Utrecht Institute of Linguistics OTS

Utrecht University

An explorative study of the error profile of children with DLD: Retention difficulties with phonemic item and order information during nonword repetition

Master Thesis

presented by

Rosanne van den Berg

The study was supervised by

Dr. Tessel Boerma (first reader)

Prof. dr. Elma Blom (second reader)

Utrecht University, the Netherlands

Student number: 6371019

Research Master of Linguistics

Date of submission: July 11, 2020

Abstract

It has been widely reported that children with developmental language disorder (DLD) are outperformed by their TD peers regarding their abilities to repeat nonwords. The poor nonword repetition of children with DLD has been suggested to reflect limitations in phonological short-term memory (pSTM), a memory system in which verbal material can temporarily be stored. However, it is unclear whether their problems with repeating nonwords are situated at the level of retaining items (i.e. phonological characteristics), order (i.e. the phonemes' position), or both. The present paper aimed to address this question, to gain more insight in the kind of retention problems that children with DLD experience, which is relevant for clinical purposes. Additionally, insight in the error patterns is theoretically of interest, as it sheds light on the question whether children with DLD differ from their TD peers by a delay or a deviance. In total 39 Dutch children with TD and 39 Dutch children with DLD participated in a nonword repetition task (NWRT) at 72 and 95 months of age. The children were matched on age and nonverbal IQ. Their nonword repetitions were analysed in terms of accuracy, and error patterns. Results revealed that with increasing age the groups improved at the same pace, but at both ages the DLD group was outperformed by the TD group. Error patterns did not change with age. Although children with DLD made relatively more combined errors, the overall error profiles of the two groups were largely the same: Item errors were most frequent, followed by combined errors, which were in turn followed by order errors. These findings suggest that the retention problems of children with DLD are similar to the retention problems of their TD peers, but that the problems of the DLD group are more severe. Retention problems were mainly situated at the level of item information, and to a lesser extent at the level of order information.

Keywords: **developmental language disorder (DLD), nonword repetition, phonological short-term memory, error analysis, development**

Acknowledgements

First and foremost, I would like to thank my supervisors Prof. dr. Elma Blom and Dr. Tessel Boerma for their critical thinking, their feedback, and their help, which definitely elevated this thesis to a higher level. I am very grateful that they offered me the opportunity to use a dataset of one of their own research projects for my thesis, as I could not continue with my original project due to the circumstances surrounding COVID-19. I enjoyed working on this thesis, and it has truly strengthened my enthusiasm and interest in the field of DLD and other (language) disorders.

I also would like to thank my fellow students from the basement (Janskerkhof 13A) with whom I could not only exchange thoughts, but who made writing this thesis also more fun. Finally, I would like to thank my family and friends for just listening to me thinking out loud, for moments of relaxation, and for their support in general.

Table of contents

1. Introduction

For typically developing (TD) children, acquiring a first language is a seemingly quick and effortless process (e.g. Leonard, 2014). In contrast, for children suffering from a developmental language disorder (DLD)[1], language development is accompanied by difficulties, and these children do not learn their mother tongue effortlessly. DLD is a neurodevelopmental language disorder, affecting approximately 7% of the population, with males being more frequently diagnosed than females (Leonard, 2014; Tomblin et al., 1997). The disorder is characterised by limitations in language abilities in different linguistic domains, like phonology, morphology, syntax, semantics, or pragmatics, and the language profiles of individuals with DLD are heterogeneous (e.g. Botting & Conti-Ramsden, 2004; Brooks & Kempe, 2012; Leonard, 2014). Their language difficulties cannot be explained by low nonverbal IQ scores, hearing impairment, neurological damage, or deprived language input.

The nature of the disorder is still subject to debate, but a storage limitation within children's phonological short-term memory (pSTM) has widely been suggested to underlie DLD (e.g. Gathercole & Baddeley, 1990; Coady & Evans, 2008). pSTM capacity can be tapped by the nonword repetition task (NWRT), in which children have to repeat a nonword immediately after hearing it. Repeating this novel sequence requires the retention of the phonemic items, as well as their order.

Many studies have demonstrated that children with DLD are outperformed by their TD peers on repeating nonwords (e.g. Graf Estes, Evans, & Else-Quest, 2007). However, few studies have looked at what types of errors children with DLD make relative to TD children. To the best of my knowledge no studies have yet explored the error profiles of children with DLD as categorised by retention problems with phonemic item information (i.e. phonological

---

[1] DLD has also been referred to as Specific Language Impairment (SLI), Language Impairment (LI), or Primary Language Impairment (PLI). For a detailed discussion, see Reilly, Bishop, and Tomblin (2014).

characteristics) and/or order information (i.e. the phoneme's position). The main purpose of the present study is to address this issue. NWRT performance accuracy and error patterns will be analysed by focussing on consonants. Additionally, the effect of syllable length and age (i.e. development from 72 to 95 months of age) on accuracy and error patterns will be examined.

Insight in the abilities to retain item and order information in pSTM during nonword repetition contributes to a better understanding of the kind of retention problems that children with DLD experience. This insight is crucial to develop effective interventions. Moreover, if differences between the error patterns of the two groups emerge, this could be used as a clinical marker to identify children with DLD. Insight in the differences between the error patterns of children with and without DLD is also theoretically of interest, as it will reveal whether children with DLD differ from their TD peers in terms of a delay or a deviance.

The remainder of this paper is as follows. In chapter 2 a theoretical framework will be provided. The typical phonological development of speech perception and speech production will briefly be described. This is relevant for interpreting the difficulties that children with DLD could experience in the light of typical development. Next, an overview of the phonological difficulties that children with DLD could experience will be given. Thereafter, a speech processing model based on the model of Baddeley and Hitch (2019), and Jacquemot and Scott (2006) will be discussed, which will be related to the question how pSTM is related to nonword repetition performance. Then, previous research on the abilities of children with DLD to retain item and order information will be discussed, as both types of information need to be retained during nonword repetition. In chapter 3 the present study will be described, followed by a description of the method in chapter 4. Subsequently, in chapter 5 the results will be presented, which will be discussed in chapter 6. This chapter also contains limitations of the present study, directions for future research, and clinical implications. Finally, chapter 7 finishes with the conclusion, including a summary of the main findings.

2. Theoretical background

2.1. Typical phonological development

2.1.1. Speech perception

Newborns show a preference to listen to speech over non-speech (Vouloumanos & Werker, 2004, 2007). They can discriminate essentially between all phonetic contrasts that exist within all languages, but between 6 and 12 months of age this ability becomes limited to the sounds of their native language (Werker, 1989). Statistical distributional information within the linguistic input helps the child to discover these sounds of the native phonemic inventory (Maye, Werker, & Gerken, 2002; Fikkert, 2007). Due to this language-specific categorical perception all irrelevant variations within different productions of a particular phoneme can be ignored, which is beneficial for language processing.

Another step in the development of speech perception is to detect words within the speech stream, which infants are capable of doing from around 6 to 8 months (Bortfeld, Morgan, Golinkoff, & Rathbun, 2005; Jusczyk & Aslin, 1995). Different cues could help to detect word boundaries. Infants of 8 months can use transitional probabilities between syllables; the statistical likelihood that a particular syllable follows the preceding syllable (Saffran, Aslin, & Newport, 1996). For example, in the Dutch phrase *lieve baby* 'sweet baby' the probability of the syllable *lie* being followed by *ve* is higher than the probability of the syllable *ve* being followed by *ba*. In addition, 8-month-old infants can use prosodic cues like lexical stress (Johnson & Jusczyk, 2001). By 9 months of age, infants can use patterns of phonotactic probabilities, which is the statistical likelihood that particular sequences of phonemes co-occur (Mattys & Jusczyk, 2001). To illustrate for Dutch, the sequence *kui* (e.g. *kuiken* 'chick') occurs less frequently than the sequence *raa* which is very frequent (e.g. *raam* 'window', *raaf* 'raven', *raadsel* 'riddle') (Adriaans, 2006, as cited in Rispens & Baker, 2012). Which segmentation

strategy is preferred changes with age, and when infants grow older they will rely more heavily on an integration of multiple cues (Thiessen & Saffran, 2003; Johnson & Seidl, 2009).

Detecting words and word learning is supported by phonological knowledge, but at the same time an expanding vocabulary facilitates phonological development (Stoel-Gammon, 2011). Some researchers suggest that the child's *phonological representations*, or "sound-based codes" with which words are stored in the lexicon, are not adult-like from the start, but rather holistic, or phonemically undetailed (Claessen & Leitão, 2012, p. 212; Bonte & Blomert, 2004; Coady & Aslin, 2003, 2004; Treiman & Baron, 1981; Walley, 1993). During lexical development the phonological representations become more refined with increasing detail on the level of syllables and phonemes. In contrast, other researchers suggest that the child's phonological representations are phonemically detailed from the start (Jusczyk & Aslin, 1995; Swingley & Aslin, 2002). The differences between these accounts will not be elaborated upon, but for the purpose of the present paper it is important that both accounts claim that by the time the child has "acquired a considerable native language vocabulary" these phonological representations are segmented into sequences consisting of individual phonemes (Schraeyen, 2018, p. 9).

2.1.2. Speech production

The development of speech production starts with utterances like cries or coughs (Stoel-Gammon & Sosa, 2007). At the age of 2 to 3 months infants begin to produce vowel-like sounds, and around 6 to 8 months the babbling period starts, in which they produce consonant-vowel (CV) sequences of one or two syllables. Between 6 and 12 months, the infant's repertoire of consonants expands substantially, and becomes more language-specific (Zsiga, 2013). Around 1 year of age infants generally produce their first meaningful words, after which

babbling and the early word productions co-exist for a few months (Stoel-Gammon & Sosa, 2007).

The child's early word productions are characterised by a great variability (Sosa & Stoel-Gammon, 2012; Zsiga, 2013). A particular sound may be pronounced correctly in one word, but incorrectly in another, or a specific word may be pronounced in different ways. Overall, the developmental trajectory of vowel acquisition seems to be more accurate from the outset compared to the acquisition of consonants (Donegan, 2013; Mennen, Levelt, & Gerrits, 2007). For instance, from the start front vowels (i.e. /i/) are produced in the front of the mouth, and back vowels (i.e. /ɑ/) in the back. Consonants are more prone to errors regarding their place of articulation. For example, children commonly replace /k/ (produced in the back) by /t/ (produced in the front), which is called *fronting* (e.g. **k**aas 'cheese' becomes **t**aas) (Beers, 1995). Phoneme substitutions commonly occur in the child's speech (Beers, 1995, 2003; Zsiga, 2013). Likewise, simplification processes like cluster reduction (e.g. **s**ter 'star' becomes **t**er), weak syllable deletion (i.e. **ge**daan 'done' becomes daan), final consonant deletion (e.g. da**k** 'roof' becomes da), or reduplication (e.g. bal 'ball' becomes baba), are typically applied to simplify the phonological structure of a word. The use of such simplification processes gradually decreases, and at around 3 to 4 years of age the child's phonological system appears to be similar to the adults' one (Beers, 1995, 2003; Zsiga, 2013).

2.2. DLD and phonological problems

Many children with DLD experience phonological problems in comprehension, production, or both (Leonard, 2009; Schwartz, 2009). Compared to TD children they have poorer speech perceptual abilities, comprising weaknesses in categorical perception (Gerrits & de Bree, 2009; Schwartz, Scheffler, Lopez, 2013), perception of lexical stress (Richards & Goswami, 2015), perception of speech at fast rate (Guiraud, Bedoin, Krifi-Papoz, Herbillon, Caillot-Bascou,

Gonzalez-Monge, & Boulenger, 2018), or problems with noise exclusion during speech identification (Ziegler, Pech-Georgel, George, Alario, & Lorenzi, 2005).

Moreover, their phonological representations are argued to be holistic (Criddle & Durkin, 2001; Maillart, Schelstraete, & Hupet, 2004), or of lower quality (Claessen & Leitão, 2012). Although the causation and nature of these inaccurate representations are differently described (i.e. holistic, low quality), which is out of the scope of this paper to elaborate upon, these accounts have in common that the representations of children with DLD are argued to lack detailed phonological information (Claessen & Leitão, 2012; Criddle & Durkin, 2001; Maillart et al., 2004; Sutherland & Gillon, 2005).

