	Y	Y	Severe	Moderate	WNL	WNL	NR
21	11;7	17q21 del	Y	Y	Y	WNL	Moderate	Mild	Severe	Appropriate
22	12;3	17q21 del	NR	Y	Y	Severe	Severe	NA	NA	Moderate[a]
23	12;5	c.2699_2702dup, p.(Ser901Argfs)	Y	Y	Y	Severe	Severe	NA	NA	Mild
24	12;8	17q21 del	Y	Y	Y	Severe	Severe	Severe	Severe	Mild
25	15;6	17q21 del	Y	Y	Y	Severe	Severe	NA	NA	Mild
26	16;11	c.1652+1G>A, p. (?)	Y	Y	Y	Severe	Severe	NA	NA	Severe[a]
27	21;0	17q21 del	NR	NR	NR	Severe	Severe	Severe	Moderate	Moderate
28	25;11	17q21 del	Y	Y	Y	Severe	Severe	NA	NA	Moderate
29	27;0	17q21 del	Y	Y	Y	Severe	Severe	Severe	Severe	Moderate

NR not reported by local speech pathologist or collated health professional reports, *NA* not assessed as not developmentally appropriate or not able to be tested, *WNL* within normal limits, *Y* feature present

[a]Autistic traits
[b]Childhood apraxia of speech not reported, but child not using 3–5 word phrases, dropping plurals and word endings, nasal errors, inconsistent speech errors

Communication phenotype

Oral-motor

Three individuals had cleft lip and palate (ID 15, 18, 26). Hypodontia (*n* = 7/7; 100%), macroglossia (*n* = 5/12; 42%) and malocclusion (cross-bite or underbite) (n = 8/13; 62%) were noted in a subset of participants. High arch palate was reported in half of the group (*n* = 8/16, 50%) (Supplementary Table 1).

Abnormal oral-motor function deficits were evident in all participants assessed (*n* = 26/26, 100%) to some degree. Specifically, reduced range and precision of single mandibular, labial-facial, laryngeal and lingual movements was noted (e.g., poke out your tongue; blow a kiss). Praxis deficits were equally common across more complex multimovement non-speech oral and speech sequences.

Speech

The most common speech diagnosis was CAS (speech apraxia) (*n* = 24/24, 100%) (Table 2). The speech profile was characterised by exceptionally delayed onset of first words, limited babbling, reduced phonetic inventories (i.e., had not acquired all English sounds) relative to typical peers, more errors on vowels than consonants, inconsistency of errors, addition and omission errors in attempts to simplify syllable structures including cluster reduction, simplified syllable structures relative to age, and prosodic errors. Instances of dysfluency (stuttering), manifesting as syllable, word or phrase level repetitions, were seen in some participants (*n* = 3/18; 17%). A proportion (*n* = 14/15; 93%) had dysarthria, typically characterised by low pitch, hypernasality, monotonous, monoloud and flaccid, slow speech.

Four children (IDs 1–3,29) had few spoken words at assessment, relying on alternative forms of communication such as gesture, Makaton sign and technological supports, such as iPads to support expressive speech and language. The remaining 25 children assessed on single word performance demonstrated articulation (phonetic distortion) errors. Delayed (i.e., e.g., 'stopping' d for th in feader for feather) or atypical (i.e., sound preference substitution) phonological speech sound processes were also present across this group. All the 25 children were reported to have used early sign language, non-verbal gestures or communication devices to supplement or facilitate communication prior to intelligible speech development. Intelligible speech was obtained only after explicit teaching of sound imitation, syllable generation, syllable combinations, increasingly complex words, short phrases, sentences and spontaneous speech. Each stage required extensive work to acquire each skill and significant ongoing follow-up work to maintain the skill. Therapeutic focus emphasised language and literacy at mid-to-late school age, once speech was intelligible and children could fluently produce phrases or sentences. Frequency of speech therapy reduced, however, once children had acquired intelligible speech.

Language

Expressive and receptive language abilities were commensurate in most participants (*n* = 23/29; 79%). In a small subset, expressive performance was lower than receptive (*n* = 5/29; 17%) or vice versa (*n* = 1/29; 3%). Further reliable comparison across linguistic subdomains (e.g., semantics, morphology, syntax) was not possible given the range of tools used to assess language that differed in items elicited across domains, and due to the broad chronological and developmental age range examined. Language abilities were noted as commensurate with cognition by the treating clinicians. There was no clear distinction in language performance in individuals with and without seizures (Tables 1 and 2).

Literacy

Almost half the cohort (*n* = 13/29; 45%) were too young (<5 years) for literacy testing. Testing was not conducted in a further subset determined as developmentally premature for literacy assessment (*n* = 6/29; 21%) or where the clinician did not have access to the WRAT assessment (*n* = 3/ 29; 10%). Both reading and spelling performance was variable in the seven individuals assessed (Table 2). All children with typical or mildly impaired reading skills were school-aged. Two of the three individuals with more severely affected reading, relative to peers, were adults (aged 21, 27 years, respectively).

Social skills—pragmatic language

Only 5/29 (17%) children underwent formalised testing with the CCC-2. Sixteen reports were based on subjective clinician judgement and data were absent for 8/29 (27%) individuals. A range of abilities were reported (Table 2). Autistic traits were seen in few participants (*n* = 3/19, 16%). Traits included sensory skill deficits and the need to follow a consistent routine. No child had a confirmed ASD diagnosis. Overall, children had a keen desire to communicate, good initiation, appropriate turn-taking and intact basic social skills of eye contact and non-verbal gestures. Whilst available data were limited, a widening gap in social skills relative to peers with increasing age was observed (Table 2).

Discussion

Our linguistic phenotyping in a genetically confirmed cohort with KdVS revealed a distinctive communication profile. The most striking feature was the presence of CAS and delayed onset of first words. There was no evidence for better receptive than expressive language. Literacy was commonly impaired and social pragmatic skills were varied. Whilst children have a keen desire to communicate with appropriate eye contact, turn-taking, and non-verbal gestures, higher-level pragmatic language deficits were identified in a subset of the cohort. With only four participants with nonsense intragenic *KANSL1* variants, we are unable to draw definitive genotype–phenotype conclusions, yet no striking communication differences were seen between individuals with intragenic *KANSL1* variants vs. those with the standard 17q21.31 microdeletion, in line with previous reports (Koolen et al., 2016). Group performance and treatment indications for each domain of communication are discussed below.

Oral-motor

Oral-motor dysfunction was pervasive and impacted by both hypotonia and oral praxis. Early feeding issues were influenced differentially across the group by laryngomalacia or tracheomalacia, weak suck due to hypotonia, gastroesophageal reflux generating negative food associations, and poor lip seal due to malocclusion. Drooling was influenced by the degree of hypotonia and presence of macroglossia/malocclusion and would benefit from specific therapies (AACPDM, 2016).

Chewing delays in managing solid or lumpy textures was also influenced by hypotonia, oral-motor praxis, and the presence of reflux causing negative food associations and food refusal. The delayed trajectory of feeding milestones may lead to some children with KdVS missing the 'critical period' for chewing practice or learning to manage solids (Fujishita et al., 2015) as seen in other neurodevelopmental conditions (Sanchez, Spittle, Slattery, & Morgan, 2016). Focused oral feeding interventions could mitigate these issues (van den Engel-Hoek, van Hulst, van Gerven, van Haaften, & de Groot, 2014; Volkert, Piazza, Vaz, & Frese, 2013).

Speech: CAS, dysarthria, articulation and phonological disorder

Speech development was the core challenge in the preschool period. Almost all had CAS, often with flaccid dysarthria, and additional articulation and phonological errors. The presence of CAS with co-occurring speech diagnoses is seen in other syndromes, such as 16p11.2 deletion syndrome (Fedorenko et al., 2016), Floating Harbour Syndrome (White, Morgan, & Da Costa et al., 2010) and 7q11.23 duplication syndrome (Mervis, Morris, Klein-Tasman, Velleman, & Osborne, 2015); although the speech profile in KdVS is arguably more severe by comparison, particularly in the early years. Whether this profile is underpinned by exceptionally delayed myelination or other factors is yet to be determined. Impairment of the broader motor system in KdVS, and/or deficits of the corpus callosum impacting on inter-hemisphere communication are likely factors restricting neuroplasticity and contributing to the protracted period of speech motor development. Yet there is a remarkable ongoing propensity for speech learning, and intelligible speech is acquired by the middle school years. The sociable nature of the children, with their strong desire to communicate and high tolerance for frustration (Egger et al., 2013) are positive indicators for continuing to practice speech and achieve functional outcomes. This is in contrast to other conditions, where children initiate conversation less frequently (e.g., cerebral palsy; Pennington & McConachie, 1999) and where speech may plateau at lower levels of achievement at a younger age.

The priority for clinical management of communication in individuals with KdVS is to manage co-morbidities that may impact speech and language (e.g., optimising hearing, controlling epilepsy, repair of cleft palate and corrected malocclusion). For speech-specific intervention in children younger than 3 years of age with few words, a Core Vocabulary treatment programme could be a suitable approach (Crosbie, Holm, & Dodd, 2005; Iuzzini & Forrest, 2010). For children from 4 years of age with CAS, RCT-supported evidence exists for the Nuffield Dyspraxia Programme Version 3 and the Rapid Syllable Repetition programme, although these therapies have not been trialled in children with ID (Murray, McCabe, & Ballard, 2015). As apraxia resolves and children begin to acquire fluent, consistently intelligible speech, dysarthric features become more apparent and targeted dysarthria treatments may be indicated (Pennington, Parker, Kelly, & Miller, 2016).

Language

Receptive and expressive language abilities were typically commensurate and severely affected in our cohort, in agreement with a previous report (Egger et al., 2013). A lack of data on severity of non-verbal cognition precluded reliable correlational analyses between language and cognitive functioning here. No pattern of relative strengths and weaknesses in language domains was noted (i.e., across semantics, morphology and syntax). With a dearth of language intervention research in other genetic syndromes, let alone KdVS, selection of approaches to trial will likely be guided by intervention studies in the general developmental language disorder literature (Bishop, Snowling, Thompson, & Greenhalgh, 2016a, 2016b; Law, Garrett, & Nye, 2003).

Literacy

The range of reading and spelling abilities seen here are likely the result of differential impairments across skills contributing to literacy, including: impacts of speech sound disorder and language impairment on phonological awareness for literacy; reduced phonological awareness skills due to frequent otitis media (Winskell, 2006) and sensorineural hearing impairments (Park, Lombardino, & Ritter, 2013); motor praxis issues, which may impact on written spelling; and the high prevalence of visual deficits with impacts on visual integration and/or visuo-motor integration for reading and written spelling. Targeted assessment of sound awareness by a speech pathologist, visual ability by an optometrist and visual-integration for literacy by an occupational therapist is critical to tailor an intervention programme appropriate to the individual child. A therapeutic goalsetting challenge will be optimising early precursors to literacy alongside the goal of obtaining fluent, intelligible speech.

Social skills

Individuals with KdVS have been reported as 'hypersociable' due to their desire to communicate (Egger et al., 2013). Indeed our cohort had strengths in initiation of communication, desire to communicate, appropriate eye contact, non-verbal skills and turn-taking. No child had a formal diagnosis of autism and few had autistic traits. Overall, data support prior observations (Egger et al., 2013) that, relative to many genetic intellectual disability syndromes, social skills are a strength in KdVS. Here we extend the social phenotype, demonstrating challenges in narrative/story telling and in providing contextual information. Greater linguistic sophistication is required in social interactions with age, and whilst preliminary, data showed a trend for greater pragmatic impairment with increasing age. Intervention focused on narrative storytelling (Adams, Lockton, & Freed et al., 2012) and provision of context may support limitations in this area.

Limitations and future directions

Speech, oral-motor and language functioning was thoroughly characterised here, using a consistent approach. By contrast, few children were formally assessed for literacy or pragmatic skills. Almost half our cohort was aged <5 years of age, meaning it was inappropriate to measure reading and spelling development. Further, our method of using local clinicians to acquire data was limited in that many therapists did not have access to the literacy or social pragmatic tools specified in our a-priori designed protocol, despite attempts to use universally adapted tools. Nevertheless, our preliminary data will support hypothesis generation for future larger-scale studies in this area. Recruitment bias was also possible in our study that invited participants to take part in a 'speech and language examination', potentially leading to over-estimations of communication deficits in our cohort.

Clinical indications summary

Children with KdVS should be enroled in speech therapy programmes early in life, in particular with an emphasis on the acquisition of receptive and expressive language alongside tackling the motor programming and motor planning deficits associated with speech apraxia. Implementation of multi-modal communication, such as sign language or communication devices would support language acquisition and social communication development prior to fluent speech developing. Further, therapy should target not only on speech sound production in the early years, but also the understanding of sounds and ability to visually process written text to provide an optimal foundation for reading and spelling development. Finally, narrative language therapy is indicated as developmentally appropriate, to support acquisition of more sophisticated pragmatic language skills.

Acknowledgements

Thanks to the families and clinicians who took part in this project. A special thanks to Mark Mitchell for considering feedback. A.M. is supported by a National Health and Medical Research Council (NHMRC) Practitioner Fellowship #1105008. This work is also supported by the NHMRC Centre of Research Excellence in Speech and Language NeurobioloGy (CRE-SLANG) #1116976 and NHMRC Project grant #APP1127144, awarded to A.M., S.F. This work is also supported by the Victorian Government's Operational Infrastructure Support Programme.

References

American Academy for Cerebral Palsy and Developmental Medicine (AACPDM). (2016). Sialorrhea. Retrieved from https://www.aacpdm.org/UserFiles/file/care-pathways-sialorrhea-print.pdf

Adams, C., Lockton, E., & Freed, J., et al. (2012). The Social Communication Intervention Project: a randomized controlled trial of the effectiveness of speech and language therapy for school-age children who have pragmatic and social communication problems with or without autism spectrum disorder. *International Journal of Language & Communication Disorders, 47,* 233-44.

Bernardo, P., Madia, F., & Santulli, L., et al. (2016). 17q21.31 microdeletion syndrome: description of a case further contributing to the delineation of Koolen-de Vries syndrome. *Brain & Development, 38,* 663-668.

