

Reciprocal Relationships between Nonword Repetition and Vocabulary during the
Preschool Years

Josje Verhagen^{1,2*}, Jan Boom^{3*}, Hanna Mulder¹, Elise de Bree⁴, & Paul Leseman¹

¹Utrecht University, Department of Special Education: Cognitive and Motor Disabilities

Heidelberglaan 1, 3584 CS, Utrecht, the Netherlands

²University of Amsterdam, Amsterdam Center for Language and Communication,

Spuistraat 134, 1012 VB, Amsterdam, the Netherlands

³Utrecht University, Department of Developmental Psychology

Heidelberglaan 1, 3584 CS, Utrecht, the Netherlands

⁴University of Amsterdam, Institute of Child Development and Education

Nieuwe Achtergracht 127, 1001 NG, Amsterdam, the Netherlands

*Co-first author

Corresponding author:

Dr. Josje Verhagen,

University of Amsterdam, Amsterdam Center for Language and Communication

Spuistraat 134, 1012 VB, Amsterdam, the Netherlands

Phone number: +31-(0)20-2523881, Email address: J.Verhagen@uva.nl

Abstract

The aim of this longitudinal study is to evaluate three views on the relationship between nonword repetition and vocabulary: (i) the storage-based view that considers nonword repetition, a measure of phonological storage, as the driving force behind vocabulary development, (ii) the lexical restructuring view that considers improvements in nonword repetition as the result of vocabulary growth, and (iii) the “combined” view that assumes that both storage-based learning and lexical restructuring play a role, resulting in reciprocal relationships between nonword repetition and vocabulary during language development.

Data are analyzed from 471 monolingual Dutch children who performed tasks assessing nonword repetition and vocabulary at yearly intervals, from ages two to five. Latent Change Score (LCS) modeling of Item Response Theory (IRT)-scaled scores was used to investigate the relationships between nonword repetition and vocabulary growth over time. Additionally, the statistical techniques used in earlier work – cross-lagged and latent growth modeling – were applied to see whether the results changed as a function of the analytical technique used.

Results from a bivariate LCS model showed positive reciprocal influences from nonword repetition on vocabulary between two and five years. Such positive cross-influences also emerged from the cross-lagged and latent growth models. Predictive relationships from vocabulary to nonword repetition were stronger than vice versa. These results indicate that both storage-based learning and lexical restructuring play a role in vocabulary learning, at least in early stages of language development, with the clearest support found for lexical restructuring.

Keywords: nonword repetition, vocabulary, phonological storage, lexical restructuring, Latent Change Score modeling, preschoolers

Reciprocal Relationships between Nonword Repetition and Vocabulary during the
Preschool Years

Vocabulary knowledge is pivotal in young children's development and an important predictor of later academic skills, in particular early literacy and reading (Duff, Reen, Plunkett, & Nation, 2015; Lee, 2011; Storch & Whitehurst, 2002). Children typically show fast rates of vocabulary learning throughout the preschool period, but there is substantial variation both in the level of children's vocabulary knowledge in their early years and the rate at which this knowledge increases (Huttenlocher, Haight, Bryk, Seltzer, & Lyon, 1991; Rowe, Raudenbusch, & Goldin-Meadow, 2012). Given the predictive value of young children's vocabulary knowledge for later academic attainment, it is important to investigate the determinants of differences in vocabulary knowledge in early stages of language development, when interventions may already be effective in increasing children's vocabulary skills (Fricke, Bowyer-Crane, Haley, Hulme, & Snowling, 2013; for a review, see Marulis & Neuman, 2010).

One determinant of children's vocabulary skill that has been proposed in earlier work is children's ability to repeat novel words, that is, nonword repetition. However, in previous work, it has been debated whether nonword repetition is indeed the driving force behind vocabulary growth (Baddeley, Gathercole, & Papagno, 1998) or, alternatively, the result of it (Metsala, 1999). These two views are not necessarily mutually exclusive. In fact, a third view has been proposed that assumes that nonword repetition and vocabulary mutually influence each other during development (Rispen & Baker, 2012; Snowling, 2006). The aim of the current study is to evaluate these views by investigating the directionality of the relationships in a longitudinal study of Dutch two- to five-year-old children.

As stated above, three views on the relationship between nonword repetition and vocabulary knowledge have been posited in earlier work. The first is that phonological storage, tapped with nonword repetition tasks, is needed for learning new words (Baddeley, Gathercole, & Papagno, 1998; for a review, see Gathercole, 2006). This view, here referred to as the storage-based view, holds that nonword repetition skill reflects the ability to store unknown phonological material, and as such underlies learning of new words. Specifically, upon encountering a novel word, children need to store this novel word in short-term memory, from where they can extract its relevant phonological features (Gathercole, 2006). Numerous cross-sectional studies have found associations between nonword repetition and vocabulary knowledge in children of different ages, including children as young as two years of age (Hoff, Core, & Bridges, 2008; Stokes & Klee, 2009; Verhagen, et al., 2017). Further evidence comes from word learning studies, showing that nonword repetition ability and novel word learning are correlated (Gathercole & Baddeley, 1990).