Regarding the production modality, the speech of children with DLD is typically less accurate relative to TD children (e.g. Gerrits & de Bree, 2009). Although the severity of these problems differs per child, their incorrect phonological realisations can lead to unintelligible speech, which may cause problems in communication (Beers, 1995; Bishop & Snowling, 2004; Brooks & Kempe, 2012).

The production of weak syllables, consonant clusters, and word-final consonants have been reported to be problematic for children with DLD across different languages (e.g. Beers, 1992; Bortolini & Leonard, 2000; Gallon, Harris, & Van der Lely, 2007; Gerrits & de Bree, 2009; Gerrits, 2010; Orsolini, Sechi, Maronato, Bonvino, & Corcelli 2001). As aforementioned, these sound patterns are also commonly simplified by TD children. Analyses that focussed on simplification processes in the productions of children with DLD revealed that their speech is characterised by processes that are common for typical development, as well as processes that are uncommon for typical development (e.g. *backing; foto* 'photo' becomes *foko*, or *metathesis* (transposition or exchange of sounds); *fiets* 'bicycle' becomes *fiest*) (Beers, 1995; Leonard, 1996; Aguilar-Mediavilla, Sanz-Torrent, & Serra-Raventos, 2002).

It has been attempted by many researchers to describe the phonological development of children with DLD in terms of a delay or deviance compared to their TD peers (e.g. Gerrits & de Bree, 2009; Gerrits, 2010; Aguilar-Mediavilla et al., 2002; Zwitserlood, 2014). Quantitative differences (i.e. more phonological errors in the DLD group) between the productions of children with DLD and TD children have led researchers to the conclusion that the phonological development of children with DLD is delayed, whereas qualitative differences (i.e. errors that are atypical[2] for TD children) have been interpreted as a deviance in the phonological development of children with DLD (for a review, see Leonard, 2014).

2.3 Limitations in phonological short-term memory: An underlying deficit of DLD

2.3.1   A pSTM integrated speech processing model

2.3.1.1 Underlying principles of the pSTM integrated speech processing model

As aforementioned, a deficit in pSTM has been suggested to underlie DLD (e.g. Gathercole & Baddeley, 1990; Coady & Evans, 2008). For a better understanding of how this limitation can affect speech processing, and thus resulting in the errors that children with DLD make with repeating nonwords, Jacquemot and Scotts's (2006) pSTM integrated speech processing model, which is updated in this paper based on Baddeley and Hitch (2019), can be useful. This updated version of the model is depicted in Figure 1.

---

[2] Atypical patterns are patterns that occur infrequently in typical development, but it should be noted atypical patterns have also been reported in the speech of TD children (Beers, 1995; Leonard, 2014). Therefore, it is questionable whether these processes indeed indicate a deviance in the phonological development of children with DLD.
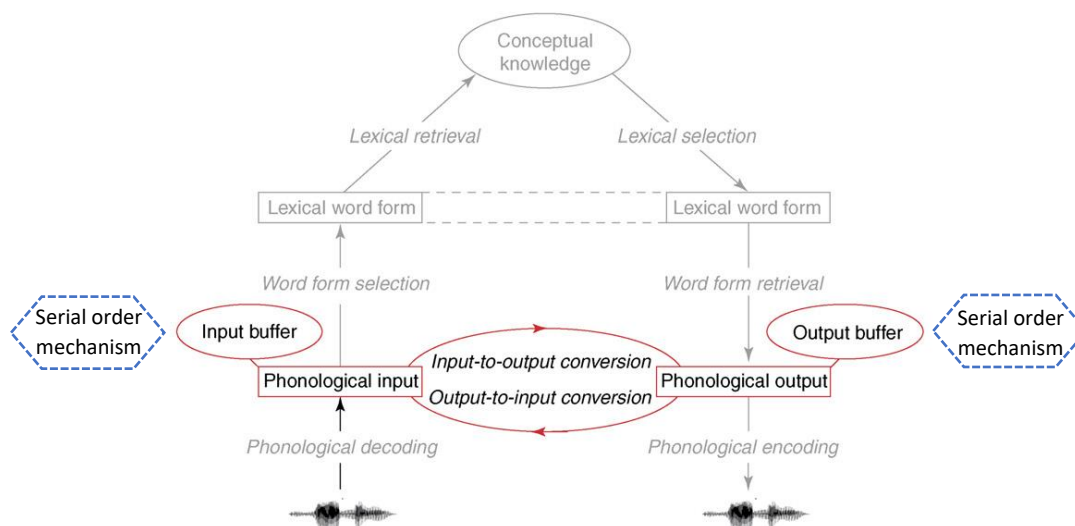
Figure 1. An adapted version of the pSTM integrated speech processing model by Jacquemot and Scott (2006, p. 481). The red part comprises pSTM. In blue the mechanism for the retention of serial order information (Attout et al., 2012; Baddeley & Hitch, 2019; Hurlestone & Hitch, 2014; Hurlestone et al., 2015) is added.

The model by Jacquemot and Scott (2006) is based on the pSTM model by Baddeley and colleagues (Baddeley & Hitch, 1974, as cited in Gathercole & Baddeley, 1990, Baddeley, 2003; Baddeley, Lewis, & Vallar, 1984). pSTM or the *phonological loop* is part of working memory, and has been proposed to be a mechanism that is responsible for the retention of verbal information. As shown in Figure 1, in the model by Jacquemot and Scott (2006) pSTM comprises three components: (1) the phonological input buffer, and (2) the phonological output buffer in which phonological representations can be temporarily stored and refreshed by rehearsal, and (3) the process of information conversion between these two buffers. Although it is subject to debate whether the input and output buffer are two separate buffers, this model adopts the approach that these are separate but interconnected in a way that information can be converted. This conversion of information between the two buffers allows for the repetition of nonwords, without the need to process its semantic meaning (Jacquemot & Scott, 2006).

In recent models of (p)STM the view is adopted that item and order information are distinctly stored (Attout, Van der Kaa, George, & Majerus, 2012; Hurlstone, Hitch, & Baddeley, 2014; Hurlstone & Hitch 2015). The original model of pSTM did not include a distinction

between item and order memory systems. However, as this original model could not explain "why items tend to be recalled in the wrong order when capacity is exceeded", the model has been updated by including such a distinction (Baddeley & Hitch, 2019, p. 94).

The code that is used for the storage of item information has not been clearly specified, but is has been described as an "articulatory" and "acoustic" code (Baddeley & Hitch, 2019, p. 94), or information comprising "phonological, lexical, and semantic characteristics" (Attout et al., 2012, p. 357). Generalising this to the phonemic level it can be argued that these verbal items are stored by means of their phonological characteristics.

The storage of order information refers to the serial order of items in a sequence (e.g. phonemes in a word) (Attout et al., 2012). The mechanism for the storage of order information seems to be connected to but separate from the language system, although controversy exists about how it exactly operates, which is out of the scope of the present paper to elaborate upon (cf. Attout et al., 2012; Hurlstone et al., 2014; Hurlstone & Hitch, 2015).

2.3.1.2. The process of repeating nonwords: Perception and production

In the case of repeating a nonword, speech processing starts witch perception (see the left part of the model in Figure 1). First, during phonological decoding the incoming nonword is segmented into smaller units (i.e. syllables, phonemes). Whilst the phonological representations (i.e. phonological input) are being constructed and manipulated, these representations are temporarily retained within the input buffer**.** Although not specified within the model, these representations can be speculated to contain item and order information. Item information is retained within the input buffer of pSTM by means of their phonological characteristics, whereas the serial order retention mechanism is responsible for the retention of the phonemes' order (Baddeley & Hitch, 2019).

Whereas the process would continue with lexical recognition if real words are presented, this does not hold for nonwords, since nonwords have no semantical meaning[3]. Instead, the phonological representation of the nonword, that contains item and order information, is converted from the input to the output buffer, and the process continues with the process of speech production (see the right part of the model in Figure 1).

In the output buffer the converted phonological representation is retrieved and encoded into a motor programme. Although Jacquemot and Scott (2006) do not specify how phonological representations are encoded, they argue that the output buffer is responsible for the storage of phonological elements that are strung together. Therefore, it can be assumed that an in-between step in phonological encoding is necessary prior to motor programming, presumably being as follows. During the encoding stage, the segments corresponding to the phonemes of the phonological representation are selected, as well as the metrical frame that contains information about the number of syllables and the stress pattern (Biran & Friedmann, 2004; Levelt, 1992). It can be speculated that at this stage order information is retrieved as well. Subsequently, the segments are integrated into the metrical frame in their correct serial order (i.e. corresponding to the order as specified in the phonological representation). Finally, the process continues by activating the motor programme, resulting in the production of the nonword.

2.3.2   pSTM and its relation to a nonword repetition deficit

The suggestion that a deficit in pSTM is one of the underlying deficits of DLD has received much support from research with nonword repetition tasks (NWRTs) as a measure of storage in pSTM (for a review, see Coady & Evans, 2008). It has consistently been shown that children

---

[3] Note that the suggestion that lexical recognition does not play a role during nonword repetition should be nuanced, as lexical knowledge seems to influence nonword repetition to some extent (Graf Estes et al., 2007; Burke & Coady, 2015; Coady & Evans, 2008; Riches, Loucas, Baird, Charman, & Simonoff, 2011).

with DLD show more difficulties with NWRTs compared to TD children (Graf Estes et al., 2007). Repeating a nonword involves multiple processing stages: A new phonological representation has to be created during phonological decoding, this representation needs to be temporarily stored in pSTM, the representation has to be retrieved, and finally it has to be sequenced during phonological encoding (e.g. Coady & Evans, 2008; Rispens & Parigger, 2010; Schraeyen, 2018). Disruption at all of these levels can lead to errors during the repetition of a nonword. For children with DLD the suggested pSTM deficit presumably affects the maintenance of phonological representations in the phonological buffer, which could lead to problems at different stages in the speech perception and/or production process.

It should be noted though that factors as wordlikeness and phonotactic probability influence nonword repetition performance (e.g. Archibald & Gathercole, 2006; Munson, Kurtz, & Windsor, 2005; Coady, Evans, & Kluender, 2010), indicating that lexical knowledge in long-term memory plays a role to some extent as well (Graf Estes et al., 2007; Burke & Coady, 2015; Coady & Evans, 2008; Riches et al., 2011). Although the NWRT is not a pure measure of one phonological process, it is widely accepted as a measure of pSTM capacity (Coady & Evans, 2008).

That errors in nonword repetition are caused by limitations in pSTM is evidenced by the robustly found *length effect* (e.g. Bishop, North, & Donlan, 1996; Boerma et al., 2015; Briscoe Bishop, Norbury, 2001; Dollaghan & Campbell, 1998; Gathercole & Baddeley, 1990; Graf Estes et al., 2007; Marton & Schwartz, 2003). This effect holds that with increasing length higher demands are placed on pSTM capacity, which results in a decrease of performance accuracy in nonword repetition. Relative to their TD peers, children with DLD perform more poorly on longer nonwords (i.e. three syllables or more) compared to shorter nonwords (i.e. one or two syllables) (e.g. Dollaghan & Campbell, 1998; Graf Estes et al., 2007; Marton & Schwartz, 2003). This implies that the pSTM capacities of children with DLD are "more readily

overwhelmed by long nonwords", indicating that pSTM appears to be limited in children with DLD (Graf Estes et al., 2007, p. 183).

2.3.2.1. Development of pSTM and NWRT performance

Given that the pSTM abilities of children with TD and DLD are still developing, it is not surprising that they do not perform flawless on NWRTs. The capacity of pSTM of TD children increases linearly from 4 years of age to early adolescence, and levels off at around 14-15 years of age (Gathercole, Pickering, Ambridge, & Wearing, 2004).