Bishop, D. V. (2003). *Children's communication checklist* (2nd ed.). Sydney: Pearson.

Bishop, D. V., Snowling, M. J., Thompson, P. A., & Greenhalgh, T. (2016a). CATALISE: a multinational and multidisciplinary Delphi consensus study of problems with language development: Terminology. *The Journal of Child Psychology and Psychiatry, 58*(10), 1068-1080.

Bishop, D. V., Snowling, M. J., Thompson, P. A., & Greenhalgh, T. (2016b). Phase 2 of CATALISE: a multinational and multidisciplinary Delphi consensus study. Identifying language impairments in children. *PLoS One, 11*(12), 1-26.

Christie, J., Clark, W., & Mortensen, L. (1986). *Mt. Wilga higher level language assessment*. Sydney: Speech Pathology Department, Mount Wilga Rehabilitation Centre.

Crosbie, S., Holm, A., & Dodd, B. (2005). Intervention for children with severe speech disorder: a comparison of two approaches. *International Journal of Language & Communication Disorders, 40*, 467-491.

Dodd, B., & Morgan, A. (2017). *Intervention case studies of child speech impairment*. London, UK: R Press.

Egger, J. I. M., Wingbermuhle, W. M. A., & Verhoeven, W. M. A., et al. (2013). Hypersociability in the behavioral phenotype of 17q21.31 microdeletion syndrome. *American Journal of Medical Genetics Part A, 161A,* 21-26.

Fedorenko, E., Morgan, A., & Murray, E., et al. (2016). A highly penetrant form of childhood apraxia of speech due to deletion of 16p11.2. *European Journal of Human Genetics, 24*, 302-306.

Fujishita, A., Koga, Y., Utsumi, D., Nakamura, A., Yoshimi, T., & Yoshida, N. (2015). Effects of feeding a soft diet and subsequent rehabilitation on the development of the masticatory function. *Journal of Oral Rehabilitation, 42*, 266-274.

Goldman, R., & Fristoe, M. (2000). *Goldman-Fristoe test of articulation*. Bloomington, MN: Pearson.

Iuzzini, J., & Forrest, K. (2010). Evaluation of a combined treatment approach for childhood apraxia of speech. *Clinical Linguistics & Phonetics, 24*, 355-345.

Keen, C., Samango-Sprouse, C., Dubbs, H., & Zackai, E. H. (2017). 10-year-old female with intragenic KANSL1 mutation, no KANSL1-related intellectual disability, and preserved verbal intelligence. *American Journal of Medical Genetics Part A, 173*, 762-765.

Koolen, D. A., & de Vries, B. B. A. (2010). KANSL1-related intellectual disability syndrome. In: Pagon, R.A., Adam, M.P., Ardinger, H.H., et al., editors. *GeneReviews® [internet].* Seattle: University of Washington.

Koolen, D. A., Kramer, J. M., & Neveling, K., et al. (2012). Mutations in the chromatin modifier gene KANSL1 cause the 17q21.31 microdeletion syndrome. *Nature Genetics, 44,* 639-641.

Koolen, D. A., Pfundt, R., & Beunders, G., et al. (2016). The Koolen-de Vries syndrome: a phenotypic comparison of patients with a 17q21.31 microdeletion versus a KANSL1 sequence variant. *European Journal of Human Genetics, 24*, 652-659.

Koolen, D. A., Sharp, A. J., & Hurst, J. A., et al. (2008). Clinical and molecular deleniation of the 17q21.31 microdeletion syndrome. *Journal of Medical Genetics, 45,* 710-720.

Koolen, D. A., Vissers, L. E., & Pfundt, R., et al. (2006). A new chromosome 17q21.31 microdeletion syndrome associated with a common inversion polymorphism. *Nature Genetics, 38*, 999-1001.

Kort, W., Schittekatte, M., & Compaan, E. (2008). *CELF-4-NL: clinical evaluation of language fundamentals*. Amsterdam: Pearson Assessment and Information B.V.

Law, J., Garrett, Z., & Nye, C. (2003). Speech and language therapy interventions for children with primary speech and language delay or disorder. *Cochrane Database of Systematic Reviews, 3*.

Maassen, B., van Haaften, L., Diepeveen, S., Terband, H., van den Engel-Hoek, L., Veenker, T., & De Swart, B. (2019). *Computer Articulation Instrument*. Amsterdam: Boom Uitgevers.

Mervis, C. B., Morris, C. A., Klein-Tasman, B. P., Velleman, S. L., & Osborne, L. R. (2015). 7q11.23 duplication syndrome. In: Pagon, R.A., Adam, M.P., Ardinger, H.H., et al., editors. *GeneReviews® [internet]*. Seattle: University of Washington.

Morgan, A. T., & Liégeois, F. (2010). Re-thinking diagnostic classification of the dysarthrias: a developmental perspective. *Folia Phoniatrica et Logopaedica, 62*, 120-126.

Morgan, A. T., Liégeois, F., & Liederkerke, C., et al. (2011). Role of cerebellum in fine speech control in childhood: persistent dysarthria after surgical treatment for posterior fossa tumour. *Brain and Language, 117*, 69-76.

Morgan, A. T., Mei, C., & Da Costa, A., et al. (2015). Speech and language in a genotyped cohort of individuals with Kabuki syndrome. *American Journal of Medical Genetics Part A, 167*, 1483-1492.

Murray, E., McCabe, P., & Ballard, K. J. (2015). A randomized controlled trial for children with childhood apraxia of speech comparing rapid syllable transition treatment and the Nuffield Dyspraxia Programme-Third Edition. *Journal of Speech Language and Hearing Research, 58*, 669-686.

Park, J., Lombardino, L. J., & Ritter, M. (2013). Phonology matters: a comprehensive investigation of reading and spelling skills of school-age children with mild to moderate sensorineural hearing loss. *American Annals of the Deaf, 158*, 20-40.

Pennington, L., & McConachie, H. (1999). Mother-child interaction revisited: communication with non-speaking physically disabled children. *International Journal of Language & Communication Disorders, 34,* 391-416.

Pennington, L., Parker, N. K., Kelly, H., & Miller, N. (2016). Speech therapy for children with dysarthria acquired before three years of age. *Cochrane Database of Systematic Reviews, 7.*

Reilly, S., Skuse, D., & Wolke, D. (1999). *Schedule for oral motor assessment*. London, UK: Wiley.

Robbins, J., & Klee, T. (1987). Clinical assessment of oropharyngeal motor development in young children. *Journal of Speech and Hearing Disorders, 52*, 271-277.

Sanchez, K., Spittle, A. J., Slattery, J. M., & Morgan, A. T. (2016). Oromotor feeding in children born before 30 weeks' gestation and term-born peers at 12 months' corrected age. *Journal of Pediatrics, 178*, 113-8.e1.

Schlichting, J. E. P. T., & Lutje Spelberg, H. C. (2010). *Schlichting Test voor Taalproductie-II*. Houten: Bohn, Stafleu van Loghum.

Semel, E., Wiig, E. H., & Secord, W. A. (2003). *Clinical evaluation of language fundamentals: Australian standardised edition*. Sydney: Pearson.

Tan, T. Y., Aftimos, S., & Worgan, L., et al. (2009). Phenotypic expansion and further characterisation of the 17q21.31 microdeletion syndrome. *Journal of Medical Genetics, 46*, 480-489.

van den Engel-Hoek, L., van Hulst, K. C., van Gerven, M. H., van Haaften, L., & de Groot, S. A. F. (2014). Development of oral motor behavior related to the skill assisted spoon feeding. *Infant Behavior and Development, 37*, 187-191.

Volkert, V. M., Piazza, C. C., Vaz, P. C., & Frese, J. (2013). A pilot study to increase chewing in children with feeding disorders. *Behavior Modification, 37*, 391-408.

White, S. M., Morgan, A., & Da Costa, A., et al. (2010). The phenotype of Floating-Harbor syndrome in 10 patients. *American Journal of Medical Genetics Part A, 152A*, 821-829.

Wiig, E. H., Secord, W. A., & Semel, E. (2006). *Clinical evaluation of language fundamentals-preschool: Australian standardised edition*. Sydney: Harcourt Assessment.

Wilkinson, G. S., & Robertson, G. J. (2006). *Wide range achievement test*. Lutz: Psychological Assessment Resources.

Winskell, H. (2006). The effects of an early history of otitis media on children's language and literacy skill development. *British Journal of Educational Psychology, 76*, 727-744.

Zimmerman, I. L., Steiner, V. G., & Pond, R. E. (2011). *Preschool language scales: Australian and New Zealand language adapted edition*. Sydney: Pearson.

Zollino, M., Marangi, G., & Ponzi, E., et al. (2015). Intragenic KANSL1 mutations and chromosome 17q21.31 deletions: broadening the clinical spectrum and genotype–phenotype correlations in a large cohort of patients. *Journal of Medical Genetics, 52*, 804-814.

SEVEN

CHAPTER 7

SUMMARY AND GENERAL DISCUSSION

Summary

A major task for speech-language pathologists (SLPs) is to differentiate children with delayed or disordered speech development from typically developing children. Prevalence rates for speech delay or speech sound disorders (SSDs) are reported ranging from 2.3% to 24.6% (Wren, Miller, Peters, Emond, & Roulstone, 2016). To be able to correctly identify and diagnose children with SSD, it is important to have reliable and valid assessment tools with representative normative data. In addition, these tools are important for clinical decision making and treatment planning. Children with SSD form a heterogeneous group, showing variability in severity, etiology, proximal causes, speech characteristics, and response to treatment (Dodd, 2014; Ttofari Eecen, Eadie, Morgan, & Reilly, 2018). Currently available tests yield detailed descriptions of speech symptoms. They do not directly assess the underlying processes involved: lemma access, word form selection, phonological encoding, speech motor planning and programming, and speech motor execution (Terband, Maassen, & Maas, 2019). Insight into the deficits that might be the underlying causes of an SSD, requires an extensive analysis of a child's performance on a range of speech tasks that reflect different underlying processes. Based on these premises, the Computer Articulation Instrument (CAI) was developed (Maassen et al., 2019). The overall aim of this thesis was to characterize the speech profiles of typically developing children and children with speech delay or SSD, such that a first, significant step towards process-oriented diagnostics was taken. The following aims were addressed:

1. To construct a speech production test battery that comprises a range of speech tasks, for the assessment of a comprehensive speech profile.
2. To collect normative data on the development of speech profiles in typically developing children and analyse the psychometric properties of the CAI in terms of interrater and test-retest reliability and different aspects of construct validity.
3. To provide a comprehensive analysis of speech sound development in Dutch-speaking children based on a large cross-sectional study as a background for interpreting clinical test results.
4. To describe and analyse clinical speech characteristics of speech delay and specific speech deficits in children with diverse neurodevelopmental disorders.

In **Chapter 1** background information of speech development, speech disorders, and measurement of speech is presented. In addition, the aims and outline of this thesis are given. This thesis is divided into two parts. **Part One** comprises **chapter 2**, **3**, and **4** and focusses on the construction of a new speech assessment tool, called Computer Articulation Instrument (CAI), and the description of typical

speech sound development in Dutch-speaking children. **Part Two** comprises **chapter 5** and **6** describing the clinical implication of the CAI.

Chapter 2 focuses on the psychometric evaluation of the CAI. The CAI is a computer-based speech production test battery with a broad set of tasks that allows the detection of signs of delay or deviance in several speech production characteristics such that a norm-referenced speech profile for Dutch-speaking children can be obtained. The CAI assesses both phonological and speech motor skills in 4 tasks: (1) picture naming, (2) nonword imitation, (3) word and nonword repetition, and (4) maximum repetition rate (MRR). The task *picture naming* consists of 60 items and taps into the whole chain of speech processes, from preverbal visual-conceptual processing to lemma access, word form selection, phonological encoding, speech motor planning and programming, and speech motor execution (Maassen & Terband, 2015). With the task *nonword imitation* a child is asked to reproduce nonwords (or nonsense words). In contrast to picture naming, a child cannot revert to its lexicon during this task and thus the child either needs to analyse the phonological structure of the nonword directly, addressing the phonological decoding and encoding system, or follows the auditory-to-motor-planning pathway. The task *nonword imitation* in the CAI consists of 80 items for the 2- to 3-year-old children, and 33 items for the 4-to-7-year-old children. In word and *nonword repetition* a child is asked to repeat 5 words or nonwords 5 times. This task aims to assess variability in speech production, which occurs when a child uses multiple productions of the same word or nonword. *MRR* is a pure motor task (articulo-motor planning and programming) and does not require any knowledge of words, syllables, or phonemes. The *MRR-task* consists of three monosyllabic sequences (/pa/, /ta/, /ka/), two bisyllabic sequences (/pata/, /taka/), and one trisyllabic sequence (/pataka/). Psychometric properties were reported in terms of description of the normative sample, interrater and test-retest reliability, content validity, and two aspects of construct validity. A total of 1,524 typically developing Dutch-speaking children aged between 2;0 and 7;0 participated in the normative study. Fourteen age groups were created with a range of 4 months for children aged 2;0-5;11 and a range of 6 months for those aged 6;0-6;11. The sample was representative of the general Dutch population in terms of gender, geographic region, degree of urbanization and socioeconomic status. The four tasks of the CAI were administered to all children. A set of parameters was extracted for the four tasks. Segmental and syllabic accuracy were measured with *picture naming* and *nonword imitation.* These parameters aimed to measure phoneme repertoire and the ability of producing combinations of speech sounds (e.g., percentage of consonants correct, percentage of reduction of consonant clusters, percentage of syllable structure correct). Consistency was measured with the task *word and nonword repetition.* The parameter proportion of whole-word variability (PWV) was calculated (Ingram, 2002) by dividing the total number of different forms by the total number of productions. *MRR* was calculated as the number of syllables per second, and aims to measure speech motor planning and programming.

Interrater reliability and test-retest reliability were analysed using subgroups of the normative sample and studied by estimating intraclass correlation coefficients (ICCs). ICCs for interrater reliability ranged from sufficient to good, except for percentage of vowels correct of *picture naming* and *nonword imitation* and for the *MRRs* for bisyllabic and trisyllabic items (MRR-BiTri). The ICCs for test-retest reliability were sufficient for the parameters of *picture naming* and *nonword imitation.* Insufficient ICCs were found for the parameters of *word and nonword repetition* and *MRR*. The scores on these tasks improved significantly the second time round due to natural development and learning effects.