The second view on the relationship between nonword repetition and vocabulary holds that better nonword repetition is the result of growth in vocabulary knowledge rather than its underlying cause (Bowey, 2001; Metsala, 1999; Metsala & Walley, 1998). According to this view, very young children represent words initially in a relatively holistic manner, without much phonological detail. As their vocabularies increase, children shift towards more analytic representations at either the syllable or phoneme level. These more detailed phonological representations facilitate performance on phonological tasks, such as phoneme awareness tasks and nonword repetition. Thus, in this so-called lexical restructuring view (Metsala, 1999; Metsala & Walley, 1998), vocabulary knowledge drives nonword repetition rather than vice versa.

Finally, it has been proposed that nonword repetition and vocabulary influence each other during development (Hoff, Core, & Bridges, 2008; Rispens & Baker, 2012; Snowling, 2006), such that storage-based learning and lexical restructuring work in parallel. In this account, an increase in phonological storage capacity underlies vocabulary growth. This growth in vocabulary, in turn, leads to more fine-grained phonological representations and, subsequently, to an increased ability to represent phonological information, which surfaces as better performance on nonword repetition tasks (Hoff, Core, & Bridges, 2008; Snowling, 2006). Support for this idea comes from a cross-sectional study by Rispens and Baker (2012) who found that both a measure of phonological storage (i.e., digit span) and a measure of phonological representation (i.e., nonword discrimination) were significant predictors of nonword repetition ability, at least in the five-year-olds tested in their study.

To investigate in more detail how nonword repetition and vocabulary are interrelated during children's development, longitudinal studies are needed. Unlike cross-sectional studies, longitudinal studies provide insight into how one variable is related to another over time, and whether this relationship is reciprocal, and as such, can yield information about the direction of effects. Although a wealth of studies have investigated the relationships between nonword repetition and vocabulary concurrently, only two longitudinal studies hitherto have tested whether nonword repetition influences vocabulary in children's development, and/or vice versa (Gathercole, Willis, Emslie, & Baddeley, 1992; Melby-Lervåg et al., 2012). These studies have provided mixed results.

In the first study, Gathercole et al. (1992) assessed nonword repetition and receptive vocabulary in a sample of 80 English-speaking children at four time points, between four and eight years of age. Using cross-lagged correlations, the authors found that nonword repetition at

age four predicted vocabulary at age five. At later ages, vocabulary predicted nonword repetition, but not vice versa. To explain these findings, the authors argued that children's growing vocabularies enabled them to increasingly rely on lexical strategies rather than phonological storage. They also proposed that reading ability, which develops at the oldest ages investigated in their study, might have increasingly predicted children's vocabularies and, hence, overshadowed the contribution of phonological storage.

In the second study, Melby-Lervåg et al. (2012) replicated the work by Gathercole et al. with a sample of 219 Norwegian children, assessed longitudinally from four to seven years. Applying two types of analyses to their data (cross-lagged models and latent growth models), these authors found that the only cross-relationship that approached significance was from nonword repetition at age four to vocabulary at age five ($\beta = 0.17, p = .079$). Furthermore, when applying cross-lagged models and latent growth modeling to the data collected by Gathercole et al. (1992), Melby-Lervåg et al. found that no significant cross-loadings between nonword repetition and vocabulary remained. Based on these results, Melby-Lervåg et al. concluded that the idea that vocabulary learning is constrained by nonword repetition should be questioned, as well as the broader theory that the phonological loop acts as a language-learning device (Baddeley, Gathercole, & Papagno, 1998). Note that the results of this study also speak against a lexical restructuring view, because no predictive relationships from vocabulary to nonword repetition were found either.

A factor that might explain, at least in part, the near lack of significant cross-domain predictive relationships in the above-reviewed studies is the fact that children beyond age four were tested. Both the storage-based and lexical restructuring views assume that relationships between nonword repetition and vocabulary are most prominent at early stages of language

development and become weaker over time. On a storage-based account, the assumption is that storage-mediated learning is increasingly replaced with lexicon-based learning mechanisms in the course of development, as children's vocabularies become more substantial (Gathercole, 1995, 2006). Support for this idea comes from cross-sectional studies showing that the relationship between nonword repetition and vocabulary is stronger for younger than for older children (Gathercole, 1995; Gathercole et al., 2005), and stronger in beginning than in more advanced second language learners (Cheung, 1996; Gathercole, 1995; Masoura & Gathercole, 2005).

Lexical restructuring is also considered an early process. Previous findings suggest that it is still ongoing at ages five and six (Rispen & Baker, 2012; Storkel, 2002) and ends around age eight (Walley, Metsala, & Garlock, 2003). In their cross-sectional study, already referred to above, Rispen and Baker (2012) found, for instance, that a measure of phonological representation explained a significant portion of the variance in nonword repetition skills in five-year-olds, and in fact, was a stronger predictor than digit span, a measure of phonological storage. However, this measure did not explain variance in eight-year-olds' nonword repetition. This suggests that there is a developmental shift between the ages of five and eight, such that the quality of phonological representations, which is assumed to be positively affected by growth in children's vocabularies, becomes increasingly less predictive of children's nonword repetition ability.

The aim of the current study is to evaluate the above-discussed views by investigating the directionality of the relationships between nonword repetition and vocabulary. A longitudinal investigation will be presented, which complements existing studies in two ways. First, younger children than in earlier work will be assessed: preschoolers from two to five years of age.