For children with DLD the developmental trajectory of pSTM has not been examined longitudinally across the same time span as it has been for children with TD. However, some studies have addressed pSTM development for children with DLD, showing that between 3 and 4 years of age their pSTM abilities improve, and between 4 and 6 years this ability appears to remain stable (Gray, 2006). From 7 to 13 years of age (i.e. Grade 2 to Grade 8[4]) their pSTM also improves (Catts, Adlof, Hogan, & Weismer, 2005). Despite improvement, these studies together showed that across the age from 3 to 13, children with DLD showed poorer NWRT performance compared to their TD peers. This indicates that the DLD group has weaker pSTM abilities relative to their TD peers. The pace at which pSTM develops has been demonstrated to be similar for children with DLD compared to their TD peers (for a meta-analysis, see Graf Estes et al., 2007)

Whereas TD children reach adult-like levels of pSTM by early adolescence, this does not hold for children with DLD: A comparison between nonword repetition accuracy of children with DLD at the age of 11, and the age of 14 revealed that their pSTM abilities did not improve during this time period (Conti-Ramsden & Durkin, 2007). In fact, "the mean scores at both time points were equivalent to what would be expected of an average 6-year-old" (Conti-

---

[4] Although no mean ages were specified, it can be assumed that children in Grade 2 and Grade 8 are approximately 7 and 13 years of age respectively.

Ramsden & Durkin, 2007, p. 149). Hence, there was no evidence for a catch-up in pSTM capacity. However, when the whole group of children with DLD was divided into groups with and without reading problems, this revealed for both ages that the former group was outperformed by the latter. This suggests that the development of nonword repetition abilities might be influenced by literacy skills, which has been supported by other research (Bishop, McDonald, Bird, & Hayiou-Thomas, 2009; Catts et al., 2005; Rispens & Parigger, 2010; but see Vandewalle, Boets, Ghesquière, and Zink, 2012).

2.3.2.2. Difficulties with the retention of item, order, or both

Previous research has shown that children with DLD are outperformed on nonword repetition by their TD peers, but less is known about the level at which the retention problems of children with DLD are situated. For instance, in order to correctly repeat a nonword the phonological items and their serial order should be retained in pSTM.

To the best of my knowledge error patterns in nonword repetition as categorised by problems with retaining item (i.e. the phonemes' phonological characteristics) and/or order information (i.e. the phonemes' position) have not been analysed for children with DLD. Noteworthy, Schraeyen (2018) has recently investigated this issue for children with dyslexia, a disorder that frequently co-occurs with DLD (Bishop & Snowling, 2004). Relative to TD children the dyslectic group appeared to experience relatively more difficulties with the retention of serial order, while no significant differences were found regarding the retention of item information.

Regarding the ability of children with DLD to retain item information, previous research provides conflicting evidence. For instance, the phonological representations of children with DLD are demonstrated to lack phonemic detail (Claessen & Leitão, 2012; Criddle & Durkin, 2001; Maillart, et al., 2004). These studies led to this conclusion based on the finding that

relative to TD children, children with DLD were poorer in detecting phonemic changes when they were presented pairs of (non)words that differed for example in terms of substituted, omitted, or added phonemes. Hence, these results imply difficulties at the level of phonemic item information.

However, the view that children with DLD experience problems with the retention of item information has been challenged. Majerus et al. (2009) conducted a recognition experiment to assess phonemic item retention in pSTM. French-speaking children with and without DLD were presented short lists of nonwords, followed by separate probe nonwords that contained either the same phonemes as the target nonword, or that differed from the target nonword by one phoneme. The participants had to judge whether the probe nonword had been presented in the list or not. No significant differences were found between the two groups, and hence the suggestion that children with DLD are impaired in their retention of phonological items in pSTM was not supported by these results.

Children with DLD might also experience problems with the retention of order information, which is partially supported by previous research. For example, when sequences consisting of four words (Hsu & Bishop, 2014), or sequences containing 3 to 7 digits (Majerus et al., 2009) are auditorily presented, children with DLD are demonstrated to be outperformed by their TD peers to overtly recall verbal items in their correct position. This implies that the pSTM abilities of children with DLD to retain serial order information is limited.

Other support for retention problems with serial order may be derived from a recent study by de Almeida, Ferré, Barthez, and Dos Santos (2019). French children with and without DLD participated in a NWRT that included nonwords made up of consonant clusters. Patterns of error types (i.e. metathesis: transposition or exchange of sounds, substitution, omission) were analysed and appeared not to differ significantly between the two groups. However, the groups did differ in quantity (i.e. the DLD group made more errors). Most errors were caused by

metathesis, which is caused by retention loss of order information, but preservation of item information (e.g. exchange: *sugar* produced as *surag*, transposition: *sugar* produced as *surga*). Therefore, this error type might be speculated to reflect retention problems with serial order. The authors did not interpret the results as such, but they argued that this error type was made to create phonologically less complex syllable structures. Further research is needed to examine whether metathesis will also frequently occur in phonological less complex syllable structures. If this would turn out to be the case, this would imply that the order errors were in fact caused by difficulties with the retention of serial order, rather than by difficulties with phonological complexity.

In contrast to de Almeida et al. (2019) other research found that errors in phoneme order were infrequent for children with TD and DLD (i.e. 6% and 7% respectively) within the set of observed error types (i.e. substitutions, cluster simplifications, omissions, additions, phoneme order errors) (Marton & Schwartz, 2003). This implies that children with DLD experienced less difficulties with the retention of order information relative to item information, similar to TD children.

Taken together, it is unclear whether phonological errors that children with DLD produce are mainly characterised by difficulties with the retention of item, order, or both types of information. The present paper will further explore this issue.

3. The present study

The nonword repetition accuracy of children with DLD has exhaustively been examined, revealing that these children are outperformed by their TD peers (e.g. Graf Estes et al., 2007). However, as far as I am aware no studies have investigated the types of errors that are made during NWRTs by focussing on problems with the retention of phonemic items (i.e. the phonemes' phonological characteristics) and/or their serial order (i.e. the phonemes' position

in the nonword). For example, to repeat the nonword *kuimup*, each individual item should be retained (i.e. each individual item /k/, /u/, /i/, /m/, /u/, /p/), as well as each item's position (e.g. /k/ is in initial position, whereas /p/ is in final position).

Insight in the error profiles of children with and without DLD contributes to a better understanding of the level at which problems with nonword repetition are situated (i.e. retention loss of item, order, or both), which is useful for the development of interventions. Additionally, if differences between the error patterns of the two groups emerge, this could be used as a clinical marker to strengthen the diagnosis of DLD. Insight in the differences between the two groups is also theoretically of interest, as it reveals whether children with DLD differ from their TD peers by means of a delay or a deviance.

The main purpose of the present paper is to address the characterisation of error patterns of children with and without DLD. Productions of consonants are focussed on, as children with DLD and TD have been demonstrated to make relatively more consonant than vowel errors when repeating nonwords (Girbau & Schwartz, 2007, 2008, but see Burke & Coady, 2015). First, performance accuracy on the NWRT will be investigated. The effect of length on accuracy will be considered, as children with DLD have, relative to their TD peers, greater difficulties with lengthier nonwords (i.e. three, four, five syllables) than with shorter ones (i.e. two syllables) (e.g. Dollaghan & Campbell, 1998; Graf Estes et al., 2007; Marton & Schwartz, 2003). Additionally, the effect of increasing age (i.e. 72 to 95 months of age) on accuracy will be examined, as only limited studies have examined the nonword repetition development of children with DLD relative to TD children. However, studies that did examine the development of nonword repetition demonstrated that NWRT accuracy, and thus pSTM capacity, improves with age for children with and without DLD (Bishop et al., 2009; Catts et al., 2005; Gathercole et al., 2004; Gray, 2006). Moreover, it will be explored how the error profiles of children with and without DLD are characterised by means of proportions of item, order, and combined (i.e.

combination of item and order) errors, and how these patterns are related to syllable length and age. These research goals are captured by the following questions:

1. Do children with DLD differ from children with TD in their accuracy at repeating nonwords, and if so, which factors moderate this between-group difference?
   a. To what extent is the accuracy difference between children with DLD and children with TD related to nonword length?
   b. To what extent is the accuracy difference between children with DLD and children with TD dependent on age?
2. Do children with DLD differ in the types of errors they make compared to TD children, and if so, which factors moderate this between-group difference?
   a. To what extent is the difference between children with DLD and children with TD related to nonword length?
   b. To what extent is the difference between children with DLD and children with TD dependent on age?

To answer the first research question, the groups will be compared on their accuracy scores. It is expected that the robust finding that children with DLD perform less accurately on NWRTs than their TD peers will be replicated (Graf Estes et al., 2007). In line with previous research it is predicted that the difference in performance accuracy between the two groups will be larger for longer nonwords (i.e. three, four, five syllables) compared to shorter ones (two syllables). This effect of length is expected, as pSTM of children with DLD is more readily exceeded by longer nonwords than it is for TD children, resulting in a larger performance decrease for the DLD group relative to the TD group (e.g. Dollaghan & Campbell, 1998; Graf Estes et al., 2007; Marton & Schwartz, 2003). Regarding the effect of age, it is expected that the NWRT accuracy

of all children will improve from 72 to 95 months of age, as their pSTM abilities are still developing (Bishop et al., 2009; Catts et al., 2005; Gathercole et al., 2004; Gray, 2006). It is possible that the two groups improve at an equal pace (Graf Estes et al., 2007), such that the gap in performance accuracy between the groups remains the same. Alternatively, the gap may become larger. The ability to repeat nonwords is facilitated by reading skills (Bishop et al., 2009; Catts, 2005, Conti-Ramsden & Durkin, 2007; Rispens & Parigger, 2010), and because many children with DLD develop reading problems (e.g. Bishop & Snowling, 2004), their improvement may be less strong than the improvement of their TD peers.

To answer the second question, the proportions of item, order, and combined errors of children with DLD relative to children with TD will be investigated. Previous research is not conclusive about whether children with DLD experience more problems with the retention of item information or order information. Therefore, only tentative predictions were formulated.

Order errors might intuitively be speculated to be 'milder' compared to the other error types, as this error type reflects only retention loss of order information while item information is preserved. Although the same reasoning (i.e. retention loss of item information while order information is preserved) might be applied to the occurrence of item errors, this does not hold. For example, in cases that all consonants of a nonword were produced incorrectly, such that the nonword was scored as item error, it can be argued that item nor order information was preserved. So, item errors may not reflect retention loss of purely item information. Last, combined errors reflect retention loss of item and order information. So, order errors are the only type based on which it can be concluded with certainty that at least one type of information has been preserved. Following this line of reasoning, it may tentatively be predicted that the proportion of order errors within the set of incorrect repetitions will be lower for the DLD group than for the TD group.

With increasing syllable length, and thus larger processing demands, it is intuitively predicted that the amount of item and order errors within each nonword will increase, leading to a higher proportion of combined errors relative to the other error types for both groups. With increasing age it may be assumed that pSTM improves, such that it is tentatively predicted for both groups that within the set of incorrect repetitions the proportion of 'milder' order errors will become higher relative to the other error types.

4. Method

4.1. Participants

In total 78 children participated, including 39 children with DLD, and 39 children with TD. Children were tested during a time period of three years with a frequency of once a year, starting at 5 and 6 years of age (Research project CoDEmBi, see Boerma et al., 2015). Children with TD were recruited via Dutch regular elementary schools. Children with DLD were recruited via two national organisations (Royal Dutch Kentalis and Royal Auris Group) that provide educational facilities for children with DLD. Children with DLD were diagnosed by licensed professionals based on a standardised procedure (Stichting Siméa, 2014).

Children who were included in the present study were selected based on the following selection criteria: (1) The child was monolingual, (2) the child participated in the first and last wave (henceforth T1 and T2), (3) the child did not give too many non-responses[5] either at T1 or T2 (n ≤ 3), or in total (n ≤ 6), and (4) the child was not diagnosed with a disorder, or with an additional disorder in the case of children with DLD. Children with DLD were matched to their TD peers as close as possible on age in months, nonverbal IQ, and gender. The group characteristics are displayed in Table 1.