Content validity of the CAI is demonstrated by the description of the test domain, its four speech tasks, and their items. The rationale of these four tasks was described. The items of the four tasks each measured different aspects of speech production.

The first aspect of construct validity concerns the requirement that the outcome of a developmental test shows a correlation with age. We hypothesized that the selected parameters would reflect typical speech development and would thus show a monotonous improvement with age. This aspect was investigated by comparing the raw scores of the four tasks in the normative sample (n = 1,524) of typically developing Dutch-speaking children aged between 2;0 and 7;0 years. Continuous norms showed developmental patterns for all CAI parameters. The second aspect of construct validity, structural validity, was determined by testing all the parameters of the instrument. Structural validity means that the parameters of a task behave as expected according to an underlying model. In order to examine this aspect of construct validity, factor analyses were conducted based on the assumption that clusters of the selected parameters would reflect different aspects of speech production, either within or across tasks. Factor analyses on a total number of 20 parameters revealed five meaningful factors: all picture-naming parameters (PN), the segmental parameters of nonword imitation (NWI-Seg), the syllabic structure parameters of nonword imitation (NWI-Syll), (non)word repetition consistency (PWV), and all MRR parameters (MRR-mono and MRR-BiTri). These results confirmed the distinctiveness of the four tasks of the CAI. Each task reflects different aspects of speech production. Furthermore, the construct validity was underlined by the weak correlations between CAI factor scores, indicating the independent contribution of each factor to the speech profile.

Based on its overall sufficient to good psychometric properties, we concluded that the CAI is a reliable and valid instrument for the assessment of speech development in Dutch children aged between 2;0 and 7;0 years and that it can be used to gauge typical and atypical speech development.

Chapter 3 and **4** of this thesis are devoted to the description of a comprehensive analysis of speech sound development in Dutch-speaking children based on a large cross-sectional study. **Chapter 3** gives a detailed description of the speech sound development of 1,503 typically developing Dutch-speaking children from the normative study of chapter 2, aged between 2;0 and 7;0 years. In recent years, many studies have been conducted worldwide to investigate speech sound development in different languages, including several that explored the typical speech sound development of Dutch-speaking children, but none of these latter studies explored both phonetic and phonological progress within a sufficiently wide age range and a sufficiently large sample. This study serves to fill this gap by providing normative cross-sectional results on informative parameters of speech development: percentage of consonants correct (PCC-R) and percentage of vowels correct (PVC), consonant, vowel and syllabic structure inventories, degrees of complexity (phonemic feature hierarchy), and phonological simplification processes. The picture-naming task of the CAI was used to obtain these parameters. PCC-R and PVC significantly increased with age. The consonant inventory was found to be complete at 3;7 years of age for the syllable-initial consonants, with the exception of the voiced fricatives /v/ and /z/, and the liquid /r/. All syllable-final consonants were acquired before the age of 4;4 years. At the age of 3;4 years, all children had acquired a complete vowel inventory and at the age of 4;7 years they produced most syllable structures correctly, albeit that the syllable structure CCVCC was still developing. All phonological contrasts were produced correctly at 3;8 years of age. Children in the younger age groups used more phonological simplification processes than the older children and by the age of 4;4 years, all simplification processes had disappeared, except for the initial cluster reduction from three to two consonants and the final cluster reduction from two to one consonant. This detailed description of typical Dutch speech sound development provides SLPs with pertinent information to determine whether a child's speech development progresses typically or is delayed or disordered.

Normative data for the MRR development of Dutch-speaking children based on a large cross-sectional study is provided in **Chapter 4**. In this study a group of 1,014 typically developing children aged 3;0 to 6;11 years performed the MRR task of the CAI. A standardised protocol was used that was proposed by our research group in a previous study (Diepeveen, Van Haaften, Terband, De Swart, & Maassen, 2019). The number of syllables per second was calculated for mono-, bi-, and trisyllabic sequences (MRR-pa, MRR-ta, MRR-ka, MRR-pata, MRR-taka, MRR-pataka). For all MRR sequences the MRR increased significantly with age. MRR-pa was the fastest sequence, followed by respectively MRR-ta, MRR-pata, MRR-taka, MRR-ka and MRR-pataka. Overall MRR scores were higher for boys than for girls, for all MRR sequences. With the normative data presented in this study, clinicians are able to distinguish typically developing children from children with SSD. In addition, MRR can contribute to a first step in differential diagnosis between different types of SSD, and especially between different types of motor speech disorders (MSD). The in-depth description of typical speech sound development and MRR in Chapter 4 and 5 can serve as a background for interpreting clinical test results.

The second part of this thesis concerns the aim to describe and analyse clinical speech characteristics of speech delay and specific speech deficits in children with diverse (neuro)developmental disorders. The studies described in **Chapter 5** and **6** are two examples of the clinical implication of the CAI. In **Chapter 5** data from two studies on clinical groups of children with speech language impairments and SSDs were presented in order to determine the known-group validity of the CAI. *Study 1* examined known-group validity by comparing the scores of 93 children diagnosed with speech-language difficulties on one task of the CAI (picture naming) with intelligibility judgments given by SLPs. For this study, the parameter "percentage of consonants correct" of the picture naming task was used (PN-PCC). Performance on picture naming was significantly affected by intelligibility level, and there was a highly significant correlation between the intelligibility levels and PN-PCC in the expected direction. Neither intelligibility level nor picture naming performance was related to nonverbal intelligence and language scores.

Study 2 aimed to determine the diagnostic power of all four CAI tasks by comparing scores on the CAI factors resulting from the study in Chapter 2: PN, NWI-seg, NWI-syll, PWV, MRR-Mono and MRR-BiTri with clinical judgments of severity of speech difficulties given by SLPs of 41 children diagnosed with SSD. Most of the children ($n = 36$) were diagnosed with a phonological disorder (PD), two children with childhood apraxia of speech (CAS), and three children had an unknown diagnosis because no details were available about the children's speech apart from the fact that their SSD was severe. A severity scale with three categories was used: mild, moderate and severe. Significant differences were found between the moderate and severe groups for the CAI factors based on picture naming, nonword imitation, and the bisyllabic and trisyllabic sequences of MRR. No significant differences were found between the moderate and severe groups for the factor word and nonword proportion of whole-word variability, and the monosyllabic sequences of the MRR. These results suggest that, especially the tasks PN, NWI, and the bisyllabic and trisyllabic sequences of MRR are most sensitive for diagnosing SSDs. Known-group validity of the CAI is supported by the findings of these two studies. Together with the results of Chapter 2, we can conclude that the CAI is a reliable and valid tool for assessment of children with SSDs.

Chapter 6 describes the communication phenotype associated with Koolen de Vries syndrome. In this prospective study the CAI was administered to the Dutch children who participated in this study. In total, 29 participants with Koolen de Vries syndrome (four with KANSL1 variants, 25 with 17q21.31 microdeletion) were assessed for oral motor, speech, language, literacy, and social functioning. The linguistic phenotyping revealed a distinctive communication profile. The participants showed an overriding 'double hit' of oral hypotonia and apraxia in infancy and preschool, associated with severely delayed speech development. Remarkably, however, speech prognosis was positive; apraxia resolved, and although dysarthria persisted, children were intelligible by mid-to-late childhood. In contrast, language and literacy deficits persisted, and pragmatic deficits were apparent.

In the next part I will elaborate on these findings by discussing two themes that emerged from this thesis (1) the importance of process-oriented diagnosis in speech sound disorders, and (2) future perspectives: the next step in process-oriented diagnosis.

General discussion

In the general introduction of this thesis the case study of Tom was presented, a 3-year-old boy with speech difficulties. His parents were concerned about his speech sound development. He is difficult to understand for his parents without a context, and his speech is unintelligible for unfamiliar listeners. Tom's speech is characterised by omissions and substitutions of sounds. His parents wonder whether speech language therapy is indicated and if so, which therapy approach is appropriate. In this case, the role of the SLP is to (1) consider Tom's situation, (2) collect cues and information (history taking, observation, standardised tests), and (3) process this information, thereby taking the first steps in the cycle of clinical reasoning (Higgs, Jones, Loftus, & Christensen, 2008). By following these steps, an SLP is able to identify and classify an SSD, and to make the next step in clinical reasoning: decision making for treatment planning. A reliable and valid speech assessment based on normative data is essential to be able to collect and process information, and to establish a differential diagnosis of SSD. With the availability of the CAI and Dutch normative data of speech development, described in chapter 2, 3 and 4 of this thesis, the important first elements in the cycle of clinical reasoning in the therapeutic process of children with speech difficulties are covered.

The importance of process-oriented diagnosis in speech sound disorders

Current diagnostic process in Dutch SLT practice

A study by our research group (Diepeveen, Van Haaften, Terband, De Swart, & Maassen, in press) gave insight in the clinical reasoning (diagnosis and intervention) of SLPs working with children with SSD in the Netherlands. Semi-structured interviews containing nondirective, open-ended questions were conducted with 33 SLPs, and 137 other SLPs filled out a questionnaire on the same topics. The results indicated that Dutch SLPs use a variety of assessments to diagnose SSD, complemented by observation and often case history. The vast majority of SLPs in the Netherlands uses 'LOGO-Art Dutch Articulation Assessment' (Nederlands Articulatieonderzoek (NAO; Baarda, de Boer-Jongsma, & Jongsma, 2013), and the Dutch version of the Metaphon Screening Assessment (Leijdekker-Brinkman, 2002). Other reported speech assessments were the Dutch version of the Hodson Assessment of Phonological Patterns (Van de Wijer-Muris & Draaisma, 2000), spontaneous speech samples, 'Dyspraxia Program' similar to the Nuffield Dyspraxia Program (Eurlings- van Deurse, Freriks, Goudt-Bakker, Van der Meulen, & Vries, 1993), oral motor assessments, a qualitative observation based on the Motor Speech Hierarchy framework used for PROMPT therapy: Verbal Motor Production Assessment for Children (VMPAC; Hayden, 2004), the Articulation subtest of the TAK (subtest Klankarticulatie, Taaltoets Alle Kinderen), a Dutch Language Proficiency Test for All Children (Verhoeven & Vermeer, 2001), and own (custom-made) speech assessments. None of these assessments are norm-based for the Dutch population or provide information about reliability and validity except for the TAK (Verhoeven & Vermeer, 2001). Moreover, all tests measure only one aspect of speech production. The production of speech sounds is a complex process that comprises lemma access, word form selection, phonological encoding, speech motor planning and programming, and speech motor execution skills. Children with SSD can experience problems with one of more of these speech processes. Speech assessment should evaluate these different aspects of SSD to be able to obtain a complete speech profile. With the available assessments, SLPs can only rely on their own interpretation to establish a differential diagnosis of the speech disorder. Despite the availability of various speech assessments, the study of Diepeveen et al. (in press) revealed the need for the availability of a fast and easy-to-administer comprehensive differential diagnostic instrument with the availability of normative data. A single and comprehensive assessment that can differentiate between the various diagnostic labels was missing (Terband et al., 2019).

Diagnostic labels

Differential diagnosis refers to the process of determining the appropriate diagnostic labels for the SSD, such as phonological disorder, childhood apraxia of speech, or dysarthria. The study of Diepeveen et al. (in press) showed that Dutch SLPs use 85 different diagnostic labels to categorize SSD. These results reveal the need for a clear and consistent terminology of diagnoses in the field of paediatric SSD. To date, no validated list of speech symptoms for differential diagnosis of SSDs is available. This may be due to the fact that different types of SSD that appear similar at the behavioural level may have different causal origins. Bishop and Snowling (2004) described four levels in their causal model for developmental disorders: etiology, neurobiology, cognitive processes and observed behaviour (i.e., observed symptom patterns). Etiological factors (genetic and environmental) determine neurobiological factors, and these in turn influence both underlying cognitive processes and the behavioural level. There are also effects in the opposite direction: children's behaviour and cognition can affect the neurobiological factors.

A number of classification systems of SSD have been described in the literature featuring a variety of approaches, i.e., etiology, descriptive-linguistics, psycholinguistic and psychomotor processing (Waring & Knight, 2013). No system is currently universally accepted by SLPs. Two of the most described systems in literature are the Speech Disorders Classification System (SDCS) by Shriberg et al. (2010) and Dodd's Model for Differential Diagnosis (Dodd, 2014).

The SDCS contains two branches: Speech Disorders Classification System-Typology (SDCS-T) and Speech Disorders Classification System-Etiology (SDCS-E). The classification in SDCS-T is based on a speaker's age and current and/or prior speech characteristics. SDCS-E divides SSD into three classes, based on etiology: Speech Delay (SD), Speech Errors (SE) and Motor Speech Disorder (MSD; including dysarthria, childhood apraxia of speech (CAS) and motor speech disorder – not otherwise specified) (Shriberg et al., 2010). According to the causal model of Bishop and Snowling (2004), in the SDCS children's speech disorders are described at the level of etiology and the level of behaviour (description of speech characteristics). However, the problem with this diagnostic system is that the cross-referencing of etiological and behavioural characteristics is insufficient. This makes it difficult to interpret and to use this model in differential diagnosis. In the future, such a multi-level classification system would become more clinically useful when genetic factors associated with speech disorders will be identified (Waring & Knight, 2013).