Studying young children is important, given that earlier cross-sectional work indicates that relationships between nonword repetition and vocabulary skills are typically stronger in younger than in older children. Second, Latent Change Score (LCS) modeling will be used, which seems better suited to study the potential interrelations between abilities during development than the analytical techniques used previously (Ferrer & McArdle, 2004, 2010; Grimm, 2008; McArdle, 2009). LCS modeling allows for the assessment of change of two or more cognitive abilities as they develop over time, as well as the potential cross-lagged interrelations between changes in these abilities during their development. In previous work, LCS models have been applied to study whether one cognitive process is a predictor of subsequent change in another process. Grimm (2008), for example, used LCS modeling to show that improved reading ability predicted later improvement in mathematics during middle to late childhood.

Crucially, LCS modeling allows for statistical testing of relationships that cannot be tested with other frequently used methods for longitudinal data such as cross-lagged panel analyses and latent growth analyses (Ferrer & McArdle, 2004; McArdle, 2009). Although cross-lagged models can indicate influences across different abilities over time, they do not capture growth or decline in the data. Latent growth modeling (LGM) is a useful method for assessing overall associations between changes in two or more variables, but this technique does not provide information about which variables lead or lag in such longitudinal relationships. LCS modeling combines assessment of growth and reciprocal relationships among multiple processes. As such, it provides a useful method for statistically testing the current research question of whether (changes in) nonword repetition skills underlie (changes in) vocabulary skill, and/or vice versa, *over time*.

Our predictions are as follows. First, assuming that phonological storage is an important determinant of early stages of lexical development (Gathercole, 2006), significant predictive relationships from nonword repetition to vocabulary are expected in the period from two to five years. Second, based on earlier work showing that lexical restructuring takes place during the preschool years (Walley, Metsala, & Garlock, 2003), we expect significant predictive relationships from vocabulary to nonword repetition. Thus, we expect that there will be positive reciprocal relationships between both skills over the preschool period in line with the above-described third view on the relationship between nonword repetition and vocabulary learning.

Method

Participants

Participants were 471 monolingual Dutch children who participated in a large-scale study of preschool children in the Netherlands (pre-COOL, cf. Mulder et al., 2014; Mulder, Verhagen, van der Ven, Slot, & Leseman, 2017). The primary aim of this study was to assess the effects of early childhood education and care on young children's linguistic and cognitive development. The current participants constituted a sub-sample of all participants involved in the pre-COOL study.

The pre-COOL study included over 3000 children who were assessed with an executive function and language test battery at age(s) two and/or three years. A subset of these children were selected for further assessments at ages four and five. The main criteria for inclusion in this so-called "core cohort" were the following: (i) children had obtained a test score on the vocabulary task as well as at least three other tasks of the language and executive function test battery at ages two and/or three years, and (ii) contact information about children's schools was available. The rationale behind these criteria was to obtain a dataset for the children in the core

cohort that was as complete as possible in order to be able to address the main question guiding the pre-COOL study. A total of 751 children were included in the core cohort and assessed at ages four and five. For the current study, children were selected from this core cohort if they were from monolingual Dutch families. This resulted in a sample of 471 children.

Children were recruited from families and early childhood education and care centers that were geographically spread over the Netherlands and located in rural and urban areas.

Educational level of children's parents was assessed through a parent questionnaire, on a 4-point scale ranging from (1) 'primary school', (2) 'lower vocational training', (3) 'secondary school and/or vocational training', to (4) 'higher education (i.e., college or university). Data was available for 435 children (92%), and showed that 7.8% of children's parents were educated at the primary school or lower vocational training level and about 52.2% at the higher education level. This indicates that the higher education levels were overrepresented relative to the general Dutch population (18.6% low educated vs. 32.3% high educated, cf. Statline, 2018).

Main sample characteristics (age, gender) and task completion rates for each wave are provided in Table 1. As can be seen from this table, task completion rates were lowest at the first wave, but generally high (i.e., above 85%) at all waves.

{Insert Table 1 here}

Approval for the pre-COOL study was obtained from the Ethical Advisory Committee of the Faculty of Social and Behavioral Sciences of Utrecht University and the Ethical Advisory Committee of the Department of Education of the University of Amsterdam.

Materials

Nonword repetition. At each wave, children completed a nonword repetition task, which contained overlapping items across successive waves. This overlap in items across waves allowed for assessing the broad ability range characteristic of young children’s developing skills with relatively brief tasks at each wave, enabling us to avoid taxing young children’s short attention spans and minimize floor or ceiling effects. Item Response Theory (IRT) modeling (see Analyses) was used to scale all items on the same latent ability scale.

At wave 1, the task contained six 1-syllable and six 2-syllable nonwords. At wave 2, the six 2-syllable nonwords of wave 1 were repeated and supplemented with six 3-syllable nonwords. At wave 3, the six 3-syllable items were repeated and supplemented with six 4-syllable nonwords. At the final wave, the six 4-syllable items were repeated and supplemented with six 5-syllable nonwords. As such, there was a small number of unique items at the first and last study wave, while the remaining items overlapped across successive waves (see Table 2).