---

[5] One child from the DLD group was included who gave more non-responses (i.e. n = 6 at T1, n = 2 at T2) than the maximum amount that was allowed as defined in the selection criteria. It was decided to include this child, because matching to a child from the TD group was only possible by including this child.

The groups did not differ in nonverbal IQ ($t(76) = .264$, $p = .793$), gender ($\chi^2(1) = 2.006$, $p = .157$), age at T1 ($t(76) = .543$, $p = .589$), and age at T2 ($t(76) = .361$, $p = .719$). Socioeconomic status (SES) significantly differed between the two groups ($t(75) = 2.353$, $p = .021$), with higher SES in the TD group compared to the DLD group. Nonverbal IQ scores were measured using the short version of the Wechsler Nonverbal–NL (Wechsler & Naglieri, 2008). SES scores were measured based on the education level of both parents.

Table 1. Participant characteristics

Group	N	Age in months T1 M (SD)	Range	Age in months T2 M (SD)	Range	Nonverbal IQ M (SD)	Range	Socioeconomic status (SES) M (SD)	Range	Gender $N_{GIRLS/BOYS}$ (%)
TD	39	71.1 (7.2)	59-90	94.3 (7.2)	81-112	105.0 (14.3)	81-128	6.58 (2.0)	2.0-9.0	22/17 (56%/44%)
DLD	39[a]	72.0 (6.5)	59-83	94.9 (6.6)	81-107	104.1 (14.0)	80-131	5.53 (2.0)	2.0-9.0	28/11 (72%/28%)

[a] Information about the SES of the parents of one child with DLD was not available.

4.2. Materials and procedures

The present study was an analysis of a previously collected longitudinal dataset for the Research project CoDEmBi, which was approved by The Standing Ethical Assessment Committee of the Faculty of Social and Behavioral Sciences at Utrecht University. Parental informed consent was given for all children who participated. For the present study the NWRT data of the first and last wave of the longitudinal dataset (in the present study T1 and T2) were included of a subset of children who met the selection criteria as set out in the previous section.

The NWRT that was used was an adapted version of the NWRT which was originally developed by Rispens and Baker (2012) (see Boerma et al., 2015). The task comprised 24 nonwords that complied with the prosodic rules of the Dutch language. The set of nonwords did not include consonant clusters, and all nonwords were made up of CV...CVC sequences.

Nonwords were equally distributed over syllable lengths of two, three, four, and five syllables, and over high versus low phonotactic probability, based on the Dutch phonotactic frequency database (Adriaans, 2006, as cited in Rispens & Baker, 2012). Thus, the total set included 12 nonwords of high, and 12 nonwords of low phonotactic probability, with each subset containing 3 nonwords for each syllable length (i.e. two to five syllables). The set of nonwords is presented in Appendix A.

The nonwords were pre-recorded by a female native speaker of Dutch. The set of nonwords was auditorily presented to the child by an alien (cartoon) who wanted to teach the child a 'foreign language'. Each nonword was presented only once. The child was asked to repeat each nonword immediately after hearing it. All children were tested individually at their schools in a quiet room, and they were allowed to withdraw from the experiment at any time. The nonword repetitions were recorded, and subsequently transcribed offline. All data were anonymised.

4.3. Scoring

To examine NWRT performance accuracy of children with and without DLD, all repeated nonwords were coded on word level as either produced correctly or incorrectly. Nonresponses (2.2%) were removed from the dataset. Nonword repetitions were scored as correct if all consonants were produced correctly, and in the correct position. Nonword repetitions in which consonants were substituted or omitted were scored as incorrect. Additions were scored as correct because these added segments do not reflect a loss of information regarding the target consonants (Dollaghan & Campbell, 1998). Voicing and devoicing of the consonants /s, z/ or /f, v/ was scored as correct, since limitations in transcription reliability might have affected perception of voicing. False starts were ignored if they were followed by (partially) correct responses (e.g. *baamen-baameriejooves* for the target nonword *baarmerienooves*).

After the nonwords were coded (i.e. correct, or incorrect), accuracy percentages were calculated. For each child an accuracy percentage was calculated on word level by dividing the amount of correctly produced nonwords by the total amount of nonwords he/she produced.

A second analysis was done within the set of incorrectly repeated nonwords, in order to explore whether repetition errors were mainly characterised by a retention loss of item, order, or both types of information. Each produced consonant in the nonword was matched on the target consonant and target position. In most cases the productions and targets could easily be matched, clearly revealing whether item and/or order information was preserved or not. Thereafter, the nonwords were coded on word level. If at least one consonant was substituted and/or omitted, while the consonants' positions were preserved, erroneous nonwords were coded as phonemic item error (PI). Erroneous nonwords were coded as serial order error (SO) if at least one correctly retained consonant had migrated to another position in the nonword (i.e. metathesis: exchange or transposition). Only correctly retained phonemes were taken into account when scoring for order, because scoring every phonemic item error as a serial order error as well would result in the item errors type being a subset of the set comprising serial order errors. The last response type was a combined error (C). Erroneous repetitions were coded as such if item and order errors were both made in one nonword repetition. Nonwords that were coded as item error or order error were not included in the set of combined errors, and vice versa. Examples of the error types are provided in Table 2. Detailed procedures that were followed in the few cases that consonants could not be matched straightforwardly are presented in Appendix B.

After coding the erroneous nonwords, error percentages of item, order, and combined errors were calculated within the set of incorrectly repeated items. For each child the proportion of each error type was calculated separately on word level by dividing the amount of each error type by the total number of incorrect nonword repetitions he/she produced.

Table 2. Scoring procedures for types of errors.

Type of error	Example (Target > production)
Phonemic item error (PI)	
Consonant omission	joe.feu**m** > joe.feu
Consonant substitution	**m**ui.hu.guuf > **n**ui.hu.guuf
Consonant substitution and omission	raa.**nom** > raa.**la**
Serial order error (SO)	
Consonant exchange	voo.**p**ee.**k**et > voo.**k**ee.**p**et
Consonant transposition	nuijigeufuusu**t** > nuijugeufuu**t**su
Combined error (C)	
Consonant exchange/transposition and consonant substitution	dee.**v**oe.**n**os > dee.**m**oe.**v**os
Consonant exchange/transposition and consonant omission	veu**j**oetu**p** > fee**p**iete
Consonant exchange/transposition and consonant substitution and consonant omission	meu**f**uu**s**ui**n**ef > meu**s**uu**f**aa**m**e

4.4. Data Analysis

All statistical analyses were performed using SPSS 25 (IBM Corp., 2017), and principles by Field (2015) were followed. Outliers were not excluded from the analyses as these seemed to be valid datapoints. An exploration of the data demonstrated that a part of the data was not normally distributed, as indicated by skewness and kurtosis values smaller than -1 and greater than 1. Transforming the data did not result in a normal distribution. However, if sample sizes of groups are equal, and if sample sizes are big enough (i.e. > 30), analysis of variance (ANOVA) is robust to violations of normality (Field, 2015). As the sample sizes of the present study were equal and big enough (i.e. $N = 39$ for each group), it was assumed that the analyses that were performed were robust to violations of normality.

To answer the first research question about performance accuracy, a 4 x 2 x 2 mixed analysis of covariance (ANCOVA) was run with Syllable Length as a within-subjects factor with four levels (i.e. two, three, four, and five syllables), Time as a within-subjects factor with

two levels (i.e. T1 and T2), and Group as a between-subjects factor with two levels (i.e. TD and DLD). Nonverbal IQ and Age were added as covariates since these factors were found to correlate with performance accuracy. The included covariates were centred around the mean across all subjects before they were included, following Schneider, Avivi-Reich, and Mozuraitis (2015).

To answer the second research question about types of errors, two separate analyses were done. First, the effect of Group and Time on Error Type were analysed using a 3 x 2 x 2 mixed ANCOVA, with Error Type as a within-subjects factor with three levels (i.e. Item, Order, Combined), Time as a within-subjects factor with two levels (i.e. T1 and T2), and Group as a between-subjects factor with two levels (i.e. TD and DLD). Nonverbal IQ was added as a covariate as this factor was found to correlate with the proportion of item errors. The covariate was centred around the mean across all subjects. Significant interactions between factors that were of interest for the purpose of this study (i.e. interaction effects involving Group) were further unpacked by post hoc analyses (i.e. one-way ANCOVAs and repeated-measures ANCOVAs). Bonferroni correction was applied. Effect sizes were calculated using Partial Eta Squared (η_p^2).

To examine the effect of syllable length on error types, a descriptive analysis was done. It was decided not to test this effect by running an ANCOVA because for certain syllable lengths some participants repeated all nonwords correctly, such that in these cases no percentages could be calculated for each error type. This data would be handled as missing data by the mixed ANCOVA, such that these subjects (TD, T1: $n = 10$, T2: $n = 10$; DLD, T1: $n = 2$, T2: $n = 1$) would be excluded listwise when running the test. Consequently, due to the non-equality of samples and smaller sample sizes running the ANCOVA would be invalid.

5. Results

5.1. Performance accuracy

Table 3 shows the mean percentages and standard deviations of the performance accuracy on the NWRT of children with and without DLD across different syllable lengths at T1 and T2.

Table 3. Mean percentages of correct nonword repetitions on the NWRT for children with TD and DLD at T1 and T2.

Time point	Syllable Length	TD			DLD		
		N	M (SD)	Range	N	M (SD)	Range
1	All	39	52.9 (15.6)	25.0-83.2	39	25.7 (14.9)	4.3-63.6
	2		65.8 (15.3)	33.3-100.0		50.3 (21.7)	0.0-100.0
	3		71.0 (22.3)	16.7-100.0		33.3 (26.5)	0.0-100.0
	4		44.1 (32.0)	0.0-100.0		12.8 (19.7)	0.0-83.3
	5		29.9 (25.2)	0.0-100.0		5.1 (11.0)	0.0-50.0
2	All	39	60.6 (16.3)	33.3-91.7	39	35.3 (14.9)	12.5-70.8
	2		70.0 (18.3)	33.3-100.0		53.8 (18.9)	16.7-83.3
	3		75.9 (16.7)	33.3-100.0		49.6 (28.0)	0.0-100.0
	4		57.4 (28.2)	0.0-100.0		25.0 (21.2)	0.0-83.3
	5		38.5 (28.9)	0.0-100.0		12.1 (15.6)	0.0-66.7

NOTE. Percentages of individual participants were rounded before the mean accuracy for all participants was calculated. Therefore, percentages may not correspond perfectly to the percentage when calculated by dividing the number of correct repetitions by the total number of items.

Results revealed a significant main effect of Syllable Length ($F(2.91, 215.12) = 110.68$, $p < .001$, $\eta_p^2 = .60$), Time ($F(1, 74) = 40.12$, $p < .001$, $\eta_p^2 = .37$), and Group ($F(1, 74) = 77.03$, $p < .001$, $\eta_p^2 = .51$). Age ($F(1, 74) = 4.89$, $p = .030$, $\eta_p^2 = .062$) and Nonverbal IQ ($F(1, 74) = 7.99$, $p = .006$, $\eta_p^2 = .10$) were significant covariates. Significant interaction effects were found for Syllable Length $\times$ Group ($F(2.91, 498.93) = 4.74$, $p = .004$, $\eta_p^2 = .060$), and of the covariate Age $\times$ Time ($F(1, 74) = 6.72$, $p = .011$, $\eta_p^2 = .08$). Other two-way and three-way interactions were not significant.

Pairwise comparisons showed that independent of Group, or Syllable Length, both groups performed better at T2 than T1 ($p < .001$). Moreover, independent of Syllable Length or

Time, the DLD-group was outperformed by the TD-group ($p < .001$). Last, independent of Group and Time an effect of Syllable Length was found. From a nonword length of three syllables onwards performance accuracy significantly decreased as syllable length increased (all ps $< .001$).