Dodd's Model for Differential Diagnosis is a descriptive-linguistic classification system and describes the proximal causes of SSD, thereby using a more consistent method than the SDCS. The model contains five key subgroups: (1) inconsistent phonological disorder, (2) consistent atypical phonological disorder, (3) phonological delay, (4) articulation disorder, and (5) childhood apraxia of speech (CAS) (Dodd, 2014). The classification into these five subgroups is based on error patterns which are typically linked to descriptive-linguistic proximal causes. For example, children with a phonological delay are characterized by all errors being accounted for by phonological error patterns (e.g., cluster reduction and fronting) that occur during typical speech development at a younger chronological age level. Several studies report the validation of this model (Dodd, 2011; Ttofari Eecen et al., 2018; Waring & Knight, 2013), and therefore it could be a good choice to use this system in clinical practice for differential diagnosis and targeted intervention of SSD in children. Thus, in contrast to the SDCS, the model of Dodd (2014) describes only one level of causation according to the model of Bishop and Snowling (2004): symptom patterns are described at the behavioural level with a clear reference to the descriptive-linguistic proximal causes. This makes Dodd's model more clinically feasible than the SDCS. However, when the context of the elicited speech is not specified, it is inevitable that there is a large overlap of symptoms at the behavioural level between diagnostic categories and a large heterogeneity within categories (Terband et al., 2019). As a consequence, until now no validated list of symptom patterns for differential diagnosis of SSDs can be determined. Further investigations are needed into the connection between speech profiles and underlying speech processes of the proposed subgroups in the SDCS and Dodd's Model for Differential Diagnosis.

In order to reveal the proximal causes of SSD a re-orientation from behavioural diagnostics to process-oriented diagnostics is required (Terband et al., 2019). More important than determining the diagnostic label is the determination of the possibly deficient underlying speech processes, such as lemma access, word form selection, phonological encoding, speech motor planning and programming, and speech motor execution. In the past years theoretical frameworks have been developed, focusing on clearly defined underlying processes instead of only behavioural symptom patterns (Maassen & Terband, 2015; Terband et al., 2019). These frameworks give a conceptual basis to analyse speech disorders and determine the underlying deficit of SSD, and thereby form the basis for process-oriented diagnostics. A first approach to be able to clinically assess the contribution of separate processing stages to the symptom profile, is the use of a set of speech tasks that each require different steps in the speech production process (i.e., lemma access, word form selection, phonological encoding, speech motor planning and programming, and speech motor execution). It is important to know in which context, or with which speech tasks, speech production is elicited. With a set of tasks as in the CAI, different aspects of speech production can be evaluated and compared with each other in order to obtain a complete speech profile. For example, the performance on the tasks picture naming and nonword imitation can be compared. In contrast to picture naming, a child cannot revert to its lexicon during nonword imitation and thus either analyses the phonological structure of the nonword to address the phonological encoding system, or follows the auditory-to-motor planning pathway. Since nonword imitation does not appeal to the underlying speech processes of lemma access and word form selection from the lexicon, comparing performance on picture naming versus nonword imitation gives information on the role of the lexicon. Relying on this concept, the CAI was constructed and normative data on the development of speech profiles in typically developing children were collected (chapter 2). The CAI contains 4 tasks: (1) picture naming, (2) nonword imitation, (3) word and nonword repetition, and (4) maximum repetition rate (MRR). The test battery of the CAI incorporates these different tasks to be able to measure the whole speech production chain of lemma access, word form selection, phonological encoding, speech motor planning and programming, and speech motor execution of children in the ages between 2 and 7 years. With the collected normative data it is possible to compare the performance on the different tasks. The psychometric properties of the CAI were analysed in terms of interrater and test-retest reliability and different aspects of construct validity referring to development and production processes. In order to be able to interpret clinical test results, a comprehensive analysis of speech sound development in Dutch-speaking children based on a large cross-sectional study is provided (chapter 3 and 4).

Treatment planning

The CAI contributes to the ability of collecting and processing information, one of the first elements in the cycle of clinical reasoning. This information can be used to make decisions for treatment planning. The assessment procedure of the CAI provides SLPs the opportunity to describe different aspects of speech production and a complete speech profile can be obtained. A complete speech profile, and not only describing single behavioural symptoms, offers the SLP the possibility to describe the underlying processing deficits of the child's speech difficulties. In the example of Tom, presented in the case study in the general introduction of this thesis, the behavioural speech symptom is described as substitution of /k/ to [t] ([tɑp-tɔk] for /kap-stok/). However, without further information with respect to context and other error patterns, it remains unclear what causes this substitution. Both phonological encoding and speech motor planning and/or programming may be the underlying deficit. Determining the degree to which Tom has difficulties with phonological encoding, compared with having problems with speech motor planning and programming, directly affects appropriate treatment planning. This can be assessed by comparing the performance on a task that taps into the whole chain of speech processes, from lemma access to speech motor execution (e.g., a picture naming task), with the performance of a task that does not require any knowledge of words, syllables or phonemes and only assesses speech motor skills, like an MRR task. In process-oriented treatment planning, the possible treatment goals must correspond to underlying processing deficits. Children with specific speech profiles benefit from different intervention approaches. A nice example is offered in the study of Crosbie, Holm, and Dodd (2005). They compared the effects of two different types of intervention on the consistency of word production and speech accuracy of children with consistent or inconsistent speech disorder. These two speech disorders are caused by different underlying speech processing deficits. Children with a consistent speech disorder have poor understanding of the phonemic rules of a language (Crosbie et al., 2005), whereas children with an inconsistent speech disorder have difficulties with phonological planning (i.e., phoneme selection and sequencing) (Dodd & McCormack, 1995). In the study of Crosbie et al. (2005) a phonological contrast therapy aiming at reorganization of the child's phonological system, was compared with a core vocabulary therapy. In this study a minimal pair approach was used in which the child's error is contrasted with the target sound using minimal pairs of words (e.g. fit – bit, tea – key). Core vocabulary therapy targets whole word production, and by providing detailed specific information about a limited number of words strives to improve the processes of phoneme selection and sequencing. A child thus learns how to say a limited set of highly frequent, functional words consistently, and as a consequence, the ability to create phonological plans directly improves (Crosbie et al., 2005). Results indicated that a greater change was seen in children with a consistent speech disorder during phonological contrast therapy and in children with inconsistent speech disorder a greater change was seen during core vocabulary therapy. The results of this study clarify that it is essential to identify different underlying speech processing deficits to be able to choose the most appropriate intervention approach.

The study in chapter 6 of this thesis is an example of how a description of a set of deficient processes leads to specific suggestions for intervention. In children with Koolen de Vries syndrome speech development was the core challenge in the preschool period. The described phenotype includes childhood apraxia of speech, flaccid dysarthria, and additional articulation and phonological errors. For children younger than 3 years of age with a small vocabulary, a core vocabulary treatment programme (Crosbie et al., 2005) was advised. For children with childhood apraxia of speech, in the age from 4 years, two treatment programmes were advised: the Nuffield Dyspraxia Programme Version 3 and the Rapid Syllable Repetition Programme (Morgan, Murray, & Liégeois, 2018). Targeted dysarthria treatment, like developed by Pennington, Parker, Kelly, and Miller (2016), is proposed when apraxia resolves and the dysarthric features become more apparent.

Evaluation of the effectiveness of treatment

The next element in the cycle of clinical reasoning is the evaluation of the effectiveness of treatment. A range of research studies are available documenting evidence for intervention approaches for children with SSD. Baker and McLeod (2011) presented a narrative review of 134 published studies of phonological intervention relevant to children with a phonological impairment, delay, or disorder. The majority of studies (74.1%) were associated with lower levels of evidence (quasi-experimental studies, 41.5%; non experimental case studies, 32.6%). A review by Morgan, Murray and Liégeois (2018) on the efficacy of interventions in children and adolescents with CAS revealed only one well-controlled study. This study conducted by Murray, McCabe, and Ballard (2015) compared two intervention programmes: the Rapid Syllable Transitions Treatment (ReST); and the Nuffield Dyspraxia Programma-3 (NDP-3). Both programmes are based on principles of motor learning. Limited evidence is found for that both ReST and NDP-3 improve word accuracy, measured by the accuracy of production on treated and non-treated items, speech production consistency and the accuracy of connected speech. Only one small randomised control trial could be identified in this review. Therefore, additional evidence associated with higher levels of evidence (further randomised control trials) is needed for adequate clinical decision making in children with SSD, among which CAS. Furthermore, until now the effect of treatment is measured with general parameters (e.g., PCC, words produced correctly) and, as a consequence, only general effects could be indicated. In future research the results of the CAI administration can be used to evaluate the effectiveness of treatment. With its possibility to compare performance on different tasks (i.e., process-oriented assessment and diagnosis), this type of evaluation method makes it possible to measure more specific effects, thereby making the measures more sensitive to changes. In addition, the CAI offers the ability to reflect on the former decisions in the cycle of clinical reasoning (i.e., collecting and processing information), and to change an intervention approach if other deviant underlying processes become more evident during treatment. Due to the effect of treatment on one underlying process, the core problem might be shifted to another process, and with that a change in intervention approach is indicated. For example, the supposed core deficit in children with CAS comprises a reduced capacity of planning and programming the motor commands that activate speech musculature (Nijland, Maassen, & van der Meulen, 2003; Nijland, Maassen, van der Meulen, et al., 2003; Shriberg, Lohmeier, Strand, & Jakielski, 2012). Therefore, intervention programmes based on principles of motor learning are the first and most appropriate choice in children with CAS. However, in typical development infants speech develops from random babbling and sensorimotor learning to a more abstract phonological acquisition. In children with CAS, the word form lexicon and phonological encoding system (higher-level processes) are acquired with deviant motor learning capacities (lower-level processes). Bottom-up processes during speech acquisition explain a large part of the symptomatology of CAS in the phonology domain (Maassen, 2002). As a result, in children with CAS deficits in phonological skills could remain after motor learning treatment. Identification of the remaining deviant process can be identified with the CAI and based on these results treatment planning can be evaluated and adjusted.

Future perspectives: the next step in process-oriented diagnosis in speech sound disorders

Chapter 2, 3, and 4 show that the CAI can be used to evaluate and compare different aspects of speech production. With the characterization of the comprehensive speech profiles of typically developing children and children with speech delay or SSD, described in chapter 5 and 6, a first significant step towards process-oriented diagnostics is taken. The study in chapter 5 yields strong indications that comparison of different aspects of speech production provides information on the underlying speech processing difficulties of children with SSD. By comparing performance on the different tasks of the CAI, it is possible to exclude speech processes from influencing the speech deficits of a child. For example, when a child with SSD has no difficulties with word and nonword repetition and no difficulties with MRR, it is assumed that speech motor planning and programming is not involved. From this, it is only one step further to the description of the specific processing deficits that underlie the different speech profiles based on different aspects of speech production.

Manipulations of tasks

The assessment of separate speech processes can be achieved by the implementation of experimental methodology. Future research will focus on the expansion of the CAI with the possibility of manipulations of tasks, thereby making it possible to analyse different parameters in different contexts. For example, auditory feedback manipulations can be added to the assessment battery. It is hypothesized that children with CAS continuously monitor their speech through auditory feedback to minimize speech errors (Terband & Maassen, 2010; Terband, Maassen, Guenther, & Brumberg, 2009). A couple of experimental studies investigated the role of auditory feedback in children with CAS. A study by Terband, van Brenk, and van Doornik-van der Zee (2014) investigated the ability of children with SSD to compensate and adapt for perturbed auditory feedback compared to typically developing children. Effects of real-time formant-frequency perturbation was investigated. They found that typically developing children tended to compensate for the perturbation whereas children with SSD tended to follow and exaggerate the frequency shift. It was suggested that children with SSD perceived the formant shift, but did not compensate appropriately. Iuzzini-Seigel, Hogan, Guarino, and Green (2015) investigated effects of auditory feedback masking on vowel space, vowel durations and voice onset time for voiceless plosives in the speech of children with CAS, speech delay and typically speaking children. The results showed that the speech production of children with CAS was affected by masked auditory feedback, whereas the children with speech delay and typically developing children did not show an effect of masking. These studies suggest an increased reliance on auditory feedback in children with CAS, and therefore it seems useful to add auditory feedback manipulations to the battery of speech tasks in the CAI.

An example of an assessment battery with speech production and speech perception tasks is described in the study of Geronikou and Rees (2016). They used this battery to specify the underlying speech processing difficulties in children with similar speech production difficulties. Speech production tasks included picture naming, word repetition and nonword repetition, and the speech perception tasks included nonword auditory discrimination and mispronunciation detection. Four children aged 4;7-5;6 years identified with speech difficulties and five typically developing children aged 4;4-5;11 years participated in this study. An individual profile was described for all four children with speech difficulties, showing that their speech production profiles were very similar, but that their performance on one of the speech perception tasks differed. None of the children had difficulties with the non-word discrimination tasks, but differences were found on the mispronunciation detection task in that two children had difficulties in discriminating the contrasts tested in the mispronunciation detection task, while the other two children performed normally. The authors' interpretation was that for the two children with difficulties the problem resided at the level of phonological representations, and in the two children who had no difficulty with the mispronunciation detection task despite their speech production difficulties, the processing deficit seemed to reside at the level of specifying target sounds into motor programs. These findings confirm the possibility that similar speech error patterns in children may arise from different patterns of underlying speech difficulties. As discussed in the above paragraph, this kind of profiling leads to implications for intervention, and therefore it is important to add these kind of speech perception tasks to the assessment battery of the CAI.

Improving CAI software for next steps in process-oriented diagnosis

In ongoing and planned research to improve the CAI software, we will focus on the development of rules to support the analysis of sound-by-sound speech error patterns. Confusion matrices will be added to allow SLPs to select specific treatment targets. Furthermore, the current analyses of the CAI are at syllable level and in the next version of the CAI word length and word structure will be added. With a more detailed description of the speech profiles of children with different types of SSD, a step can be made from the behavioural level to the level of underlying cognitive deficits as described in the causal model of Bishop and Snowling (2004). The description of the underlying cognitive deficits (i.e. endophenotypes) is a step closer to the genetic factors that cause SSD. The underlying cognitive processes are influenced by the etiological level, including the environmental factors and genetic factors via the level of neurobiology. Identifying the genetic bases of SSD may provide improved diagnosis and new insights in early identification of children at-risk for speech disorders, which could encourage earlier interventions targeted at specific deficits. A shift from reliance on behaviourally-based classifications to gene-brain-underlying processes-behaviour relationships will impact on clinical practice by informing disorder etiology, and improving counselling and prognosis (Morgan, 2013). Until now, only a few genes are described to be associated with speech sound disorders. The only known monogenetically inherited "speech and language gene" is *FOXP2*. CAS is the dominant phenotype associated with *FOXP2*, but other speech impairments are also described, like dysarthria (Shriberg et al., 2006). Recently a number of other genes for CAS have been revealed. This is described in a study by Hildebrand et al. (2020) and they identified nine newly implicating genes (*CDK13, EBF3, GNAO1, GNB1, DDX3X, MEIS2, POGZ, UPF2, ZNF142*). Furthermore, one specific candidate gene (*FOXP1*) on chromosomal region 3p14 has been found to be associated with SSD more broadly (Morgan, 2013). However, *FOXP1* is associated with many more general neurodevelopmental disabilities, and thus is not specifically associated with a speech or language disorder (Morgan, 2013). Because of the heterogeneous phenotype of SSD and the influences of development, the search for genetic influences is challenging (Lewis et al., 2006). Therefore, it is important to identify the specific speech processing deficits that underlie different speech deficits and future additional analyses of the CAI will give more details for endophenotyping.