{Insert Table 2 here}

At waves 3 and 4, items were used that were a subset of the items of the nonword repetition task reported in Rispens and Baker (2012). These had been designed for Dutch children and manipulated for item length. For waves 2 and 3, no items were available. Therefore, new items were constructed, which were very similar to the items used at waves 3 and 4 in that they had simple syllable structure, conformed to the main Dutch stress pattern, and contained early acquired phonemes, except that they were shorter, to make them appropriate for young children (see also Verhagen et al., 2017).

All items had been pre-recorded in a soundproof room by a female speech and language therapist using a high pitch voice that is typical of child-directed speech. The recordings were then embedded in the following procedure to keep children engaged in the task: children watched short video clips in which a novel object appeared from a picture of a box on a laptop screen. At the same time, they heard a pre-recorded sentence labeling the object that encouraged them to repeat the nonword: *'Look, a [jaat]! Say [jaat]!* The purpose of playing movie clips and pre-recorded speech on a laptop was to keep children engaged in the 'game', while at the same time ensuring uniformity of input in terms of rate, pitch, volume, and other phonetic and auditory features that may otherwise vary across and within speakers. Such uniformity of input seemed especially important in a study of this scale in which different research assistants administered the task.

In all assessments, items were presented in a fixed, pseudo-randomized order in which no more than two items of the same length were presented consecutively. Two practice trials were presented to familiarize children with the procedure, and short breaks were allowed if children needed a short pause. Responses were scored online by trained assistants (i.e., immediately after each response), as either 'correct' or 'incorrect'. The option 'unclear' was used if children's responses could not be categorized as either correct or incorrect by the assistant (2% of all responses), and coded as missing in the analyses. The test showed good to excellent internal consistency at all waves (Cronbach's alphas were between .81 and .95).

Receptive vocabulary. A shortened version of the Dutch *Peabody Picture Vocabulary Test* (PPVT-III-NL; Dunn & Dunn, 2005) was used to assess receptive vocabulary. In this task, children were asked to select one out of four picture drawings after an orally presented word. Although this task is usually performed as a paper-and-pencil task, it was administered on a

laptop in the current study, to facilitate administration and scoring. Moreover, to reduce testing time, a shortened version was administered from which items had been removed that appeared not to differentiate well across children, as indicated by pilot studies with children of the relevant ages ($n = 111$ at wave 1, $n = 52$ at wave 2, $n = 33$ at wave 3, $n = 36$ at wave 4). Specifically, items were removed if they were either very easy or very difficult (i.e., average accuracy was below 30% or above 70% for these items in the pilot studies). A fixed number of items was administered to all children at each wave (i.e., 24 items at waves 1 and 2, 28 items at waves 3 and 4). Internal consistency of the task was sufficient to good at all waves (Cronbach's alphas ranged between .70 and .85).

The vocabulary task was constructed along the same lines as the nonword repetition task. In this task, too, a subset of the items was kept constant from one wave to the next, whereas the remaining items were newly added (see Table 3). This structure was chosen to enable proper measurement within waves of children's rapidly developing vocabularies while avoiding floor or ceiling effects, and, at the same time, allowing scaling of all items on a latent ability scale through IRT modeling.

{Insert Table 3 here}

Procedure

Children were tested by trained research assistants in a quiet room at their daycare centers, preschools, or at their homes. Tasks were part of a larger battery of tasks and presented in a fixed order in which the vocabulary task preceded the nonword repetition task. To optimize standardized assessment, research assistants had undergone an intensive training before they

were allowed to start data collection at each study wave. First, they had attended a full day test administration course. Second, they had received a detailed standardized test protocol with step-by-step descriptions of the procedures for each measure. Third, they had submitted a video recording of a practice session with a child of the relevant age to the principal investigators, together with their scoring sheets. The test administration procedures and scoring sheets were reviewed by the investigators, and each research assistant was sent a detailed feedback report. If the research assistant had followed the standardized protocol, they were allowed to start data collection. If administration or coding errors were observed, extensive feedback was provided and the research assistant was asked to submit a second video for another round of feedback. The investigators discussed any difficult cases until agreement was reached, and read each other's feedback reports before sending these to the assistants, to ensure that no divergence in their evaluations occurred throughout the process.

Analyses

Interrelations between nonword repetition and vocabulary across the four measurement waves were examined using Latent Change Score (LCS) modeling. A typical bivariate LCS model is illustrated in Figure 1.

{Insert Figure 1 here}

In a bivariate LCS model, two sets of scores, X and Y , are observed at t measurement waves (squares X_t and Y_t in Figure 1). Underlying these scores, latent counterparts (circles x_t and y_t) are assumed, which represent the true scores. At each wave, this true score inherits the value

from the previous wave (and, in case of the first wave, from the initial intercepts [i_x and i_y]), as in a cross-lagged model, by a fixed regression coefficient of 1. In addition, at each next wave, latent variables (Δx_t and Δy_t) are added to the model, which represent changes in the true scores. Importantly, these latent change scores are a function of three types of to be estimated fixed effect loadings: (i) α loadings (α_x, α_y) for the influence of the initial slopes (s_x, s_y), (ii) β loadings (β_x, β_y) for the influence of the scores of the same ability at the previous measurement wave (x_{t-1} and y_{t-1}), and (iii) γ loadings (γ_x, γ_y) for the influence of the scores of the other ability at the previous measurement wave (y_{t-1} and x_{t-1}). The β loadings are also referred to as self-feedback parameters; the γ loadings as coupling parameters.