The interaction between Syllable Length and Group indicated that Syllable Length affected accuracy scores differently across the two groups. This interaction was further analysed by post hoc analyses. A one-way ANCOVA showed that the TD group significantly outperformed the DLD group on all syllable lengths: Two syllables ($F(1, 74) = 23.73$, $p < .001$, $\eta_p^2 = .24$), three syllables ($F(1, 74) = 50.5$, $p < .001$, $\eta_p^2 = .41$), four syllables ($F(1, 74) = 40.89$, $p < .001$, $\eta_p^2 = .36$), and five syllables ($F(1, 74) = 38.12$, $p < .001$, $\eta_p^2 = .34$). Effect sizes showed that the difference between the TD and DLD groups was larger for longer nonwords (i.e. three, four, and five syllables) relative to shorter nonwords (i.e. two syllables), with the largest effect for 3-syllable nonwords.

Furthermore, two separate repeated-measures ANCOVAs demonstrated a significant main effect of Syllable Length for both groups separately, and the effect of Syllable Length was found to be greater for children with DLD than for children with TD: DLD group ($F(2.66, 95.92) = 77.44$, $p < .001$, $\eta_p^2 = .68$), TD group ($F(2.72, 97.76) = 42.56$, $p < .001$, $\eta_p^2 = .54$). Pairwise comparisons showed that the effect of Syllable Length differed between the two groups as follows. TD children did not perform significantly different on 2-syllable nonwords compared to 3-syllable nonwords ($p = .465$), but their performance significantly decreased between the 3- and 4-syllable nonwords ($p < .001$), and 4- and 5-syllable nonwords ($p < .001$). The performance of the DLD group did also not significantly differ between 2- and 3-syllable nonwords, although this difference was marginally significant ($p = .052$). For the other syllable lengths their performance decreased as syllable length increased, as indicated by significant

performance differences between 3- and 4-syllable nonwords ($p < .001$), and 4- and 5-syllable nonwords ($p < .001$).

5.2. Types of errors

5.2.1. Overall error patterns

Table 4 presents the mean percentages and standard deviations of each error type (i.e. Item, Order, Combined) within the set of incorrect nonword repetitions, for the two groups at T1 and T2.

Table 4. Mean percentages of error types on the NWRT for children with DLD and TD at T1 and T2.

Time point	N	Error Type	TD M (SD)	Range	DLD M (SD)	Range
1	78	Item	71.1 (18.5)	36.4-100.0	68.6 (13.4)	36.8-100.0
		Order	6.3 (7.7)	0.0-25.0	3.6 (5.2)	0.0-27.3
		Combined	22.6 (15.6)	0.0-54.5	27.9 (11.3)	0.0-52.6
2	78	Item	70.1 (15.7)	46.2-100.0	62.2 (13.7)	33.3-100.0
		Order	11.4 (10.3)	0.0-28.6	7.2 (7.4)	0.0-27.3
		Combined	18.5 (16.9)	0.0-53.8	30.6 (13.0)	0.0-61.1

NOTE. Percentages of individual participants were rounded before the mean accuracy for all participants was calculated. Therefore, percentages may not correspond perfectly to the percentage when calculated by dividing the number of occurrences per error type by the total number of incorrect repetitions.

Results revealed a significant main effect of Error Type, ($F(1.42, 105.03) = 505.42$, $p < .001$, $\eta_p^2 = .87$). The effects of Time ($F(1, 74) = .04$, $p = .838$), and Group ($F(1, 74) = .09$, $p = .760$) were not significant. Nonverbal IQ was not a significant covariate. A significant interaction effect was found for Error Type × Group, ($F(1.42, 105.03) = 7.97$, $p = .002$, $\eta_p^2 = .10$). Other two-way and three-way interactions were not significant.

Pairwise comparisons showed that for all children, independent of Group, or Time, the proportion of Item errors was significantly higher than the proportion of Order errors ($p < .001$), the proportion of Item errors was significantly higher than the proportion of Combined errors

(p < .001), and the proportion of Combined errors was significantly higher than the proportion of Order errors (p < .001).

The interaction between Error Type and Group indicated that the three error types were distributed differently across the two groups. This interaction was further analysed by post hoc analyses. A one-way ANCOVA showed that the proportion of Order errors was significantly lower for the DLD group compared to the TD group ($F(1, 74) = 6.94$, $p = .016$, $\eta_p^2 = .09$). In contrast, the proportion of Combined errors was significantly higher for the DLD group compared to the TD group ($F(1, 74) = 13.22$, $p = .001$, $\eta_p^2 = .15$). The proportion of Item errors was lower for the DLD group compared to the TD group, but this difference was only close to significant ($F(1, 74) = 3.86$, $p = .053$, $\eta_p^2 = .05$). Effect sizes showed that the difference between children with and without DLD was small for Item errors, medium for Order errors, and large for Combined errors.

Furthermore, two repeated-measures ANCOVAs showed a significant main effect of Error Type for both groups separately: DLD-group ($F(1.36, 48.89) = 286.73$, $p < .001$, $\eta_p^2 = .89$), TD-group ($F(1.45, 52.06) = 228.18$, $p < .001$, $\eta_p^2 = .86$). Pairwise comparisons showed that for both groups the proportion of Item errors was significantly higher than Order errors and Combined errors, and the proportion of Combined errors was significantly higher than the proportion of Order errors (all ps < .001).

5.2.2.   Error patterns across different syllable lengths

A separate descriptive analysis was done to examine the effect of Syllable Length on Error Type. Figure 2 visualises the mean proportions of each error type across different syllable lengths for both groups at T1 and T2. Exact means and standard deviations are presented in Appendix C. For clarity the data of both time points is presented, but the effects of Time will

not be elaborated upon, since the analysis that did not consider Syllable Length showed that the proportions of error types did not significantly change over time.
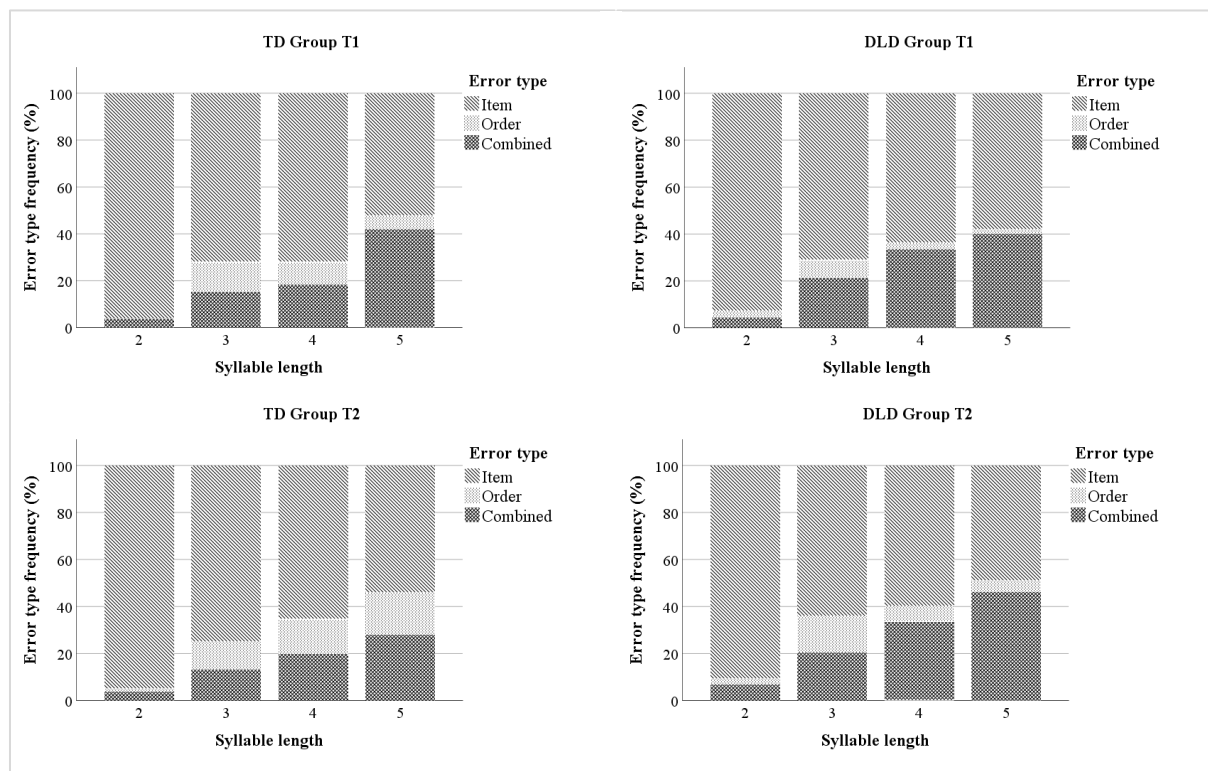
Figure 2. Proportions of error types on the NWRT across syllable lengths for children with TD and DLD at T1 and T2.

For each syllable length the overall pattern of types of errors was largely the same for both groups at both time points: The proportion of Item errors was largest, followed by the proportion of Combined errors. Item errors comprised the smallest class of error types. The proportions seemed to change with increasing syllable length for both groups at both time points. With increasing syllable length the proportion of Combined errors increased, whereas the proportion of Item errors decreased. At first glance the proportion of Order seemed to fluctuate slightly.

6. Discussion

6.1. Interpretation of the results

The main goal of the present study was to provide more insight in the widely reported poorer nonword repetition of children with DLD compared to their TD peers, and to investigate how this performance develops from 72 to 95 months of age. The effect of increasing syllable length was furthermore considered. The performance accuracy of children with and without DLD was examined, and their error profiles were explored by means of proportions of item, order, and combined errors. Insight in the error profiles reveals the kind of retention problems that children with DLD experience with repeating nonwords, and whether these problems differ from the problems of their TD peers. This insight sheds some light on the question whether children with DLD differ from children with TD by a delay or a deviance in their pSTM abilities related to nonword repetition

6.1.1. Performance accuracy

With respect to nonword repetition accuracy, previous findings were replicated that children with DLD are outperformed by TD children (e.g. Graf Estes et al., 2007). As predicted, performance accuracy was related to syllable length for all children. Specifically, for both groups, and both ages their performance significantly deteriorated between three and four, and four and five syllables. However, for children with DLD the performance decrease was marginally significant between two and three syllables, which may suggest a relatively earlier performance decrease for this group, corresponding with the findings of Marton and Schwartz (2003) for children ranging from 7 to 10 years of age. For children with DLD and children with TD they found a significant decline at 3- or 4-syllable nonwords respectively. Contrary to the nonwords used in the present study, some of the nonwords they used contained consonants clusters. It was not specified whether consonant clusters were equally distributed across

different syllable lengths. If certain syllable lengths would have contained more clusters than other syllable lengths, this might have influenced the results. The difference between the syllable structure of the nonwords that were used in the present study, and the set of Marton and Schwartz (2003) may explain why they found an accuracy decline for children with DLD at 3-syllable nonwords, whereas this decline was only found to be marginally significant in the present study.

Contrary to the prediction that the accuracy difference between the DLD and TD group would increase with length, this difference was found to be largest for three syllables, rather than for five syllables. It may tentatively be speculated that this can be explained by the trend that the performance accuracy of the DLD group declined at three syllables, whereas the TD group showed an accuracy decrease at four syllables. Regardless of this unexpected finding, the difference in accuracy between the groups was larger for longer nonwords (i.e. three, four, and five syllables) relative to short ones (i.e. two syllables). This finding is in agreement with previous research (e.g. Dollaghan & Campbell, 1998; Graf Estes et al., 2007; Marton & Schwartz, 2003). For longer nonwords pSTM seems to be more readily exceeded for children with DLD relative to TD children. This length effect indicates that the pSTM storage capacities of children with DLD are weaker compared to TD children.

At both ages the DLD group was outperformed by the TD group, and the observed effect of age was in line with the predicted pattern. At 95 months of age (T2) all children performed significantly better compared to when they were 72 months (T1), indicating a growth in pSTM. As no significant interaction was found between time (i.e. age) and group, this indicates that the groups did not differ in improvement rate, which supports the prior finding that children with DLD appear to improve at a similar pace compared to TD children (Graf Estes et al., 2007).