Several studies investigated the link between particular neurobiological findings and cognitive and behavioural speech performance. Morgan, Su, et al. (2018) found an atypical development of the left corticobulbar tract. Reduced dimensions of the corpus callosum were observed in children with SSD (Luders et al., 2017). Disruption of the dorsal language stream, associated with sound to motor speech transformations, was found in children with CAS by Liégeois et al. (2019). These promising results underline the important (causal) relation between the behavioural and cognitive level with the neurobiological level. In the future the CAI, with the possibilities to study underlying speech processes, might play an important role in these kind of studies.

Another next step would be to describe the speech profiles with the CAI for different subgroups of SSD. The classification into the five subgroups of SSD in Dodd's model for Differential Diagnosis is based on error patterns which are linked to descriptive-linguistic proximal causes (Dodd, 2014). However, until now, for this model only studies have been conducted with only one type of SSD, and no study investigated all the processes of the speech production chain in the same group of children (Waring & Knight, 2013). Future research will focus on the comparison of subgroups of children with different types of SSD using the CAI. In a study by Diepeveen et al. (submitted), a factor analyses was conducted in a heterogeneous group of children with SSD in order to link CAI speech profiles to subgroups of SSD. Speech profiles of children with for example a phonological disorder, childhood apraxia of speech or articulation disorder will be determined to further reveal the proximal causes of SSD.

Final remarks

The step from a behavioural to a process-oriented diagnosis and treatment planning requires a new mindset. During their education, SLPs are trained to think and work according to a model of diagnostic classification. In addition, most of the available speech assessments are oriented on behavioural speech symptoms only. At present, most SLPs tend to choose intervention based on the availability of materials or their own experience (Diepeveen et al., in press). The introduction of a process-oriented approach starts with a change in the education and training of professionals, providing them with the theoretical background and the clinical skills to utilize an instrument based on speech processing profiles and to interpret the results. It is our hope that with the availability of the CAI and the validation studies of this thesis, SLPs will be challenged to reframe the understanding of children's speech sound disorders.

References

Baarda, D., de Boer-Jongsma, N., & Jongsma, B. (2013). *LOGO-Art. Nederlands Articulatieonderzoek [LOGO-Art Dutch Articulation Assessment]*. Dronten, the Netherlands: Uitgeverij LOGO-Start.

Baker, E., & McLeod, S. (2011). Evidence-based practice for children with speech sound disorders: Part 1 narrative review. *Language, Speech, and Hearing Services in Schools, 42*, 102-139.

Bishop, D. V. M., & Snowling, M. J. (2004). Developmental dyslexia and specific language impairment: Same or different? *Psychological bulletin, 130*(6), 858-886.

Crosbie, S., Holm, A., & Dodd, B. (2005). Intervention for children with severe speech disorder: A comparison of two approaches. *International Journal of Language & Communication Disorders, 40*(4), 467-491.

Diepeveen, S., Van Haaften, L., Terband, H., De Swart, B., & Maassen, B. (2019). A standardized protocol for maximum repetition rate assessment in children. *Folia Phoniatrica Et Logopaedica, 71*, 238-250.

Diepeveen, S., Van Haaften, L., Terband, H., De Swart, B., & Maassen, B. (in press). Clinical reasoning for SSDs: Diagnosis and intervention in SLPs' daily practice. *American Journal of Speech-Language Pathology, 71*, 238-250.

Diepeveen, S., Van Haaften, L., Van de Zande, A., De Swart, B., Terband, H., & Maassen, B. (submitted). Towards a typology of speech sound disorders.

Dodd, B. (2011). Differentiating speech delay from disorder: does it matter? *Topics in Language Disorders, 31*(2), 96-111.

Dodd, B. (2014). Differential diagnosis of pediatric speech sound disorder. *Current Developmental Disorders Reports, 1*(3), 189-196.

Dodd, B., & McCormack, P. (1995). A model of speech processing for differential diagnosis of phonological disorders. In B. Dodd (Ed.), *Differential Diagnosis and Treatment of Children with Speech Disorder* (pp. 65-89). London: Whurr Publishers.

Eurlings- van Deurse, M., Freriks, A., Goudt-Bakker, K., Van der Meulen, S., & Vries, L. D. (1993). *Dyspraxieprogramma, therapieprogramma voor kinderen met kenmerken van een verbale ontwikkelingsdyspraxie [Dypraxia Program: A therapy program for children with childhood apraxia of speech]*. Lisse, the Netherlands: Swets & Zeitlinger.

Geronikou, E., & Rees, R. (2016). Psycholinguistic profiling reveals underlying impairments for Greek children with speech disorders. *Child Language Teaching and Therapy, 32*(1), 95-110.

Hayden, D. (2004). PROMPT: A tactually grounded treatment approach to speech production disorders. In I. Stockman (Ed.), *Movement and action in learning* *and development: Clinical implications for pervasive developmental disorders* (pp. 255-297). San Diego: Elsevier-Academic Press.

Higgs, J., Jones, M. A., Loftus, S., & Christensen, N. (2008). *Clinical reasoning in the health professions*: Elsevier Health Sciences.

Hildebrand, M. S., Jackson, V. E., Scerri, T. S., et al. (2020). Severe childhood speech disorder; Gene discovery highlights transcriptional dysregulation. *Neurology, 94*, 1-e20.

Ingram, D. (2002). The measurement of whole-word productions. *Journal of Child Language, 29*(4), 713-733.

Iuzzini-Seigel, J., Hogan, T. P., Guarino, A. J., & Green, J. R. (2015). Reliance on auditory feedback in children with childhood apraxia of speech. *Journal of Communication Disorders, 54*, 32-42.

Leijdekker-Brinkman, W. (2002). *Metaphonbox [Dutch version of the Metaphon Screening Assessment]*. Lisse, the Netherlands: Swets & Zeitlinger B.V.

Lewis, B. A., Shriberg, L. D., Freebairn, L. A., Hansen, A. J., Stein, C. M., Taylor, H. G., & Iyengar, S. K. (2006). The genetic bases of speech sound sisorders: Evidence from spoken and written language. *Journal of Speech, Language, and Hearing Research, 49*(6), 1294-1312.

Liégeois, F. J., Turner, S. J., Mayes, A., et al. (2019). Dorsal language stream anomalies in an inherited speech disorder. *Brain, 142*(4), 966-977.

Luders, E., Kurth, F., Pigdon, L., Conti-Ramsden, G., Reilly, S., & Morgan, A. T. (2017). Atypical callosal morphology in children with speech sound disorder. *Neuroscience, 367*, 211-218.

Maassen, B. (2002). Issues contrasting adult acquired versus developmental apraxia of speech. *Seminars in speech and language, 23*(4), 257-266.

Maassen, B., & Terband, H. (2015). Process-oriented diagnosis of childhood and adult apraxia of speech (CAS and AOS). In M. A. Redford (Ed.), *The Handbook of Speech Production* (Vol. First Edition, pp. 331-350): Wiley.

Maassen, B., van Haaften, L., Diepeveen, S., Terband, H., van den Engel-Hoek, L., Veenker, T., & De Swart, B. (2019). *Computer Articulation Instrument*. Amsterdam: Boom Uitgevers.

Morgan, A. T. (2013). Speech-language pathology insights into genetics and neuroscience: Beyond surface behaviour. *International Journal of Speech-Language Pathology, 15*(3), 245-254.

Morgan, A. T., Murray, E., & Liégeois, F. J. (2018). Interventions for childhood apraxia of speech. *Cochrane Database of Systematic Reviews, 5*, 1-60.

Morgan, A. T., Su, M., Reilly, S., Conti-Ramsden, G., Connelly, A., & Liégeois, F. J. (2018). A brain marker for developmental speech disorders. *The Journal of Pediatrics, 198*, 234-239.

Murray, E., McCabe, P., & Ballard, K. J. (2015). A randomized controlled trial for children with childhood apraxia of speech comparing Rapid Syllable Transition Treatment and the Nuffield Dyspraxia Programme - Third edition. *Journal of Speech, Language, and Hearing Research, 58*(3), 669-686.

Nijland, L., Maassen, B., & van der Meulen, S. (2003). Evidence of motor programming deficits in children diagnosed with DAS. *Journal of Speech, Language, and Hearing Research, 46*(2), 437-450.

Nijland, L., Maassen, B., van der Meulen, S., Gabreëls, F., Kraaimaat, F. W., & Schreuder, R. (2003). Planning of syllables in children with developmental apraxia of speech. *Clinical Linguistics & Phonetics, 17*(1), 1-24.

Pennington, L., Parker, N. K., Kelly, H., & Miller, N. (2016). Speech therapy for children with dysarthria acquired before three years of age. *The Cochrane Library, 7*, 1-37.

Shriberg, L. D., Ballard, K. J., Tomblin, J. B., Duffy, J. R., Odell, K. H., & Williams, C. A. (2006). Speech, prosody, and voice characteristics of a mother and daughter with a 7;13 translocation affecting *FOXP2. Journal of Speech, Language, and Hearing Research, 49*(3), 500-525.

Shriberg, L. D., Fourakis, M., Hall, S. D., et al. (2010). Extensions to the Speech Disorders Classification System (SDCS). *Clinical Linguistics & Phonetics, 24*(10), 795-824.

Shriberg, L. D., Lohmeier, H. L., Strand, E. A., & Jakielski, K. J. (2012). Encoding, memory, and transcoding deficits in childhood apraxia of speech. *Clinical Linguistics & Phonetics, 26*(5), 445-482.

Terband, H., & Maassen, B. (2010). Speech motor development in childhood apraxia of speech: Generating testable hypotheses by neurocomputational modeling. *Folia Phoniatrica Et Logopaedica, 62*(3), 134-142.

Terband, H., Maassen, B., Guenther, F. H., & Brumberg, J. (2009). Computational neural modeling of speech motor control in childhood apraxia of speech (CAS). *Journal of Speech, Language, and Hearing Research, 52*(6), 1595-1609.

Terband, H., Maassen, B., & Maas, E. (2019). A psycholinguistic framework for diagnosis and treatment planning of developmental speech disorders. *Folia Phoniatrica Et Logopaedica, 71*, 216-227.

Terband, H., van Brenk, F., & van Doornik-van der Zee, A. (2014). Auditory feedback perturbation in children with developmental speech sound disorders. *Journal of Communication Disorders, 51*, 64-77.

Ttofari Eecen, K., Eadie, P., Morgan, A. T., & Reilly, S. (2018). Validation of Dodd's Model for Differential Diagnosis of childhood speech sound disorders: a longitudinal community cohort study. *Developmental Medicine & Child Neurology, 61*, 689-696.

Van de Wijer-Muris, I., & Draaisma, Y. (2000). *Cursusreader 'Fonologische stoornissen', version October 2000 [Dutch version of the Hodson Assessment of Phonological Patterns]*. Unpublished document.

Verhoeven, L., & Vermeer, A. (2001). *Taaltoets Alle Kinderen [Dutch language proficiency test for all children]*. Arnhem, the Netherlands: Citogroep.

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 & Communication Disorders, 48*(1), 25-40.

Wren, Y., Miller, L., Peters, T., Emond, A., & Roulstone, S. (2016). Prevalence and predictors of persistent speech sound disorder at eight years old: Findings from a population cohort study. *Journal of Speech, Language, and Hearing Research, 59*(4), 647-673.

7

APPENDIX

APPENDICES

NEDERLANDSE SAMENVATTING

DANKWOORD
CURRICULUM VITAE
DATA MANAGEMENT FORM
PHD PORTFOLIO
PUBLICATION LIST

Nederlandse samenvatting

Profielen van normale en afwijkende spraakproductie beschreven met het Computer Articulatie Instrument (CAI)

Een belangrijke taak voor logopedisten is het onderzoeken van de spraakontwikkeling van een kind: verloopt deze normaal, vertraagd of is er sprake van een stoornis? Spraakstoornissen zijn de meest voorkomende communicatiestoornissen bij kinderen, met prevalentiecijfers die variëren tussen 2.3% en 24.6% (Wren, Miller, Peters, Emond, & Roulstone, 2016). Om kinderen met een vertraagde spraakontwikkeling of spraakstoornis goed te kunnen identificeren en diagnosticeren zijn betrouwbare en valide testinstrumenten met representatieve normaalwaarden essentieel. Daarnaast zijn deze testinstrumenten belangrijk voor het klinisch redeneren en het opstellen van een behandelplan. Kinderen met een spraakstoornis vormen een heterogene groep. De problemen verschillen in ernst, oorzaak, het karakter van de spraakfouten en het effect van de behandeling (Dodd, 2014; Ttofari Eecen, Eadie, Morgan, & Reilly, 2018). De tot nu toe beschikbare tests geven een gedetailleerde beschrijving van de symptomen van spraak. De betrokken onderliggende processen worden hiermee echter niet in kaart gebracht. De belangrijkste processen zijn: het activeren van een lemma (bevat semantische en grammaticale informatie van een woord), selecteren van een lexeem (woordvorm), fonologisch encoderen, spraakmotorische planning en spraakmotorische uitvoering (Terband, Maassen, & Maas, 2019). Een uitgebreide analyse van de prestaties van een kind op een reeks spreektaken, die ieder verschillende onderliggende processen reflecteren, is nodig om inzicht te krijgen in de problemen die mogelijk de onderliggende oorzaak zijn van de spraakstoornis. De diversiteit aan spreektaken zorgt ervoor dat verschillende processen van de spraakontwikkeling worden gemeten, hetgeen een breed en gedifferentieerd beeld oplevert. Op basis van deze uitgangspunten is het Computer Articulatie Instrument (CAI) ontwikkeld (Maassen et al., 2019). Het overkoepelende doel van dit proefschrift is het beschrijven van spraakprofielen van normaal ontwikkelende kinderen en kinderen met vertraagde spraakontwikkeling of spraakstoornis zodat de eerste stap naar procesgeoriënteerde diagnostiek gemaakt kan worden. De doelstellingen van dit proefschrift zijn:

1. De constructie van een testbatterij voor spraakproductie met een pakket spreektaken ten behoeve van het beschrijven van een uitgebreid spraakprofiel.
2. Het verzamelen van normwaarden van spraakprofielen van zich normaal ontwikkelende kinderen en het analyseren van de psychometrische eigenschappen van het CAI in termen van interbeoordelaarsbetrouwbaarheid, test-hertestbetrouwbaarheid en verschillende aspecten van constructvaliditeit.