These γ coupling parameters are especially important in light of the questions addressed in the current study, as they indicate the degree to which the level of an ability at time t predicts changes in the other ability at the next wave ($t+1$). Specifically, on a storage-based view of word learning, significant positive estimates for the coupling parameters are predicted from nonword repetition to vocabulary, but not vice versa, as nonword repetition is considered the driving force behind vocabulary development. Within the lexical restructuring account, in contrast, significant positive estimates for the coupling parameters from vocabulary to nonword repetition are predicted, but not vice versa, since growth in vocabulary is assumed to lead to more detailed phonological representations, which, in turn, facilitate nonword repetition. Finally, on the combined view, which, as we hypothesized above, may apply to the current young age group, all estimates for the coupling parameters should be positive and significant, since both storage-based learning and lexical restructuring play a role.

In LCS modeling, coupling parameters are usually held constant over time, to make them easier to interpret, but they can also be freely estimated (Ferrer & McArdle, 2004; Ferrer et al.,

2007). This allows one to investigate whether relationships between abilities change within the time period investigated. Since our study spanned three years in a period of rapid language development (i.e., children's preschool years), in which relationships between vocabulary and nonword repetition may decrease with increasing age (see Introduction), the coupling parameters were allowed to be different over waves in our model.

We report unstandardized values for the regression parameters, since standardized values are less meaningful than unstandardized values in LCS models, and difficult to interpret (McArdle, 2001; see also Butner, Berg, Baucom, & Wiebe, 2014). Note that the two initial intercepts ($i-x$ and $i-y$) and two initial slopes ($s-x$ and $s-y$) in the model are random variables and therefore different for each participant, while the loadings are the same for each participant, and that we report unstandardized averages and variances, but standardized covariances for these random variables.

In the current study, a bivariate LCS model was fitted to the data to investigate the interrelationships between the latent abilities underlying children's nonword repetition and vocabulary skill. Since all items in our study were binary and most of the items overlapped across successive waves, the LCS model was supplemented with an IRT (Item Response Theory) component (Grimm, Kuhl, & Zhang, 2013; McArdle et al., 2009), estimated *at the same time*, in one model.

IRT allows for the linkage of response accuracy on a given item to both the underlying latent ability of the participant and the latent difficulty of the item (Embretson & Reise, 2000). The larger the positive difference between this ability and the difficulty (threshold) of a particular item, the larger the likelihood of a correct response, and vice versa. An advantage of IRT modeling is that it can deal with differences in item difficulty: assuming a certain ability

level, easier items are more likely to elicit a correct response than more difficult items. Likewise, assuming a certain item difficulty level, participants with a higher ability level are more likely to pass the item than participants with a lower ability level. Importantly, for items that are the same across two successive waves, we assumed the same difficulties (thresholds). In this way, increases in ability level across waves for these items could be estimated, enabling the construction of a latent ability scale of the construct being measured, based on all – both unique and overlapping – items across waves (Grimm, Kuhl, & Zhang, 2013; McArdle et al., 2009).

Thus, the current model differed from the typical LCS model used in most studies and illustrated in Figure 1, in that, instead of the single observed variables (X_t and Y_t), children's binary response scores on the multiple items administered at each wave were loaded on four latent factors (x_t and y_t), each representing the latent ability of the construct at a given wave. *Item loadings* of items on their factor were all fixed to one. *Item thresholds* were free to differ across items, with one exception: thresholds were constrained to be equal for items that were repeated at the next wave, to ensure the same metric over these waves. By freeing the item thresholds, differences in item difficulty were taken into account. In addition, *item residual variances* of the items were free to differ between items, with one exception: residual variances were constrained to be equal for items that were repeated at the next wave, to ensure the same metric. By freeing the item residual variances, differences in item discrimination were taken into account. Note that item loadings, item thresholds, and item residual variances, although freely estimated, were the same across participants.

To facilitate comparison of nonword repetition and vocabulary abilities we fixed the average of the initial intercept at zero and the average of the initial slope at one for both abilities. This caused no problems because the scale was arbitrary and threshold estimates accommodated

such that fit was not negatively affected. In this way, we created a model, which, in its IRT part, was optimized for the specific properties of our data, and, in its LCS part, was optimized to address our research questions.

The resulting model was rather complex. To reduce model complexity and eliminate convergence problems, we first removed six items from the vocabulary task (i.e., items 34, 55, 64, 65, 66, 67), five of which were from wave four and non-overlapping with those of wave three. These six items loaded very weakly or even negatively onto the latent factor. Importantly, leaving out these items did not change the main parameters of the LCS part, but resolved convergence problems and reduced computation time considerably. Next, we removed four more items from the vocabulary task because of ceiling/floor effects. For those items, virtually all children either passed or failed the item. This left us with 141 items in total: 93 for vocabulary and 48 for nonword repetition.