6.1.2. Types of errors

Regarding types of errors, results demonstrated that independent of age for both groups the proportion of item errors was largest, followed by combined errors, which was in turn followed by order errors. Relative to the TD group, the proportion of combined errors was higher, and the proportion of order errors was lower for the DLD group. The proportion of item errors was also lower for the DLD group, although this difference was small and only close to significant.

Nonwords that were coded as combined errors were not included in the set of individual item and order errors. Consequently, a higher proportion of combined errors results in a lower proportion of individual item and order errors, such that the proportions of combined and order/item errors balance out. Hence, the relative smaller proportion of order errors of the DLD group relative to the TD group does not imply that the former group did not experience difficulties with the retention of order information. In fact, in this group order errors were made relatively more often in combination with an item error, resulting in a higher proportion of combined errors, which has also been observed for children with dyslexia (Schraeyen, 2018). This observation is in line with the tentative prediction that order errors are relatively mild.

Although order errors were relatively more often made in combination with an item error by children with DLD compared to their TD peers, the overall error pattern was largely the same for both groups (i.e. item > combined > order). As the DLD group showed poorer performance accuracy, this implies that children with DLD experience broadly the same types of retention problems as their TD peers, but that these problems are more severe in the DLD group (i.e. occurring more often). This pattern of lower accuracy for the DLD group, but similarities in the error profiles of both groups, has also been found in previous research for other error categorisations (e.g. de Almeida et al., 2019: metathesis, substitutions, omissions; Marton & Schwartz, 2003: additions, omissions, substitutions, order errors). The similarity in the error patterns between the two groups indicates that the DLD group shows a delay in the

ability to repeat nonwords, rather than a deviance compared to their TD peers, which blends in with previous findings of a delay in the phonological development of children with DLD (e.g. Gerrits & de Bree, 2009; Gerrits, 2010). This may imply that the pSTM abilities of children with DLD are weaker, or delayed, but that their pSTM operates similarly to pSTM of TD children.

A considerable proportion of the total amount of errors comprised combined errors and order errors. Corresponding with previous findings (Hsu & Bishop, 2014; Majerus et al., 2009), this implies that children with DLD experienced difficulties with the retention of order information. However, it should be emphasised that order errors were less frequent compared to item errors, which is in line with the findings of Marton and Schwartz (2003) who also found small proportions of order errors for children with and without DLD.

The majority of errors found in the nonword repetitions were situated at the level of item information. That children with DLD have problems with the retention of item information is in contradiction to Majerus et al. (2009), who demonstrated that children with and without DLD did not differ in their ability to recognise phonemic changes, implying no problems with the retention of item information for the DLD group. As the present study involved the production modality rather than just recognition, it may be speculated that item information was lost at the production stage of speech processing (see Figure 1). However, other research has demonstrated that children with DLD experience problems with the retention of item information during speech perception (e.g. Claessen & Leitão, 2012; Criddle & Durkin, 2001; Maillart, et al., 2004). Thus, it is also possible that problems with repeating nonwords were caused by pSTM limitations at a processing stage in perception. The results of the present study are not conclusive about the processing level at which problems were caused with the retention of item and/or order information.

If the pSTM deficit is situated at the perception level, this might lead to incorrect nonword repetitions as follows. The nonword may either be perceived incorrectly, or the representation may not be held long enough in pSTM to accurately specify the nonwords' characteristics (i.e. item and order information), both resulting in an inaccurately specified representation of the nonword. When these inaccurate representations are used at the production stage, this leads to incorrect nonword repetitions by means of item and/or order information. Alternatively, item and order may be accurately specified in the phonological representation during nonword perception. However, if these accurate representations cannot be retained in memory long enough for subsequent production, or if item and/or order information is lost during the information conversion between the input and output buffer, this also leads to incorrect nonword repetitions. As another possibility, nonword repetition may be hampered at the production level. During the encoding process the phonological representation of the nonword needs to be retained in pSTM. If this representation has decayed before the encoding process is finished, it may be speculated that item and/or order information could be lost. Consequently, the incorrect phonemes may be selected, and/or the phonemes may be strung together in an incorrect order, resulting in an incorrect nonword repetition. Further research is needed to examine which processing stage leads to a break-down in nonword repetition.

Regarding the development of error patterns, no differences were found between 72 and 95 months of age, contrary to the tentative prediction that the proportion of the relatively mild order error might have increased. This absence of an age effect might indicate that although retention problems became less severe (i.e. higher performance accuracy), the types of retention problems did not differ with age.

An explorative descriptive analysis of the effect of syllable length on error types showed that for all syllable lengths the overall error pattern was largely the same (i.e. item > combined > order) for both groups. Over the course of increasing syllable length, the proportion of order

errors fluctuated slightly. With increasing length the proportion of item errors decreased, and as predicted the proportion of combined errors increased. This implies that with increasing processing demands item errors were more often made in combination with an order error, which has also been pointed out by Schraeyen (2018) for dyslectic children and their TD peers. However, this does not mean that retaining combined information requires more processing capacity than retaining item or order information. As the error types were counted on word level, it was not clear whether the amount of item and order errors on the segment level also increased with increasing length. Note that these descriptive findings should be interpreted with caution as statistical support is lacking. Further research is needed to validate these results.

6.2. Limitations and future directions

The issue of scoring errors on word level rather than on segment level is a limitation of the present study. Scoring on word level does not reflect 'how much' information is exactly lost in each counted error. In one repetition one, but also more consonants may be produced incorrectly and/or in the incorrect order. Consequently, for each error it was unclear how close the production was to the target nonword, which might in fact reveal the severity of the retention problems. An additional study should be done by differentiating between the amount of errors within each nonword repetition to investigate the differences in error patterns between the two groups more thoroughly. Tentatively, it may be predicted that each nonword production of children with DLD will contain a higher amount of errors, and thus that their produced nonwords will be less close to the target nonwords relative to their TD peers. It might be worth exploring whether measuring this 'closeness' (i.e. multiple vs. single errors) between the targets and nonword repetitions could be used as a potential diagnostic marker.

Another shortcoming of the scoring procedure is that item errors did not reflect a retention loss of purely item information. For example, if all consonants were produced

incorrectly, or if consonants were omitted (e.g. *baa* for *baa.me.rie.noo.ves*) the order of each item was also no longer preserved, or it was unclear whether it had been. It might be more valid to score such productions as combined errors. Moreover, no distinction has been made when scoring substitutions. Substitutions can involve phonemes that do not occur in the nonword (e.g. voo.pee.**k**et > voo.pee.**m**et), or phonemes that occur in the nonword (e.g. anticipation: voo.pee.**k**e<u>t</u> > voo.pee.**t**et, or preservation: voo.<u>p</u>ee.**k**et > voo.**p**ee.**p**et). It may be interesting for further research to disentangle between these two types, as it may be argued that the former substitution type reflects retention problems on the level of item information, whereas anticipations and preservations reflect problems on the level of order information.

Moreover, individual variation within the DLD and TD group might have influenced the results. For instance, previous research has demonstrated that reading skills facilitate the development of nonword repetition (Bishop et al., 2009; Catts, 2005, Conti-Ramsden & Durkin, 2007; Rispens & Parigger, 2010). Reading skills are also demonstrated to be related to the types of errors that are made with repeating nonwords. Schraeyen (2018) found that children with dyslexia (i.e. reading impairment) were outperformed by their TD peers on the retention of order information, whereas no difference was found for the retention of item information. As children with DLD are at risk of reading problems (e.g. Bishop & Snowling, 2004), this variable might have influenced the results of the present study Therefore, it is recommended for future research to examine whether different results will be found between children with DLD who have reading problems, and children with DLD who have average reading skills.

As a last remark, it should be kept in mind that although the DLD group only included children who were diagnosed with DLD, this sample may not be representative for the DLD population as these children were matched to their TD peers by nonverbal IQ, resulting in a mean nonverbal IQ score for the DLD group above the normative mean (i.e. > 100). However, the typical DLD population has been demonstrated to have lower nonverbal IQ compared to

their TD peers (Gallinat & Spaulding, 2014). As nonverbal IQ was a significant covariate for performance accuracy (but not for types of errors), it may be the case that the relatively high nonverbal IQ of the DLD group has influenced the results for performance accuracy. Further research with a more representative sample of the DLD population is needed to see whether the results of the present study would be supported.

6.3. Clinical implications

The NWRT can be used as a diagnostic tool to identify children with DLD (e.g. Boerma et al., 2015; Coady & Evans, 2008). This is supported by the results of present study, which showed that children with DLD can be disentangled from their TD peers by their significantly lower performance accuracy (i.e. higher quantity of errors). The analysis of error patterns revealed that although children with DLD tended to make relatively more combined errors than the TD children, the overall error pattern of children with DLD was largely the same as the pattern of children with TD. Hence, the pattern of error types as categorised in the present study cannot be used as a reliable clinical marker to identify children with DLD.

Moreover, the findings of the present study have implications for the development of interventions. The finding that children with DLD experienced mainly difficulties with the retention of item information, and to a lesser extent with retention of order information indicates that next to retention problems with item information, problems with order information also warrant clinical attention. It is worth exploring whether interventions should target both kinds of retention problems to enhance the efficacy of treatment that aims to improve the speech production abilities of children with DLD.

7. Conclusion

The present study is the first that explored nonword repetition error patterns of children with DLD as categorised by retention problems with phonemic item information (i.e. phonological characteristics) and/or order information (i.e. the phoneme's position). Longitudinal NWRT data were analysed by focussing on the production of consonants. A comparison was made between the nonword repetitions of children with DLD and their TD peers in terms of accuracy performance, and error patterns of item, order and combined errors.

Previous findings were replicated that children with DLD are outperformed by their TD peers on nonword repetition of all lengths, with larger differences between the two groups for longer nonwords (i.e. three, four, five syllables) than shorter ones (i.e. two syllables). Both groups improved significantly throughout 72 to 95 months of age, but at both ages the DLD group was outperformed by the TD group. This indicates a relatively equal growth of pSTM for both groups, and no catch-up at 95 months for the DLD group. Findings on error types revealed that in the DLD group order errors were more often made in combination with item errors compared to the TD group. However, the overall error pattern was largely the same for both groups: Errors were mainly characterised by a retention loss of phonemic item information, followed by retention loss of combined information, and least frequently a retention loss of solely order information. The error patterns did not change with age. Taken together, these findings indicate that the retention problems during nonword repetition of children with DLD are largely the same as the problems of their TD peers, but these problems are more severe for the DLD group. This implies that the DLD group differs from the TD group by a delay in the ability to repeat nonwords, rather than a deviance, indicating that pSTM operates similarly compared to the TD group, but that it is weakened for children with DLD. This result suggests that the error pattern as characterised in the present study cannot be used as a clinical marker to

identify children with DLD, although the groups can be disentangled by the difference in quantity of errors.

This explorative study enhances our understanding of how the erroneous nonword repetitions of children with DLD are characterised relative to their TD peers. This is theoretically of interest, as it reveals whether children with DLD differ from children with TD by a delay, or a deviance in pSTM abilities related to nonword repetition. Additionally, insight in error patterns of nonword repetition contributes to a better understanding of whether the retention problems of children with DLD are mainly situated at the level of item or order information, which is a useful insight for clinical purposes.

References

Aguilar-Mediavilla, E. M., Sanz-Torrent, M., & Serra-Raventos, M. (2002). A comparative study of the phonology of pre-school children with specific language impairment (SLI), language delay (LD) and normal acquisition. *Clinical linguistics & phonetics*, *16*(8), 573-596.

Archibald, L. M., & Gathercole, S. E. (2006). Nonword repetition: A comparison of tests. *Journal of Speech, Language, and Hearing Research, 49*(5), 970-983.

Attout, L., Van der Kaa, M. A., George, M., & Majerus, S. (2012). Dissociating short-term memory and language impairment: The importance of item and serial order information. *Aphasiology*, *26*(3-4), 355-382.

Baddeley, A.D., Lewis, V., & Vallar, G. (1984). Exploring the articulatory loop. *The Quarterly Journal of Experimental Psychology Section A, 36*(2), 233-252.

Baddeley, A.D. (2003). Working memory and language: An overview. *Journal of communication disorders*, *36*(3), 189-208.