3. Het beschrijven van een uitgebreide analyse van de spraakontwikkeling van Nederlandssprekende kinderen op basis van een grote cross-sectionele studie zodat klinische testresultaten geïnterpreteerd kunnen worden.
4. Het beschrijven en analyseren van klinische spraakkenmerken van kinderen met een spraakachterstand of specifieke spraakstoornis passend bij verschillende (neurologische) ontwikkelingsstoornissen.

Achtergrondinformatie over spraakontwikkeling, spraakstoornissen en het meten van spraak wordt gepresenteerd in **hoofdstuk 1**. Daarnaast worden de doelstellingen en hoofdlijnen van dit proefschrift beschreven. Dit proefschrift bestaat uit 2 delen. **Deel 1** omvat **hoofdstuk 2, 3** en **4** en richt zich op de constructie van een nieuwe spraaktest, het Computer Articulatie Instrument (CAI), en de beschrijving van de normale spraakontwikkeling van Nederlandssprekende kinderen. **Deel 2** omvat **hoofdstuk 5** en **6** en beschrijft de klinische toepassing van het CAI.

Hoofdstuk 2 richt zich op de psychometrische evaluatie van het CAI. Het CAI is een computergestuurde testbatterij met meerdere spraakproductietaken. De diversiteit aan spreektaken zorgt ervoor dat verschillende processen van de spraakontwikkeling worden gemeten, zodat een genormeerd spraakprofiel voor Nederlandstalige kinderen opgeleverd kan worden. Het CAI meet zowel fonologische als motorische spreekvaardigheden aan de hand van 4 taken: (1) Plaatjes benoemen, (2) Nonwoordimitatie, (3) Woord- en Nonwoordrepetitie en (4) Diadochokinese (DDK). De taak *Plaatjes benoemen* bestaat uit 60 items en bij deze taak is de gehele keten van processen betrokken, van visueel-conceptuele verwerking tot lemmatoegang, woordvormselectie, fonologisch encoderen, spraakmotorische planning en spraakmotorische uitvoer (Maassen & Terband, 2015). Tijdens de taak *Nonwoordimitatie* worden kinderen gevraagd om een nonwoord (of nonsens woord) te herhalen. In tegenstelling tot de taak *Plaatjes benoemen* kan een kind tijdens deze taak geen gebruik maken van het lexicon. Het kind analyseert de fonologische structuur van het nonwoord en maakt vervolgens direct gebruik van het fonologisch encoderingssysteem, of het kiest de route van het auditief-motorische planningssysteem. De taak *Nonwoordimitatie* bestaat voor kinderen van 2 en 3 jaar uit het imiteren van 80 nonwoorden; oudere kinderen in de leeftijd van 4, 5 en 6 jaar wordt gevraagd om 33 nonwoorden te imiteren. Bij de taken *Woordrepetitie* en *Nonwoordrepetitie* dient een kind respectievelijk vijf woorden en vijf nonwoorden met een complexe structuur vijf keer te herhalen. Deze taken testen de consistentie van spraakproductie. Consistentie verwijst naar de mate waarin een uitgesproken woord hetzelfde blijft wanneer het meerdere keren wordt uitgesproken. De *Diadochokinesetaak* bestaat uit drie monosyllabische sequenties (/pa/, /ta/, /ka/), twee bisyllabische sequenties (/pata/, /taka/), en een trisyllabische sequentie (/pataka/). Het kind wordt tijdens deze taak gevraagd om reeksen van deze sequenties zo snel mogelijk uit te spreken. De *Diadochokinesetaken* zijn puur motorische taken, waarvoor geen kennis van woorden, lettergrepen of fonemen nodig is.

Om de psychometrische eigenschappen van het CAI te bepalen, zijn de steekproef voor het normeringsonderzoek, de interbeoordelaarsbetrouwbaarheid, de test-hertestbetrouwbaarheid, de inhoudsvaliditeit en twee aspecten van constructvaliditeit in kaart gebracht. In totaal namen 1524 Nederlandstalige kinderen tussen 2;0 en 7;0 jaar deel aan het normeringsonderzoek. De kinderen van 2 tot 6 jaar werden verdeeld in 12 leeftijdsranges van vier maanden. De kinderen van 6 jaar werden verdeeld in twee leeftijdsranges van een half jaar. De steekproef was representatief voor de Nederlandstalige bevolking wat betreft geslacht, geografische regio, mate van verstedelijking en sociaal-economische status. De vier taken van het CAI werden afgenomen bij alle kinderen. Een set van uitkomstmaten werd samengesteld voor de vier taken. De spraakuitingen van de taken *Plaatjes benoemen* en *Nonwoordimitatie* werden fonetisch getranscribeerd. Verschillende aspecten van spraakproductie konden hiermee onderzocht worden, zoals het foneem-repertoire, het vermogen om combinaties van spraakklanken te produceren (de syllabestructuur) en het voorkomen van fonologische processen. De resultaten van *Woordrepetitie* en *Nonwoordrepetitie* werden weergegeven als de 'proportion of whole word variability' (PWV; Ingram, 2002), die de mate van consistentie reflecteert. De PWV is het aantal verschillende woordvormen dat het kind produceert, gedeeld door het totaal aantal woordvormen. De prestaties op de *Diadochokinesetaak* werden uitgedrukt in het aantal syllabes per seconde en is bedoeld om de spraakmotorische planning en programmering te meten. De interbeoordelaarsbetrouwbaarheid en de test-hertestbetrouwbaarheid werden getest in subgroepen van de normatieve steekproef en werden berekend met behulp van de intraklassecorrelatiecoëfficiënt (ICC). De ICC's voor interbeoordelaarsbetrouwbaarheid varieerden van voldoende tot goed, met uitzondering van de PVC (percentage vocalen correct) voor *Plaatjes Benoemen* en *Nonwoordimitatie* en de productiesnelheid van bi- en trisyllabische reeksen van de *Diadochokinesetaak*. De ICC's voor test-hertestbetrouwbaarheid waren voldoende voor de uitkomstmaten van *Plaatjes benoemen* en *Nonwoordimitatie*. Er werden matige ICC's gevonden voor de uitkomstmaten van *Woord- en Nonwoordrepetitie* en *Diadochokinese*. Betere scores bij de tweede afname door natuurlijke ontwikkeling en leereffecten kunnen hiervoor een verklaring zijn.

De inhoudsvaliditeit van het CAI werd aangetoond door de beschrijving van het testdomein, de vier spreektaken en hun items. De onderbouwing van de taken werd beschreven.

Het eerste aspect van constructvaliditeit betreft de eis dat de uitkomst van een ontwikkelingstest een samenhang met leeftijd laat zien. We veronderstelden dat de geselecteerde uitkomstmaten de normale spraakontwikkeling zouden weerspiegelen en dus een toename van scores met het stijgen van de leeftijd lieten zien. Dit aspect werd onderzocht door de ruwe scores van de vier taken in de normatieve steekproef (n = 1524) van zich normaal ontwikkelende Nederlandstalige kinderen tussen 2;0 en 7;0 jaar met elkaar te vergelijken. Een continue normering liet ontwikkelingspatronen zien voor alle CAI-uitkomstmaten. Het tweede aspect van constructvaliditeit, structurele validiteit, werd bepaald door alle uitkomstmaten van het instrument te testen. Structurele validiteit betekent dat de uitkomstmaten van een taak zich gedragen zoals verwacht volgens een onderliggend model. Om dit aspect van constructvaliditeit te onderzoeken, werden factoranalyses uitgevoerd. Daarbij was de veronderstelling dat combinaties van de geselecteerde uitkomstmaten verschillende aspecten van spraakproductie zouden weerspiegelen, binnen of over de taken heen. De factoranalyse over een totaal van 20 uitkomstmaten resulteerde in vijf betekenisvolle factoren: alle uitkomstmaten van *Plaatjes Benoemen* (PB), de segmentele uitkomstmaten van *Nonwoordimitatie* (NWI-Seg), de syllabestructuur uitkomstmaten van *Nonwoordimitatie* (NWI-Syll), *Woord- en Nonwoordrepetitie* consistentie (PWV) en alle uitkomstmaten van de *Diadochokinesetaak*. Deze resultaten bevestigen het onderscheidend vermogen van de vier taken van het CAI. Elke taak weerspiegelt verschillende aspecten van de spraakproductie. Bovendien wordt de constructvaliditeit onderstreept door de zwakke correlaties tussen CAI-factorscores, wat de onafhankelijke bijdrage van elke factor aan het spraakprofiel weergeeft.

Op basis van de voldoende tot goede psychometrische eigenschappen concludeerden we dat het CAI een betrouwbaar en valide instrument is voor de beoordeling van de spraakontwikkeling bij Nederlandstalige kinderen tussen 2;0 en 7;0 jaar en dat het kan worden gebruikt om de normale en afwijkende spraakontwikkeling te meten.

In **hoofdstuk 3** en **4** van dit proefschrift wordt de beschrijving gegeven van een uitgebreide analyse van de spraakontwikkeling van Nederlandsprekende kinderen, gebaseerd op een grote cross-sectionele studie.

Hoofdstuk 3 geeft een gedetailleerde beschrijving van de spraakontwikkeling van 1503 zich normaal ontwikkelende Nederlandstalige kinderen tussen 2;0 en 7;0 jaar uit de normatieve studie van hoofdstuk 2. De laatste jaren zijn er wereldwijd vele studies uitgevoerd om de normale spraakontwikkeling in verschillende talen te onderzoeken, waaronder een aantal studies die de spraakontwikkeling van Nederlandstalige kinderen onderzochten. Geen van deze laatste studies onderzocht echter zowel de fonetische als de fonologische vooruitgang binnen een voldoende breed leeftijdsbereik en met een voldoende grote steekproef. De studie uit hoofdstuk 3 heeft als doel deze leemte te vullen door het uitvoeren van een cross-sectionele normstudie die de belangrijke uitkomstmaten van de spraakontwikkeling in kaart brengt: percentage consonanten correct-revised

(PCC-R; een aangepaste vorm van PCC waarbij distorsies van klanken als correct worden beschouwd) en percentage vocalen correct (PVC), inventarisaties van consonanten, vocalen en syllabestructuren, graden van complexiteit (fonemische kenmerkhiërarchie), en fonologische vereenvoudigingsprocessen. De taak *Plaatjes benoemen* van het CAI werd gebruikt om deze uitkomstmaten te verkrijgen. PCC-R en PVC namen significant toe met de leeftijd. Het consonant-repertoire was in deze groep compleet op de leeftijd van 3;7 jaar voor de syllabe initiale consonanten, met uitzondering van de stemhebbende fricatieven /v/ en /z/, en de liquida /r/. Alle syllabe-finale consonanten werden verworven voor de leeftijd van 4;4 jaar. Op de leeftijd van 3;4 jaar hadden de kinderen alle vocalen verworven. Op de leeftijd van 4;7 jaar produceerden ze de meeste syllabestructuren correct, hoewel de syllabestructuur CCVCC nog in ontwikkeling was. Alle fonologische contrasten werden op de leeftijd van 3;8 jaar correct geproduceerd. Kinderen in de jongere leeftijdsgroepen gebruikten meer fonologische vereenvoudigingsprocessen dan de oudere kinderen en op de leeftijd van 4;4 jaar waren alle vereenvoudigingsprocessen verdwenen, behalve de initiële clusterreductie van drie naar twee consonanten en de finale clusterreductie van twee naar één consonant. Deze gedetailleerde beschrijving van de normale Nederlandstalige spraakontwikkeling geeft logopedisten relevante informatie om te bepalen of de spraakontwikkeling van een kind normaal, vertraagd of afwijkend verloopt.

In **hoofdstuk 4** worden de normgegevens beschreven van de ontwikkeling van het vermogen tot diadochokinese van Nederlandstalige kinderen op basis van een grote cross-sectionele studie. Bij een groep van 1014 zich normaal ontwikkelende kinderen in de leeftijd van 3;0 tot 6;11 jaar werd de *Diadochokinesetaak* van het CAI afgenomen. Er werd gebruik gemaakt van een gestandaardiseerd protocol dat door onze onderzoeksgroep is voorgesteld in een eerdere studie (Diepeveen, Van Haaften, Terband, De Swart, & Maassen, 2019). Het aantal syllabes per seconde werd berekend voor de mono-, bi- en trisyllabische sequenties (DDK-pa, DDK-ta, DDK-ka, DDK-pata, DDK-taka, DDK-pataka). Voor alle DDK-sequenties nam het aantal syllabes per seconde significant toe met de leeftijd. DDK-pa was de snelste sequentie, gevolgd door respectievelijk DDK-ta, DDK-pata, DDK-taka, DDK-ka en DDK-pataka. Over het algemeen waren de DDK-scores (aantal syllabes per seconde) voor alle DDK-sequenties voor jongens hoger dan voor meisjes. Met de in dit onderzoek gepresenteerde normgegevens zijn logopedisten in staat om kinderen met problemen in de diadochokinese te herkennen. Daarnaast kan de *Diadochokinesetaak* bijdragen aan een eerste stap in de differentiaaldiagnose tussen verschillende soorten spraakstoornissen, en vooral tussen verschillende spraakmotorische stoornissen. De uitgebreide beschrijving van de normale spraakontwikkeling en diadochokinese in hoofdstuk 4 en 5 kan als achtergrond dienen voor de interpretatie van klinische testresultaten.