We used Mplus 7.4, with the Weighted Least Squares Means and Variances adjusted estimator (WLSMV), a PROBIT link, and Theta parameterization (Muthén & Muthén, 1998-2012). The PROBIT link means that the aforementioned differences between difficulty and ability are expressed in units that refer to the standard normal distribution with a mean of zero. An advantage of using the WLSMV estimator with a PROBIT link is that an absolute overall fit measure is available. This overall fit of the model was evaluated in terms of the indices Root Mean Square Error of Approximation (RMSEA), Comparative Fit Index (CFI), and Tucker Lewis Index (TLI), following commonly applied cut-off criteria. Specifically, model fit was considered good if RMSEA was below .05, and CFI and TLI were above .90 (Little, 2013). Another advantage of using the PROBIT model is that discrimination differences between items

could be added to the model (Asparouhov & Muthén, 2015). So, model fit assessed the ability to model responses across all test items for all participants.

Our analyses proceeded in three steps. First, a bivariate LCS model of nonword repetition and vocabulary skill was fitted, containing an integrated IRT component, as explained above. Of particular importance to our research question were the cross-loadings (coupling parameters) between the latent abilities underlying nonword repetition and vocabulary, because these indicate the changes in an ability from one wave to the next, caused by the other ability at the previous wave. This model had 202 free parameters.

Second, to see if there was evidence for a bidirectional relationship between vocabulary and nonword repetition three alternative models were fitted, after fitting the LCS model with all coupling parameters. In the first model, the three coupling parameters from x_t to Δy_{t+1} were removed. In the second, the three coupling parameters from y_t to Δx_{t+1} were removed. Both models had 199 free parameters. In the last model, both sets of coupling parameters were removed (for similar model comparisons, see Ferrer et al., 2007, Ferrer & McArdle, 2004; Quinn, Wagner, Petscher, & Lopez, 2015). This model had 196 free parameters. Model fit of the resulting models in which either one or both coupling parameters were left out was compared with the model containing all coupling parameters. However, in this comparison, regular chi-square difference testing could not be used due to using the WLSMV estimator (Muthén & Muthén, 1998-2012). The alternative DIFFTEST-procedure for WLSMV models provided by Mplus could not be applied to our models either, due to constraining item difficulties of overlapping items to be equal across waves. Instead, we present corrected chi-square difference testing based on running all models with the more conservative WLSM estimator. This should give the exact same values for all parameter estimates, which, indeed, was the case in our

models. If a significant result was obtained from the chi-square difference test, the more complex model (containing more coupling parameters) was considered the preferred model.

Finally, to see whether our results were specific to the type of analysis used, we checked whether the same results were obtained if we analyzed our data using cross-lagged models and LGM – the analyses used in Melby-Lervåg et al. (2012). For a more detailed description of these analyses, see the supplemental materials online.

Results

Longitudinal Relationships between Nonword Repetition and Vocabulary

Descriptive statistics of proportion correct scores for vocabulary and nonword repetition at each wave are shown in Table 4. It should be stressed here that, although these data provide general information regarding potential ceiling effects and inter-individual variation in the data, they are not informative regarding growth of nonword repetition and vocabulary skill, due to the use of partly different item sets across waves.

{Insert Table 4 here}

The relative differences in growth of children’s latent abilities underlying vocabulary and nonword repetition from waves one to four are presented in Figure 2. As mentioned above, we fixed the average of the initial intercept at zero and the average of the initial slope at one in our model for both abilities. As can be seen in this figure, vocabulary ability shows steady growth, whereas nonword repetition ability initially grows at a similarly fast rate, from wave one to wave two, but then slows down, from wave two to wave four.

{Insert Figure 2 here}

To address our research question regarding relationships between vocabulary and nonword repetition, a bivariate LCS model was fitted in which coupling parameters in both directions (from nonword repetition to vocabulary and vice versa) were estimated. This model, which is depicted in Figure 3, fitted the data well ($\chi^2(9809) = 10274.849$, RMSEA = .010, CFI = .933, TLI = .933).

{Insert Figure 3 here}

In this model, the initial intercepts (I-nwr and I-vocab in Figure 3) are zero, due to the anchoring at wave 1 for both vocabulary and nonword repetition abilities. Although these means suggest that children had the same proportions correct for vocabulary and for nonword repetition, this is not a correct interpretation, as these scores are based on arbitrarily chosen item difficulties, making direct comparisons between the two abilities not meaningful. These initial intercepts show substantial individual variation for both abilities (nonword repetition: $var = 0.932$, $SE = 0.137$, $SD = 0.97$, $p < .001$; vocabulary: $var = 0.168$, $SE = 0.025$, $SD = 0.41$, $p < .001$). The initial slope estimates (S-nwr and S-vocab in Figure 3) were fixed at 1. The variances for these estimates are modest (nonword repetition: $var = 0.080$, $SE = 0.019$, $SD = 0.28$, $p < .001$; vocabulary: $var = 0.037$, $SE = 0.007$, $SD = 0.19$, $p < .001$). There is a significant standardized covariance between the initial intercept and the initial slope for nonword repetition ($cov = 0.331$, $SE = 0.066$, $p < .001$), but not for vocabulary. The standardized covariance between the initial

intercept for vocabulary and the initial intercept for nonword repetition is significant ($cov = 0.499, SE = 0.053, p < .001$). The remaining covariances (see Figure 3: in grey) are not significant.