Baddeley, A. D., & Hitch, G. J. (2019). The phonological loop as a buffer store: An update. *Cortex*, *112*, 91-106.

Beers, M. (1992). Phonological processes in Dutch language impaired children. *Scandinavian journal of logopedics and phoniatrics*, *17*(1), 9-16.

Beers, M. (1995). *The phonology of normally developing and language impaired children* (Doctoral dissertation). Amsterdam: IFOTT.

Beers, M. (2003). Klankproductieproblemen: Een fonologische benadering. *Stem-, Spraak- en Taalpathologie*, *11*(4), 245-259.

Biran, M., & Friedmann, N. (2005). From phonological paraphasias to the structure of the phonological output lexicon. *Language and Cognitive processes*, *20*(4), 589-616.

Bishop, D. V., North, T., & Donlan, C. H. R. I. S. (1996). Nonword repetition as a behavioural marker for inherited language impairment: Evidence from a twin study. *Journal of child Psychology and Psychiatry*, *37*(4), 391-403.

Bishop, D. V. M. (1997). Cognitive neuropsychology and developmental disorders: Uncomfortable bedfellows. *The Quarterly Journal of Experimental Psychology Section A, 50(4), 899-923*

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., McDonald, D., Bird, S., & Hayiou-Thomas, M. E. (2009). Children who read words accurately despite language impairment: Who are they and how do they do it? *Child development*, *80*(2), 593-605.

Boerma, T., Chiat, S., Leseman, P., Timmermeister, M., Wijnen, F., & Blom, E. (2015). A quasi-universal nonword repetition task as a diagnostic tool for bilingual children learning Dutch as a second language. *Journal of Speech, Language, and Hearing Research*, *58*(6), 1747-1760.

Bonte, M., & Blomert, L. (2004). Developmental changes in ERP correlates of spoken word recognition during early school years: a phonological priming study. *Clinical Neurophysiology*, *115*(2), 409-423.

Bortfeld, H., Morgan, J. L., Golinkoff, R. M., & Rathbun, K. (2005). Mommy and me: Familiar names help launch babies into speech-stream segmentation. *Psychological science*, *16*(4), 298-304.

Bortolini, U., & Leonard, L. B. (2000). Phonology and children with specific language impairment: Status of structural constraints in two languages. *Journal of Communication Disorders*, *33*(2), 131-150.

Botting, N. & Conti-Ramsden, G. (2004). Characteristics of children with specific language impairment. In L. Verhoeven & H. van Balkom (Eds.), *Classification of developmental language disorders: Theoretical issues and clinical implications* (pp. 3-20). Mahwah, NJ: Lawrence Erlbaum Associates.

Briscoe, J., Bishop, D. V., & Norbury, C. F. (2001). Phonological processing, language, and literacy: A comparison of children with mild-to-moderate sensorineural hearing loss and those with specific language impairment. *The Journal of Child Psychology and Psychiatry and Allied Disciplines*, *42*(3), 329-340.

Brooks, P.J. & Kempe, V. (2012). *Language Development*. Chichester, UK: John Wiley & Sons Ltd.

Burgess, N., & Hitch, G. J. (1999). Memory for serial order: A network model of the phonological loop and its timing. *Psychological review*, *106*(3), 551-581.

Burke, H. L., & Coady, J. A. (2015). Nonword repetition errors of children with and without specific language impairments (SLI). *International journal of language & communication disorders*, *50*(3), 337-346.

Catts, H. W., Adlof, S. M., Hogan, T. P., & Weismer, S. E. (2005). Are specific language impairment and dyslexia distinct disorders?. *Journal of Speech, Language, and Hearing Research, 48*(6), 1378-1396.

Claessen, M., & Leitão, S. (2012). Phonological representations in children with SLI. *Child Language Teaching and Therapy*, *28*(2), 211-223.

Coady, J. A., & Aslin, R. N. (2003). Phonological neighbourhoods in the developing lexicon. *Journal of Child language*, *30*(2), 441-469.

Coady, J. A., & Aslin, R. N. (2004). Young children's sensitivity to probabilistic phonotactics in the developing lexicon. *Journal of Experimental Child Psychology*, *89*(3), 183-213.

Coady, J. A., & Evans, J. L. (2008). Uses and interpretations of non-word repetition tasks in children with and without specific language impairments (SLI). *International Journal of Language & Communication Disorders*, *43*(1), 1-40.

Coady, J., Evans, J. L., & Kluender, K. R. (2010). Role of phonotactic frequency in nonword repetition by children with specific language impairments. *International Journal of Language & Communication Disorders*, *45*(4), 494-509.

Conti-Ramsden, G., & Durkin, K. (2007). Phonological short-term memory, language and literacy: developmental relationships in early adolescence in young people with SLI. *Journal of Child Psychology and Psychiatry*, *48*(2), 147-156.

Criddle, M. J., & Durkin, K. (2001). Phonological representation of novel morphemes in children with SLI and typically developing children. *Applied Psycholinguistics*, *22*(3), 363-382.

de Almeida, L., Ferré, S., Barthez, M. A., & Dos Santos, C. (2019). What do monolingual and bilingual children with and without SLI produce when phonology is too complex? *First Language*, *39*(2), 158-176.

Dollaghan, C., & Campbell, T. F. (1998). Nonword repetition and child language impairment. *Journal of Speech, Language, and Hearing Research*, *41*(5), 1136-1146.

Donegan, P. (2013). Normal vowel development. In M.J Ball & F.E. Gibbon (Eds.), *Handbook of vowels and vowel disorders* (pp. 42-78). New York, NY: Psychology Press, Taylor & Francis Group.

Field, A. (2015). *Discovering Statistics Using IBM SPSS Statistics* (4th ed.). Thousand Oaks, CA: SAGE Publications.

Fikkert, P. (2007). Acquiring phonology. In P. de Lacy (Ed.), *Handbook of phonological theory* (pp. 537-554). Cambridge, MA: Cambridge University Press.

Gallinat, E., & Spaulding, T. J. (2014). Differences in the performance of children with specific language impairment and their typically developing peers on nonverbal cognitive tests: A meta-analysis. *Journal of Speech, Language, and Hearing Research*, *57*(4), 1363-1382.

Gallon, N., Harris, J., & Van der Lely, H. (2007). Non-word repetition: an investigation of phonological complexity in children with Grammatical SLI. *Clinical linguistics & phonetics*, *21*(6), 435-455.

Gathercole, S. E., & Baddeley, A. D. (1989). Evaluation of the role of phonological STM in the development of vocabulary in children: A longitudinal study. *Journal of memory and language*, *28*(2), 200-213.

Gathercole, S. E., & Baddeley, A. D. (1990). Phonological memory deficits in language disordered children: Is there a causal connection? *Journal of memory and language*, *29*(3), 336-360.

Gathercole, S., Pickering, S., Ambridge, B., & Wearing, H. (2004). The structure of working memory from 4 to 15 years of age. *Developmental Psychology*, *40*(2), 177–190.

Gerrits, E., & de Bree, E. (2009). Early language development of children at familial risk of dyslexia: Speech perception and production. *Journal of communication disorders*, *42*(3), 180-194.

Gerrits, E. (2010). Acquisition of /s/-clusters in Dutch-speaking children with phonological disorders. *Clinical linguistics & phonetics*, *24*(3), 199-209.

Girbau, D., & Schwartz, R. G. (2007). Non-word repetition in Spanish-speaking children with Specific Language Impairment (SLI). *International Journal of Language & Communication Disorders*, *42*(1), 59-75.

Girbau, D., & Schwartz, R. G. (2008). Phonological working memory in Spanish–English bilingual children with and without specific language impairment. *Journal of communication disorders*, *41*(2), 124-145.

Graf Estes, G., Evans, J. L., & Else-Quest, N. M. (2007). Differences in the nonword repetition performance of children with and without specific language impairment: A meta-analysis. *Journal of Speech, Language, and Hearing Research, 50*(1), 177-195.

Gray, S. (2006). The relationship between phonological memory, receptive vocabulary, and fast mapping in young children with specific language impairment. *Journal of Speech, Language, and Hearing Research, 49*(5), 955-969.

Guiraud, H., Bedoin, N., Krifi-Papoz, S., Herbillon, V., Caillot-Bascou, A., Gonzalez-Monge, S., & Boulenger, V. (2018). Don't speak too fast! Processing of fast rate speech in children with specific language impairment. *PloS one 13*(1), e0191808.

Hsu, H. J., & Bishop, D. V. (2014). Sequence-specific procedural learning deficits in children with specific language impairment. *Developmental science*, *17*(3), 352-365.

Hurlstone, M. J., Hitch, G. J., & Baddeley, A. D. (2014). Memory for serial order across domains: An overview of the literature and directions for future research. *Psychological bulletin*, *140*(2), 339.

Hurlstone, M. J., & Hitch, G. J. (2015). How is the serial order of a spatial sequence represented? Insights from transposition latencies. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *41*(2), 295.

IBM Corp. Released 2017. IBM SPSS Statistics for Windows, Version 25.0. Armonk, NY: IBM Corp.

Jacquemot, C., & Scott, S. K. (2006). What is the relationship between phonological short-term memory and speech processing? *Trends in cognitive sciences*, *10*(11), 480-486.

Johnson, E. K., & Jusczyk, P. W. (2001). Word segmentation by 8-month-olds: When speech cues count more than statistics. *Journal of memory and language*, *44*(4), 548-567.

Johnson, E. K., & Seidl, A. H. (2009). At 11 months, prosody still outranks statistics. *Developmental science*, *12*(1), 131-141.

Jusczyk, P. W., & Aslin, R. N. (1995). Infants′ detection of the sound patterns of words in fluent speech. *Cognitive psychology*, *29*(1), 1-23.

Leonard, L.B. (2009). Is expressive language disorder an accurate diagnostic category? *American Journal of Speech-Language Pathology*, *18*(1), 115-123.

Leonard, L.B. (2014). *Children with specific language impairment* (2nd ed.). Cambridge, MA: MIT Press.

Levelt, W. J. (1992). Accessing words in speech production: Stages, processes and representations. *Cognition*, *42*, 1-22.

Maillart, C., Schelstraete, M. A., & Hupet, M. (2004). Phonological representations in children with SLI. *Journal of Speech, Language, and Hearing Research, 47*(1), 187-198.

Majerus, S., Leclercq, A. L., Grossmann, A., Billard, C., Touzin, M., Van der Linden, M., & Poncelet, M. (2009). Serial order short-term memory capacities and specific language impairment: No evidence for a causal association. Cortex, 45(6), 708-720.

Marton, K., & Schwartz, R. G. (2003). Working memory capacity and language processes in children with specific language impairment. *Journal of Speech, Language, and Hearing Research, 46*(5), 1138-1153.

Mattys, S. L., & Jusczyk, P. W. (2001). Phonotactic cues for segmentation of fluent speech by infants. *Cognition*, *78*(2), 91-121.

Maye, J., Werker, J. F., & Gerken, L. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition*, *82*(3), B101-B111.

Mennen, I., Levelt, C., & Gerrits, E. (2007). Dutch speech acquisition. In S. McLeod (Ed.), *The International Guide to Speech Acquisition* (pp. 327–339). Clifton Park, NY: Thomson Delmar Learning.

Munson, B., Kurtz, B. A., & Windsor, J. (2005). The influence of vocabulary size, phonotactic probability, and wordlikeness on nonword repetitions of children with and without specific language impairment. *Journal of Speech, Language, and Hearing Research, 48*(5), 1033-1047.

Orsolini, M., Sechi, E., Maronato, C., Bonvino, E., & Corcelli, A. (2001). Nature of phonological delay in children with specific language impairment. *International journal of language & communication disorders*, *36*(1), 63-90.

Page, M., & Norris, D. (1998). The primacy model: a new model of immediate serial recall. *Psychological review*, *105*(4), 761-781.

Reilly, S., Bishop, D. V., & Tomblin, B. (2014). Terminological debate over language impairment in children: Forward movement and sticking points. *International Journal of Language & Communication Disorders*, *49*(4), 452-462.