Het tweede deel van dit proefschrift heeft als doel het beschrijven en analyseren van klinische spraakkenmerken van een vertraagde spraakontwikkeling en specifieke spraakstoornissen bij kinderen met diverse (neurologische) ontwikkelingsstoornissen. De studies beschreven in **hoofdstuk 5** en **6** zijn twee voorbeelden van de klinische toepassing van het CAI. In **hoofdstuk 5** worden gegevens uit twee studies naar klinische groepen kinderen met spraak-taalstoornissen en spraakstoornissen gepresenteerd om de 'known-group' validiteit van het CAI vast te stellen. 'Known-group' validiteit gaat over het vermogen van een meetinstrument om te discrimineren tussen onderscheidende subgroepen. In *studie 1* werd de 'known-group' validiteit onderzocht door de scores van 93 kinderen met spraak-taalproblemen op één taak van het CAI (*Plaatjes benoemen*) te vergelijken met het door logopedisten gegeven oordeel over de verstaanbaarheid. Voor deze studie werd de uitkomstmaat percentage consonanten correct van *Plaatjes benoemen* gebruikt (PB-PCC). Er bleken significante verschillen in PB-PCC voor de drie onderscheiden verstaanbaarheidsniveaus: goed (n=22), matig (n=46), laag (n=25). Ook was er een significante correlatie tussen de verstaanbaarheidsniveaus en PB-PCC in de verwachte richting. Noch het niveau van verstaanbaarheid, noch de prestaties op de taak Plaatjes benoemen waren gerelateerd aan non-verbale intelligentie- en receptieve en expressieve taalscores. *Studie 2* had tot doel de diagnostische waarde van de vier CAI-taken te bepalen door gebruik te maken van de scores op de CAI-factoren die uit het onderzoek in hoofdstuk 2 naar voren kwamen: PB, NWI-seg, NWI-syll, PWV, DDK-Mono en DDK-BiTri werden vergeleken met het klinisch oordeel van logopedisten over de ernst van de spraakmoeilijkheden van 41 kinderen met een gediagnosticeerde spraakstoornis. De meeste kinderen (n = 36) werden gediagnosticeerd met een fonologische stoornis, twee kinderen met spraakontwikkelingsdyspraxie, en drie kinderen hadden geen diagnose omdat details over de spraak van de kinderen niet beschikbaar waren, afgezien van het feit dat hun spraakstoornis ernstig was. Er werd een ernstschaal met drie categorieën gebruikt: mild, matig en ernstig. Er bleken significante verschillen tussen de matige en de ernstige categorie voor de CAI-factoren op basis van *Plaatjes benoemen*, *Nonwoordimitatie* en de bisyllabische en trisyllabische sequenties van de *Diadochokinesetaak*. Er werden geen significante verschillen gevonden tussen de matige en ernstige categorie voor de factor *Woord- en Nonwoordrepetitie* consistentie (PWV) en de monosyllabische sequenties van de *Diadochokinesetaak*. Deze resultaten suggereren dat vooral de taken PB, NWI en de bisyllabische en trisyllabische sequenties van de *Diadochokinesetaak* het gevoeligst zijn voor de diagnose van een spraakstoornis. De 'known-group'-validiteit van het CAI wordt ondersteund door de bevindingen van deze twee studies. Samen met de resultaten van hoofdstuk 2 kunnen we concluderen dat het CAI een betrouwbaar en valide instrument is voor de beoordeling van kinderen met spraakstoornis.

Hoofdstuk 6 beschrijft het communicatiefenotype dat geassocieerd wordt met het syndroom van Koolen de Vries. In dit prospectieve onderzoek werd bij alle deelnemende, Nederlandstalige kinderen het CAI afgenomen. In totaal werden 29 deelnemers met het Koolen de Vries syndroom (vier met KANSL1-varianten, 25 met 17q21.31 microdeletie) beoordeeld op mondmotoriek, spraak, taal, geletterdheid en sociaal functioneren. De linguïstische fenotypering liet een onderscheidend communicatieprofiel zien. De deelnemers toonden een doorslaggevende 'double hit' van orale hypotonie en spraakontwikkelingsapraxie in de kleuter- en voorschoolse leeftijd, samen met een sterk vertraagde spraakontwikkeling. Opmerkelijk genoeg was de spraakprognose echter positief; in de late kinderleeftijd was de spraakontwikkelingsdyspraxie niet meer aanwezig, en hoewel de dysartrie bleef bestaan, waren de kinderen beter verstaanbaar. Daarentegen bleven de taal- en leesachterstanden bestaan en waren er pragmatische taalproblemen aanwezig.

Referenties

Diepeveen, S., Van Haaften, L., Terband, H., De Swart, B., & Maassen, B. (2019). A standardized protocol for maximum repetition rate assessment in children. *Folia Phoniatrica Et Logopaedica, 71*, 238-250.

Dodd, B. (2014). Differential diagnosis of pediatric speech sound disorder. *Current Developmental Disorders Reports, 1*(3), 189-196.

Ingram, D. (2002). The measurement of whole-word productions. *Journal of Child Language, 29*(4), 713-733.

Maassen, B., & Terband, H. (2015). Process-oriented diagnosis of childhood and adult apraxia of speech (CAS and AOS). In M.A. Redford (Ed.), *The Handbook of Speech Production* (Vol. First Edition, pp. 331-350): Wiley.

Maassen, B., van Haaften, L., Diepeveen, S., Terband, H., van den Engel-Hoek, L., Veenker, T., & De Swart, B. (2019). *Computer Articulation Instrument.* Amsterdam: Boom Uitgevers.

Terband, H., Maassen, B., & Maas, E. (2019). A psycholinguistic framework for diagnosis and treatment planning of developmental speech disorders. *Folia Phoniatrica Et Logopaedica, 71*, 216-227.

Ttofari Eecen, K., Eadie, P., Morgan, A. T., & Reilly, S. (2018). Validation of Dodd's Model for Differential Diagnosis of childhood speech sound disorders: a longitudinal community cohort study. *Developmental Medicine & Child Neurology, 61*, 289-696.

Wren, Y., Miller, L., Peters, T., Emond, A., & Roulstone, S. (2016). Prevalence and predictors of persistent speech sound disorder at eight years old: Findings from a population cohort study. *Journal of Speech, Language, and Hearing Research, 59*(4), 647-673.

Dankwoord

Daar zit ik dan: lekker in het zonnetje op mijn dak, uitkijkend op de skyline van Nijmegen en eindelijk klaar om het dankwoord van mijn proefschrift te schrijven. Wat begon als een klein onderzoeksproject groeide uit tot een groter project en leidde uiteindelijk tot dit proefschrift en mijn promotie. Het was geen vooropgezet plan, maar ik ben blij dat het zo gegaan is.

Het ontwerp van de omslag van dit boekje is geïnspireerd op de horizon en kleuren van de zee (dank je wel Ilse Schrauwers van isontwerp.nl!). De letters die de titel vormen zijn nog niet 'af' of net anders dan gebruikelijk, net als de spraak van jonge kinderen in ontwikkeling of de spraak van kinderen met een spraakstoornis. Hier op mijn dak is door de gebouwen en bomen de horizon niet te zien, maar afgelopen jaren heb ik door mijn promotietraject mijn horizon kunnen verbreden, mijn interesse in wetenschappelijk onderzoek kunnen ontdekken en geprobeerd mijn onderzoekersvaardigheden steeds meer 'af' te maken. Dit is gelukt doordat in ons project de klinische toepasbaarheid van het onderzoek steeds op de voorgrond stond. De combinatie van het werken als onderzoeker en het klinisch werken als logopedist met jonge kinderen en hun ouders hebben er voor gezorgd dat ik dit proefschrift met plezier heb kunnen afronden. De fijne samenwerking met collega's en inspirerende omgeving hebben daar ontzettend bij geholpen!

Allereerst wil ik mijn promotor Prof. dr. B.A.M. Maassen bedanken. Beste Ben, dank je wel voor de fijne begeleiding in de afgelopen jaren. Zonder jou was ik niet op dit pad terecht gekomen en onder jouw begeleiding ben ik uitgegroeid tot de onderzoeker die ik nu ben. Ik ben onder de indruk van jouw kennis en altijd had ik na een overleg met jou weer inspiratie om verder te gaan met schrijven. Fijn dat je me het vertrouwen gaf dat het geen 'rocket science' hoefde te worden en een toegepast onderzoek kon blijven.

Dr. B.J.M. De Swart, beste Bert. Als hoofd van onze afdeling logopedie en mijn copromotor, heb je ontzettend veel betekend en geregeld voor dit onderzoeksproject. Door jouw inzet heb ik afgelopen jaren de ruimte gekregen voor het doen van onderzoek en afmaken van dit traject. Bedankt voor je vertrouwen!

Het klinkt cliché, maar dit boekje zou er ECHT nooit zijn geweest zonder 'ons team' logopedie-kinderen: Lenie van den Engel-Hoek, Marjo van Gerven, Sandra de Groot, Karen van Hulst en Marloes Lagarde. Ik voel me zo thuis bij jullie! Ik ben dankbaar om te kunnen werken in een team met zoveel aandacht voor de patiënten, ouders, het vak logopedie en elkaar.

Lieve Lenie, wat een geluk dat ik jou als copromotor, daily supervisor, collega, mentor, maar vooral ook als kamergenoot heb gehad de afgelopen jaren. Onze 6

oktober kamer! Vanaf het begin van mijn carrière zit ik al naast je en dat voelt heel vertrouwd. Ik heb zoveel van je geleerd; over het werk en het leven. Ik geniet van onze gesprekken over de kinderen en hun ouders. Van hoe we hun problemen vanuit alle kanten kunnen belichten. We hebben soms 's ochtends gelijk bij binnenkomst al diepgaande gesprekken, maar we kunnen ook goed kletsen over de kleine dingen. En ik weet dat ik het niet van je mag zeggen: maar wat ga ik je ontzettend missen als je in januari met pensioen gaat!

Lieve Marjo, bedankt voor de tijd die je altijd hebt voor een gesprek, dank je voor je luisterend oor, je goede raad, je kritische blik, bemoedigende adviezen en vooral je humor. Dat geeft me altijd veel inzicht! Heel fijn om het onderwerp 'spraak(motoriek)' met je te kunnen delen.

Lieve Sandra, we zijn 17(!) jaar geleden samen begonnen in het Radboudumc. Vanaf dat moment trekken we samen op en daar zijn we niet meer mee gestopt. Van je enthousiasme, energie en besluitkracht leer ik veel. Ik heb veel zin in de projecten in de kliniek die we kunnen gaan oppakken nu mijn proefschrift klaar is. Zullen we nog heel lang ons jaarlijkse jubileum vieren met lekker eten, een wijntje (chardonnay of was het toch sauvignon blanc?) en eindeloos kletsen?

Lieve Karen, jouw altijd positieve uitstraling heeft mij ontzettend geholpen tijdens dit promotietraject. Jij blijft in elke lastige situatie de mogelijkheden zien. Wat was het fijn om de laatste jaren samen op te kunnen gaan in het schrijven van artikelen en de laatste hoofdstukken van onze proefschriften. Onze gezamenlijke schrijfdagen (met veel lekkers!) waren heel motiverend. Heel fijn om aan het eind van de dag / begin van de avond bij je binnen te kunnen lopen om van alles te bespreken. Dank je wel voor al je inspirerende woorden en adviezen!

Lieve Marloes, dank je wel voor je oprechte betrokkenheid en dat je zo een fijne collega bent. Samenwerken met jou gaat zo makkelijk en vanzelfsprekend en is daarmee fijn en bijzonder. Het is heel fijn om met jou te brainstormen over alles wat komt kijken bij een promotie (en over waar je het beste kunt eten in Nijmegen, huisinrichting enz.…). Op naar jouw boekje en feestje!

Mijn dank gaat natuurlijk ook uit naar Sanne Diepeveen. Lieve Sanne, we zijn sinds de start van dit project een echt team. Een goed duo in de koppel-structuur van de HAN en de afdeling Revalidatie van het Radboudumc. We hebben samen best al wat mooie producten afgeleverd: het CAI, de CAI-cursus en 8 artikelen. Dank je wel voor alle lange-, korte- en tussendoor-overlegmomenten! Ik kijk uit naar jouw promotie!

De leden van de manuscriptcommissie wil ik hartelijk bedanken voor de tijd en aandacht die zij aan dit proefschrift hebben geschonken: Prof. dr. M.A.A.P Willemsen, Prof. dr. J.P.M. Fikkert en Dr. R. Jonkers.

Ook dank ik alle medeauteurs van de afzonderlijke artikelen voor de waardevolle suggesties, aanvullingen, nieuwe ideeën en hulp bij de statistische berekeningen. Bedankt dat jullie dit proefschrift naar een hoger niveau hebben getild.

Mijn interesse in de wetenschap is aangewakkerd tijdens het schrijven van mijn master scriptie die werd begeleid door Lian Nijland. Lian, dank je wel dat je mij hebt aangestoken met jouw enthousiasme voor onderzoek naar spraakstoornissen bij kinderen.

Lieve dames van het secretariaat, Marieke, Daniëlla en Christine, het is heel fijn om bij jullie binnen te lopen en altijd een luisterend oor te kunnen vinden. Dank jullie wel daarvoor!

Naast bovengenoemde collega's wil ik ook alle collega's van de afdeling kinderrevalidatie bedanken. Collega's van de kinderfysiotherapie, ergotherapie en kinder-revalidatieartsen, dank voor de steun tijdens mijn promotieperiode en de fijne samenwerking. Ook de collega-logopedisten van de 'volwassen-kant' wil ik bedanken voor hun interesse en samenwerking. Het is prettig brainstormen tijdens het logopedie-onderzoekersoverleg!

Prof. dr. Sander Geurts, bedankt voor de ruimte die ik heb gekregen om op de afdeling Revalidatie mijn onderzoek te kunnen doen.