The wave-specific growth estimates (Δ_{nwr_t} and Δ_{vocab_t} in Figure 3) are influenced in three ways: (i) by a random initial slope, just discussed, weighted by a fixed loading of 1, (ii) by a negative weighted influence from the previous level of the ability itself (i.e., the self-feedback parameter), and (iii) by positive weighted influences from the other variable (i.e., the coupling parameter). The negative self-feedback parameters, which were held constant over subsequent waves indicate that the growth rate decreased from one wave to the next for both abilities ($\beta_{nwr_t} = -0.949, p < .001$ for nonword repetition; $\beta_{vocab_t} = -0.325, p < .001$ for vocabulary). The coupling parameters from vocabulary to nonword repetition were positive and significant at all three intervals ($\gamma_{nwr_1 \rightarrow \Delta_{vocab_2}} = 0.551, p < .001$; $\gamma_{nwr_2 \rightarrow \Delta_{vocab_3}} = 0.367, p < .011$; $\gamma_{nwr_3 \rightarrow \Delta_{vocab_4}} = 0.301, p < .001$). The coupling parameters from nonword repetition to vocabulary were positive and significant or showed a trend towards significance ($\gamma_{vocab_1 \rightarrow \Delta_{nwr_2}} = 0.123, p = .005$; $\gamma_{vocab_2 \rightarrow \Delta_{nwr_3}} = 0.172, p = .062$; $\gamma_{vocab_3 \rightarrow \Delta_{nwr_4}} = 0.219, p = .033$). Thus, these parameters indicate positive effects on the growth rate from the other ability for both abilities, from one wave to the next, albeit most clearly so for the cross-relationship from vocabulary to nonword repetition. Note, however, that, since growth of each variable is a function of three different influences which may work in different directions and can be proportional or linear, the sizes of the effects of vocabulary on nonword repetition ability and vice versa cannot simply be compared.

Taken together, these findings indicate that there are cross-lagged interrelations between both variables over time: changes in an ability from one wave to the next, over the period from two to five years, are not only dependent on the level of ability itself at the previous wave, but

also on the level of the other ability at the previous wave. Due to the proportional nature of these estimates, and the fact that the coupling parameters become larger or smaller over waves, growth is not linear (see Figure 2). Average latent vocabulary shows a substantial increase from one wave to the next compared to the variation in this ability within waves, while the average latent NWR ability shows a smaller increase as compared to the variation in this ability within waves.

In order to address whether the cross-effect of vocabulary on nonword repetition ability was stronger than vice versa, we compared the model in Figure 3 with three alternative models in which one or both cross-effects (from vocabulary to nonword repetition and vice versa) were removed (see Ferrer et al., 2007, Ferrer & McArdle, 2004 for similar procedures). This model comparison indicated that our initial model with coupling parameters in both directions showed significantly better fit (WLSM $\Delta\chi^2(3) = 13.543, p = .004$) than the model with the coupling parameters regressing vocabulary on nonword repetition only ($\chi^2(9812) = 10293.254, \text{RMSEA} = .010, \text{CFI} = .931, \text{TLI} = .930$) and better fit (WLSM $\Delta\chi^2(6) = 17.241, p = .008$) than the model containing neither of the two coupling parameters ($\chi^2(9815) = 10296.399, \text{RMSEA} = .010, \text{CFI} = .931, \text{TLI} = .930$). It did not show better fit (WLSM $\Delta\chi^2(3) = 2.753, p = .431$) than the model with the coupling parameters regressing nonword repetition on vocabulary only ($\chi^2(9812) = 10276.120, \text{RMSEA} = .010, \text{CFI} = .933, \text{TLI} = .933$).

Summarizing, the current results show positive and significant reciprocal relationships over time between nonword repetition and vocabulary during the preschool period. The model in which both cross-loadings (i.e., coupling parameters) were included provided a significantly better fit to the data than the models in which cross-loadings from vocabulary to nonword repetition or both cross-loadings were removed, which shows that there are positive bidirectional relationships between nonword repetition and vocabulary in the period from two to five years.

However, it did not show better fit than a model in which cross-loadings from nonword repetition to vocabulary were removed, suggesting that especially vocabulary skill was predictive of nonword repetition skill in the current study, rather than vice versa.

Checking the Results against Cross-Lagged and Latent Growth Models

The current findings differ from those in Melby-Lervåg et al. (2012) in showing reciprocal relationships between nonword repetition and vocabulary over time, at least in our main model with coupling parameters in both directions. One explanation of this discrepancy in results might be that we used a different type of analysis. Recent work has shown that LCS modeling may sometimes yield different results than cross-lagged models (cf. Usami, Hayes, & McArdle, 2015, 2016). To see whether our results would differ as a function of the analyses used, we conducted two additional analyses by adapting the parameter settings of our above-described LCS model: cross-lagged modeling (i.e., relating two simplex models for vocabulary and nonword repetition) and latent growth modeling (LGM), as used in Melby-Lervåg et al. (2012). Here, only a brief summary of the main results is given. For more details on the models and the analytical steps taken, see online supplemental materials.

The cross-lagged model (see Figure A1 in online supplemental materials) fitted our data well ($\chi^2(9812) = 10325.570$, RMSEA = .011, CFI = .926, TLI = .926), but not as well as the LCS model. Importantly, in this model, the cross-lagged paths from nonword repetition to vocabulary were significant across successive waves ($b = 0.091$, $p < .001$) and the same held for the cross-lagged paths from vocabulary to nonword repetition ($b = 0.397$, $p < .001$), indicating bidirectional relationships between both abilities.