Richards, S., & Goswami, U. (2015). Auditory processing in specific language impairment (SLI): Relations with the perception of lexical and phrasal stress. *Journal of Speech, Language, and Hearing Research*, *58*(4), 1292-1305.

Riches, N. G., Loucas, T., Baird, G., Charman, T., & Simonoff, E. (2011). Non-word repetition in adolescents with specific language impairment and autism plus language impairments: A qualitative analysis. *Journal of communication disorders*, *44*(1), 23-36.

Rispens, J., & Baker, A. (2012). Nonword repetition: The relative contributions of phonological short-term memory and phonological representations in children with language and reading impairment. *Journal of Speech, Language, and Hearing Research, 55*(3), 683-694.

Rispens, J., & Parigger, E. (2010). Non-word repetition in Dutch-speaking children with specific language impairment with and without reading problems. *British journal of developmental psychology*, *28*(1), 177-188.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, *274*(5294), 1926-1928.

Schneider, B. A., Avivi-Reich, M., & Mozuraitis, M. (2015). A cautionary note on the use of the Analysis of Covariance (ANCOVA) in classification designs with and without within-subject factors. *Frontiers in Psychology*, *6*, 474.

Schraeyen, K. (2018*). Item and serial order retention in phonological short-term memory representations in relation to reading performance* (Doctoral dissertation). Leuven: KU Leuven.

Schwartz, R. (2009). Specific language impairment. In R. Schwartz (Ed.), *Handbook of child language disorders* (pp. 3-43). Hove, UK: Psychology Press.

Schwartz, R. G., Scheffler, F. L., & Lopez, K. (2013). Speech perception and lexical effects in specific language impairment. *Clinical linguistics & phonetics*, *27*(5), 339-354.

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.

Stichting Siméa. (2014). Indicatiecriteria: auditief en/of communicatief beperkte leerlingen [Criteria for diagnosis: Auditorily and/or communicatively impaired students]. Retrieved from http://www.simea.nl/dossiers/passend-onderwijs/brochures-po/14-simea-brochure-indicatiecriteria-digitaal.pdf

Stoel-Gammon, C., & Sosa, A.V. (2007). Phonological development. In E. Hoff & M. Shatz (Eds.). *Blackwell handbook of language development* (pp. 238-256). Oxford, UK: John Wiley & Sons.

Stoel-Gammon, C. (2011). Relationships between lexical and phonological development in young children. *Journal of child language*, *38*(1), 1-34.

Sutherland, D., & Gillon, G. T. (2005). Assessment of phonological representations in children with speech impairment. *Language, speech, and hearing services in schools*.

Swingley, D., & Aslin, R. N. (2002). Lexical neighborhoods and the word-form representations of 14-month-olds. *Psychological science*, *13*(5), 480-484.

Thiessen, E. D., & Saffran, J. R. (2003). When cues collide: use of stress and statistical cues to word boundaries by 7-to 9-month-old infants. *Developmental psychology*, *39*(4), 706-716.

Tomblin, J. B., Records, N. L., Buckwalter, P., Zhang, X., Smith, E., & O'Brien, M. (1997). Prevalence of specific language impairment in kindergarten children. *Journal of speech, language, and hearing research*, *40*(6), 1245-1260.

Treiman, R., & Baron, J. (1981). Segmental analysis ability: Development and relation to reading ability. *Reading research: Advances in theory and practice*, *3*, 159-198.

Vandewalle, E., Boets, B., Ghesquière, P., & Zink, I. (2012). Development of phonological processing skills in children with specific language impairment with and without literacy delay: A 3-year longitudinal study. *Journal of Speech, Language, and Hearing Research, 55*(4), 1053-1067.

Vouloumanos, A., & Werker, J. F. (2004). Tuned to the signal: the privileged status of speech for young infants. *Developmental science*, *7*(3), 270-276.

Vouloumanos, A., & Werker, J. F. (2007). Listening to language at birth: Evidence for a bias for speech in neonates. *Developmental science*, *10*(2), 159-164.

Walley, A. C. (1993). The role of vocabulary development in children′s spoken word recognition and segmentation ability. *Developmental review*, *13*(3), 286-350.

Wechsler, D., & Naglieri, J. A. (2008). Wechsler Nonverbal Scale of Ability. Purchased from http://www.pearsonclinical.nl/wnvnl-wechsler-non-verbal.

Werker, J. F. (1989). Becoming a native listener. *American Scientist*, *77*(1), 54-59.

Ziegler, J. C., Pech-Georgel, C., George, F., Alario, F. X., & Lorenzi, C. (2005). Deficits in speech perception predict language learning impairment. *Proceedings of the National Academy of Sciences*, *102*(39), 14110-14115.

Zsiga, E.C. (2013). *The sounds of language: An introduction to phonetics and phonology*. Oxford, UK: Wiley-Blackwell.

Zwitserlood, R. (2014). *Language growth in Dutch school-age children with specific language impairment* (Doctoral dissertation). Utrecht: LOT Publications.

Appendices

Appendix A. The Language-Specific Nonword Repetition Task (Rispens & Baker, 2012), adopted from Boerma et al. (2015, p. 1760)

Syllable length	Phonotactic probability	Orthography	International Phonetic Alphabet
2	high	raanom	rɑnɔm
		daanes	dɑnɛs
		woosel	wosɛl
	low	luubuf	lybʏf
		kuimup	koeymʏp
		joefeum	jʊfø:m
3	high	kaaroodin	kɑrodɪn
		voopeeket	vopekɛt
		deevoenos	devʊnɔs
	low	veujoetup	vø:jʊtʏp
		nuigeusup	noeyxø:sʏp
		muihuuguf	moeyhyxʏf
4	high	liekoovoepar	likovʊpar
		kooviewaalan	koviwɑlan
		liejootaanig	lijotɑnɪx
	low	guiweusoegeer	xoeywø:sʊxɪr
		meufuusuinef	mø:fysoeynɛf
		juuvuigoowuf	jyvoeyxowʏf
5	high	wookaaloemoodon	wokɑlʊmodɔn
		baamerienooves	bɑmɛrinovɛs
		tieloniedaanag	tilɔnidɑnax
	low	fuugiwuinoefep	fyxɪwoeynʊfɛp
		geumuwoekuubir	xø:mʏwʊkybɪr
		nuijigeufuusut	noeyjɪxø:fysʏt

Appendix B. Detailed additions to NWRT scoring procedures.

(a) If the produced nonword contained more consonants than the target nonword, the order of a correctly produced consonant was scored correct if its position was the same as the position of that consonant in the target, when counted from either the beginning or the end of the sequence. If at least one consonant did not obey this principle, the production was scored as order error. For example, if *mui.huu.guf* was produced as *maa.lu.gij.juf*, order was preserved: When counting from the beginning of the nonword, /m/ is in onset position, /g/ has remained its position in the third syllable, and when counting from the end of the sequence, /f/ is in final position. In contrast, if *fuu.gi.wui.noe.fep* was produced as *fie.nie.jui.gee.fu.fep*, order was considered as incorrect due to the migration of /g/.

(b) If the produced nonword contained less consonants than the target nonword, the relative order of the consonants was considered to determine whether order was preserved. For example, in *juu.gij* for *juu.vui.goo.wuf* the relative order was preserved, and thus this was not scored as order error. In contrast, in a production like *lie.noo.taa* for *lie.joo.taa.nig* the relative order of the produced consonants has been changed (i.e. /n/ proceeds /t/ rather than the other way around), and this was scored as order error.

(c) In cases that the position of a correctly produced consonant within a syllable (i.e. onset or coda) was ambiguous, the child's production was interpreted in such a way that it matched the target nonword to the greatest extent that could be reasoned. Put differently, the child was given the benefit of doubt. For example, in cases that *nui.ji.geu.fuu.sut* was produced as *nuihegeefuus*, it was assumed that the position of the last produced consonant was preserved, and thus the position of /s/ was interpreted as onset of the fifth syllable (i.e. *nui.he.gee.fuu.s*), rather than the coda consonant of the fourth syllable. Thus, this production was not scored as a serial order error.

(d) In a few cases less syllables were produced than the target consisted of, ánd the consonants that were produced were not present in the target. In these cases it was assumed that the child tried to produce the nonword in serial order from its beginning to its end. For example, if *fuu.gi.wui.noe.fep* was produced as *fuu.goe.l*, this production was matched on the target sequence *fuu.gi.w*.

(e) If the repetition consisted of a single produced syllable, it could not clearly be determined whether serial order was affected or not. In these cases, it was assumed that the child had obtained serial order, and thus these errors were not coded as order errors (e.g. *fuu.gi.wui.noe.fep* produced as *wu*).

Appendix C. Mean percentages of error types on the NWRT across different syllable lengths for children with TD and DLD at T1 and T2.

Time point	Error Type	Syllable Length	TD			DLD		
			n	*M (SD)*	Range	*n*	*M (SD)*	Range
1	Item	All	39	71.1 (18.5)	36.4-100.0	39	68.6 (13.4)	36.8-100.0
		2	37	96.2 (11.5)	50.0-100.0	38	92.9 (15.9)	50.0-100.0
		3	33	72.2 (36.9)	0.0-100.0	38	71.3 (25.8)	0.0-100.0
		4	36	72.2 (28.8)	0.0-100.0	39	63.8 (22.2)	20.0-100.0
		5	38	52.3 (29.7)	0.0-100.0	39	58.0 (25.4)	0.0-100.0
	Order	All	39	6.3 (7.7)	0.0-25.0	39	3.6 (5.2)	0.0-25.0
		2	37	0.0 (0.0)	0.0-0.0	38	2.6 (9.7)	0.0-50.0
		3	33	12.5 (24.2)	0.0-100.0	38	7.2 (18.9)	0.0-100.0
		4	36	9.4 (20.9)	0.0-100.0	39	2.9 (7.7)	0.0-25.0
		5	38	5.9 (11.3)	0.0-40.0	39	1.9 (5.7)	0.0-20.0
	Combined	All	39	22.6 (15.6)	0.0-54.5	39	27.9 (11.3)	0.0-52.6
		2	37	3.8 (11.5)	0.0-50.0	38	4.5 (10.7)	0.0-33.3
		3	33	15.4 (33.6)	0.0-100.0	38	21.5 (21.3)	0.0-80.0
		4	36	18.5 (20.6)	0.0-66.7	39	33.3 (20.2)	0.0-80.0
		5	38	41.8 (29.7)	0.0-100.0	39	40.1 (24.8)	0.0-100.0
2	Item	All	39	70.1 (15.7)	46.2-100.0	39	62.2 (13.7)	33.3-100.0
		2	36	94.9 (14.8)	50.0-100.0	39	90.8 (19.4)	0.0-100.0
		3	34	75.0 (37.4)	0.0-100.0	38	64.1 (33.2)	0.0-100.0
		4	34	65.5 (35.4)	0.0-100.0	39	60.0 (25.8)	0.0-100.0
		5	39	54.1 (31.6)	0.0-100.0	39	49.1 (25.6)	0.0-100.0
	Order	All	39	11.4 (10.3)	0.0-28.6	39	7.2 (7.4)	0.0-27.3
		2	36	1.4 (8.3)	0.0-50.0	39	2.3 (8.3)	0.0-33.3
		3	34	11.8 (27.7)	0.0-100.0	38	15.4 (25.0)	0.0-100.0
		4	34	14.6 (23.8)	0.0-100.0	39	6.5 (11.9)	0.0-40.0
		5	39	18.0 (26.7)	0.0-100.0	39	4.9 (10.4)	0.0-40.0
	Combined	All	39	18.5 (16.9)	0.0-53.8	39	30.6 (13.0)	0.0-61.1
		2	36	3.7 (12.7)	0.0-50.0	39	18.4 (6.8)	0.0-100.0
		3	34	13.2 (30.9)	0.0-100.0	38	26.7 (20.4)	0.0-100.0
		4	34	19.9 (28.5)	0.0-100.0	39	25.9 (33.5)	0.0-100.0
		5	39	27.9 (29.1)	0.0-100.0	39	46.0 (25.8)	0.0-100.0