Veel dank gaat uit naar alle kinderen, ouders, kinderdagverblijven/peuterspeelzalen en basisscholen die mee hebben gewerkt aan het normeringsonderzoek van het CAI. Ik wil de logopedisten en studenten die betrokken waren bij dit project ontzettend bedanken: er is door jullie bergen werk verricht (zoals testafnames, transcriberen, syllabes tellen, en nog een keer transcriberen, analyseren). Zonder jullie waardevolle bijdrage was het nooit gelukt om zo een grote groep kinderen (1658!) te onderzoeken.

Lieve vrienden en familie, bedankt voor jullie interesse in mijn promotie en vooral voor alle gezelligheid en fijne gesprekken in de afgelopen jaren tijdens onder andere etentjes, wandelingen en weekendjes weg.

Hanneke, dank je wel voor alle borrels, bitterballen en voor wie je bent! Ik heb geluk met een vriendin zoals jij!

Nienke, onze wegen zijn zich vanaf de opleiding logopedie gaan kruisen. Samen logopedie-student, samen TSP-student en allebei onze plek gevonden in Nijmegen. Dank je wel voor de warme vriendschap!

Meiden van de volley, Ellis, Janine, Linda, Lotte, Manon, Mieke, Paulien en Wieteke, een avond met jullie is de beste afleiding van het schrijven van een proefschrift. We delen lief en leed, maar vooral heel veel lol! Dank jullie wel!

Mirande, lieve buuf, dank je wel voor al onze hardlooprondjes en de tussendoor-gesprekjes, allebei hangend in de deurpost van onze voordeur.

Annemiek, onze surf-reisjes naar Bali en Marokko waren een geweldige onderbreking van het schrijfwerk. Wanneer zullen we weer? Net zoals het jaarlijkse weekendje weg met de moeders, tantes, zussen en nichtjes. Die dagen zijn zo gezellig en waardevol. Ik kwam elke keer met energie terug. Hopelijk houden we deze traditie nog lang in stand!

Marlon, dank je wel voor de fijne online yogalessen tijdens de laatste fase. Door jouw lessen vroeg (!) in de ochtend heb ik de discussie van dit proefschrift met een goede focus kunnen afmaken!

Lieve Richard, dank je wel voor je interesse, voor je rake vragen over mijn proces waardoor ik aan het denken werd gezet en weer verder kon. Dank je wel ook voor alle keren dat ik bij jullie aan tafel op vrijdagmiddag het weekend kon inluiden. Ik ben blij met jou als zwager!

Liefste Janne en Moos, misschien hebben jullie er nog wel het meeste voor gezorgd dat ik ontspannen door alle promotie-perikelen heen ben gekomen. Ik ben supertrots dat ik 'tante / tant / tantie' ben van het liefste en grappigste nichtje en neefje ooit!

Lieve Pap en mam, dank jullie wel voor jullie onvoorwaardelijke steun en de stevige "Zeeuwse" basis die jullie me gegeven hebben! Nieuwsgierigheid, het hebben van een brede blik en 'afmaken waarmee je begint' zijn met de paplepel ingegoten en die eigenschappen hebben zeker geholpen bij het afronden van dit proefschrift! Heel fijn dat we alles (en nu ook letterlijk van dichtbij) met elkaar kunnen delen.

Het laatste stukje is voor mijn twee paranimfen. Ik ben zo trots en blij dat jullie dicht naast me staan tijdens de verdediging en vooral ook in het leven.

Lieve Lotte, vanaf de eerste dag op de opleiding logopedie is het 'aan' tussen ons. En sindsdien hebben we elkaar niet meer losgelaten. Natuurlijk ben jij mijn paranimf! Dank je wel voor onze onvoorwaardelijke vriendschap!

Lieve Kobie, lieve Ko, hoe kan ik nu opschrijven wat je voor me betekent of waarvoor ik je allemaal wil bedanken? Het is ontzettend veel, en daarom houd ik het maar simpel: ik kan me geen betere zus en nu ook paranimf wensen!

Dit boekje is klaar, op naar nieuwe horizonnen en nieuwe plannen!

Curriculum vitae

Leenke van Haaften was born in Breda, the Netherlands, on October 6th, 1980. She grew up in Zegge and Kapelle. After her graduation from secondary school (Het Goese Lyceum in Goes) in 1999, she started her bachelor training in Speech and Language Therapy at the HAN University of Applied Science in Nijmegen. After graduation in 2003 she started working as a speech-language therapist (SLT), first at the department of Pediatric Neurology, and since 2011 at the department of Rehabilitation of the Radboud university medical center, Amalia Children's Hospital, in Nijmegen. The paediatric SLT team is involved in research and diagnostic assessment and treatment of infants and children with complex oral motor disorders and speech and language disorders. Based on scientific research the team has developed several disease specific assessment and treatment trajectories for children with neurological disorders and syndromes. The team combines patient care with scientific research, that resulted in several studies and many publications.

During the first years of her career, Leenke combined working as an SLT with studying Speech and Language Pathology at the Radboud University in Nijmegen. She obtained her Master degree in 2007. After that, she became involved in Ben Maassen's research project on the development of the Computer Articulation Instrument (CAI). This gave her the opportunity to start with a PhD trajectory in combination with her clinical work. In parallel with this PhD trajectory, the CAI was published and became available for Dutch SLTs in 2019. Together with her colleague Sanne Diepeveen, Leenke developed a CAI post-graduated course and they implemented the CAI in the bachelor programme Speech and Language Therapy at the HAN University of Applied Sciences. At the moment, Leenke is participating in a number of research projects in which the CAI is used to describe speech production abilities in children with developmental disorders and syndromes.

Data management form

General information about the data collection

This research project involves human subject data. Participants volunteered to participate, and anonymity and confidentiality were assured. Written informed consent for collecting these data was obtained from all parents or legal representatives of the participants. The research ethics committee of the Radboud university medical center, Nijmegen stated that this research project (Chapter 2-5) does not fall within the remit of the Medical Research Involving Human Subjects Act (WMO). Therefore, the studies could be carried out (in the Netherlands) without an approval by an accredited research ethics committee. Data were collected and stored at the Radboud university medical center and the HAN University of Applied Sciences.

FAIR principles

Findable

The raw and processed data and accompanying files (descriptive files, syntax files etc.) of this research project are stored in a folder on the server of the department of Rehabilitation at Radboud university medical center (Q:\Research\041 CAI). This folder is only accessible by the main researchers of this project. Documentation to describe the datasets is provided on the department server. The privacy of the participants is warranted by use of encrypted and unique individual subject codes.

Accessible

Only members of the research group have access to the databases. Paper data are stored in the archive of the HAN University of Applied Sciences.

It is not yet possible to make the data available in a public repository because participants only gave informed consent to use their data for purposes as explained on the signed informed consent form. However, requests for data can be made by contacting the staff secretary of the Department of Rehabilitation of the Radboud university medical center (secretariaatstaf.reval@radboudumc.nl). A suitable way to share the data will then be sought. In the future it will be explored how our data can be published in a public repository.

Interpretable

Documentation was added to the data sets to make them interpretable. The documentation contains links to publications, references to the location of the data sets and descriptions of the data sets. The data was stored in SPSS format. No existing data standards were used such as vocabularies, ontologies or thesauri.

Reusable
The data will be stored for at least 10 years and can therefore also be reused in this time period.

PhD Portfolio

Name PhD student	Leenke van Haaften
Department	Radboud university medical center, Department of Rehabilitation
Graduate school	Donders Graduate School for Cognitive Neuroscience
Promotor	Prof. Dr. B.A.M. Maassen
Co-promotors	Dr. L. van den Engel-Hoek Dr. B.J.M. de Swart

Activities	Year	ECTS
Courses and workshops		
Donders introduction course, Radboud university medical center	2015	0.6
Academic writing, Radboud university medical center	2015	3
Schrijven van wetenschappelijke teksten	2013	3
Mindfullness-based stress reduction voor promovendi	2018	2
Writing week, department of Rehabilitation, Radboudumc	2017, 2018, 2019	6
Teaching (Guest lectures, workshops and supervision of student-trainees)		
Bachelor Speech Language Therapy, HAN University of Applied Sciences	2010-2020	50
Master Speech and Language Pathology, Radboud University Nijmegen	2010-2020	20
Symposia, Congresses and Conferences		
Oral presentations		
Annual congress NVLF – Ede	2009	0.5
PAOG Logopediesymposium – Nijmegen	2011	0.5
8th CPLOL congress – The Hague	2012	1
29th World congress IALP – Torino, Italy	2013	1
Het WAP symposium EHealth4com – Nijmegen	2013	0.5
Annual congress NVSST – Utrecht	2013	0.5
Simeacongres – Lunteren	2013	0.5
Annual congress NVLF - Nieuwegein	2017	0.5
Noorddag logopedie – Groningen	2017	0.5
PAOG Logopediesymposium – Nijmegen	2017	0.5
FENAC conference – Dalfsen	2018	0.5

Poster presentations

7th CPLOL congress – Ljubljana, Slovenia	2009	1
28th International IALP congress – Athens, Greece	2010	1
6th International Conference on Speech Motor Control - Groningen	2011	1
Annual congress NVLF – Nieuwegein	2011	0.5
8th CPLOL congress – The Hague	2012	1
34th Annual Congress VVL – Berchem, Belgium	2013	0.5
7th International Conference on Speech Motor Control – Groningen	2017	1

Publication list

1. Van den Engel-Hoek, L., van Hulst, K., van Gerven, M., **van Haaften, L.**, & De Groot, S. (2014). Development of oral motor behaviour related to the skill assisted spoon feeding. *Infant Behaviour and Development, 37(2)*, 187-191.
2. Morgan, A. T., **van Haaften, L.**, van Hulst, K., Edley, C., Mei, C., Tan, T. Y., Amor, D., Fisher, S.E., & Koolen, D. A. (2018). Early speech development in Koolen de Vries syndrome limited by oral praxis and hypotonia. *European Journal of Human Genetics, 26*, 75-84.
3. Van den Engel-Hoek, L., van Gerven, M., **van Haaften, L.**, De Groot, S., Lagarde, M., & van Hulst, K. (2018). *Eet- en drinkproblemen bij jonge kinderen: een leidraad voor logopedisten en ander hulpverleners in de gezondheidszorg.* Assen: Uitgeverij Koninklijke Van Gorcum.
4. Diepeveen, S., **van Haaften, L.**, Terband, H., De Swart, B., & Maassen, B. (2019). A standardized protocol for maximum repetition rate assessment in children. *Folia Phoniatrica Et Logopaedica, 71*, 238-250.
5. Maassen, B., **van Haaften, L.**, Diepeveen, S., Terband, H., van den Engel-Hoek, L., Veenker, T., & De Swart, B. (2019). *Computer Articulation Instrument*. Amsterdam: Boom Uitgevers.
6. Van den Engel-Hoek, L., Lagarde, M., van Gerven, M., **van Haaften, L.**, Admiraal, R.J.C., & Erasmus, C.E. (2019). Mastication problems and dysphagia in 4 patients with Treacher Collins syndrome due to affected orofacial muscles. *Radiology and Medical Diagnostic Imaging, 2(3)*, 1-5.
7. **Van Haaften, L.**, Diepeveen, S., van den Engel-Hoek, L., Jonker, M., de Swart, B., & Maassen, B. (2019). The psychometric evaluation of a speech production test battery for children: The reliability and validity of the computer articulation instrument (CAI). *Journal of Speech, Language, and Hearing Research, 62(7)*, 2141-2170.
8. **Van Haaften, L.**, Diepeveen, S., Terband, H., Vermeij, B., van den Engel-Hoek, L., de Swart, B., & Maassen, B. (2019). Profiling speech sound disorders for clinical validation of the computer Articulation instrument. *American Journal of Speech-Language Pathology, 28(2S)*, 844-856.
9. Diepeveen, S., **van Haaften, L.**, Terband, H., De Swart, B., & Maassen, B. (accepted for publication) Clinical reasoning for SSDs: Diagnosis and intervention in SLPs' daily practice. *American Journal of Speech-Language Pathology.*
10. **Van Haaften, L.**, Diepeveen, S., van den Engel-Hoek, L., De Swart, B., & Maassen, B. (accepted for publication). Speech sound development in Dutch typically developing children: A normative cross-sectional study. *International Journal of Language & Communication Disorders.*

11. **Van Haaften, L.**, and Diepeveen, S., Terband, H., van den Engel-Hoek, L., De Swart, B., & Maassen, B. (revised version submitted). Maximum repetition rate in a large cross-sectional sample of typically developing Dutch-speaking children*. International Journal of Speech-Language Pathology.*

Donders Graduate School for Cognitive Neuroscience

For a successful research Institute, it is vital to train the next generation of young scientists. To achieve this goal, the Donders Institute for Brain, Cognition and Behaviour established the Donders Graduate School for Cognitive Neuroscience (DGCN), which was officially recognised as a national graduate school in 2009. The Graduate School covers training at both Master's and PhD level and provides an excellent educational context fully aligned with the research programme of the Donders Institute.

The school successfully attracts highly talented national and international students in biology, physics, psycholinguistics, psychology, behavioral science, medicine and related disciplines. Selective admission and assessment centers guarantee the enrolment of the best and most motivated students.

The DGCN tracks the career of PhD graduates carefully. More than 50% of PhD alumni show a continuation in academia with postdoc positions at top institutes worldwide, e.g. Stanford University, University of Oxford, University of Cambridge, UCL London, MPI Leipzig, Hanyang University in South Korea, NTNU Norway, University of Illinois, North Western University, Northeastern University in Boston, ETH Zürich, University of Vienna etc.. Positions outside academia spread among the following sectors: specialists in a medical environment, mainly in genetics, geriatrics, psychiatry and neurology. Specialists in a psychological environment, e.g. as specialist in neuropsychology, psychological diagnostics or therapy. Positions in higher education as coordinators or lecturers. A smaller percentage enters business as research consultants, analysts or head of research and development. Fewer graduates stay in a research environment as lab coordinators, technical support or policy advisors. Upcoming possibilities are positions in the IT sector and management position in pharmaceutical industry. In general, the PhDs graduates almost invariably continue with high-quality positions that play an important role in our knowledge economy.

For more information on the DGCN as well as past and upcoming defenses please visit: ***http://www.ru.nl/donders/graduate-school/phd/***