The latent growth model, replicated from Melby-Lervåg et al. (2012) (see model A2 in the online supplemental materials) fitted the data reasonably well ($\chi^2(9817) = 10394.656$,

RMSEA = .011, CFI = .917, TLI = .916). However, it fitted the data less well than the LCS model, as evidenced by lower CFI/TLI values, as well as a significant outcome of a chi-square difference test for nested models (WLSM $\Delta\chi^2(8) = 73.009, p < .001$). Importantly, in this LGM model too, there were significant cross-loadings between both abilities. Specifically, the path from the intercept of nonword repetition to the slope of vocabulary was significant ($b = 0.147, p = .015$), as well as the path from the intercept of vocabulary to the slope of nonword repetition ($b = 0.037, p = .013$).

Discussion

In this study, we investigated the developmental relationships between nonword repetition and vocabulary over the preschool years in a sample of 471 Dutch-speaking children. Our aim was to test three different views on how nonword repetition and vocabulary growth may be interrelated in children's development: (i) a storage-based view, which assumes that improvements in nonword repetition drive vocabulary learning (Gathercole, 2006), (ii) a lexical restructuring view, which assumes that improved nonword repetition is the result of vocabulary growth rather than its cause (Bowey, 2001; Metsala, 1999), and (iii) a "combined" view according to which both processes play a role, resulting in reciprocal relationships between the two abilities over time (Rispen & Baker, 2012; Snowling, 2006).

In earlier longitudinal studies on this topic, outcomes of children between four and seven or eight years of age were examined, and no clear evidence was found for either view (Gathercole et al., 1992; Melby-Lervåg et al. 2012). We hypothesized that this might be due to the relatively "old age" of the participants in these studies, and predicted that, in our study on preschool children, bidirectional relationships between nonword repetition and vocabulary would be found. Specifically, based on earlier evidence which suggests that young children rely on

phonological storage as a basic learning mechanism more strongly than older children during word learning (Gathercole, 1995; Jarrold et al., 2004), we expected that nonword repetition would predict vocabulary skill in our sample of two- to five-year-olds. Moreover, we predicted that, in children this young, vocabulary would predict nonword repetition skill, based on earlier work showing that lexical restructuring is an early process (Edwards, Beckman, & Munson, 2004; Storkel, 2002; Verhagen et al., 2017).

Our results indeed indicated that there were bidirectional relationships between nonword repetition and vocabulary. Specifically, using Latent Change Score (LCS) modeling, a model in which cross-paths in both directions (i.e., from vocabulary to nonword repetition and vice versa) were estimated showed that vocabulary significantly predicted nonword repetition over time, from ages two to three, three to four, and four to five. Interestingly, this relationship became weaker over time, in line with our prediction that lexical restructuring would become less important with age. Nonword repetition showed a less clear pattern: it significantly predicted vocabulary for two out of three time intervals, from ages two to three and from ages four to five. The prediction between ages three and four only showed a trend towards significance ($p = .062$).

A comparison between this ‘full model’ and three alternative models was then performed: (i) a model in which only the coupling parameters from nonword repetition to vocabulary were included (in line with the phonological storage view), (ii) a model in which only the coupling parameters from vocabulary to nonword repetition were included (in line with the lexical restructuring view), (iii) and a model in which no coupling parameters were included. Comparing the model with all coupling parameters (representing the ‘combined view’) to the other three models, we found that the former model had a significantly better fit to the data than the model with coupling parameters from vocabulary to nonword repetition representing the

lexical restructuring view and the model without any coupling parameters. It did not show better fit than the model with coupling parameters from nonword repetition to vocabulary representing the storage-based view, however.

Taken together, these results indicate that receptive vocabulary level positively and significantly contributed to growth rate changes in nonword repetition from one wave to the next. Evidence for the reverse relationship was less strong. Significant predictive relationships were found for two out of three age intervals and a model in which only these relationships were retained, did not show significantly worse fit than the model with both cross-relationships. Thus, with respect to the storage-based view of word learning, the results are less conclusive.

To rule out that the significant cross-domain relationships found in our full model were an artifact of the type of modeling used and, for this reason, differed from the findings in earlier work, we re-analyzed our data using cross-lagged and latent growth modeling, the techniques used by Melby-Lervåg et al. (2012) on both their own data and those of Gathercole et al. (1992). These additional analyses yielded very similar results and confirmed the presence of bidirectional relationships in our data. This indicates that our findings are robust.

Our motivation for using LCS modeling in our primary analysis was that LCS modeling is better suited to assess the dynamic relationships between two growing abilities than the analytical techniques used previously (Ferrer & McArdle, 2004, 2010; Grimm, 2008; McArdle, 2009). We found the fit of the LCS model to be better than the fit of both cross-lagged and LGM models that were based on the same data. Note, however, that these more classic models differ in important respects from an LCS model, which might have led to under- or overestimations of the cross-effects (see also Berry & Willoughby, 2017). Specifically, in the LGM, the cross-effect may have been underestimated, perhaps because this model only contains cross-connections in

the global, not wave-specific, intercept and slope parameters (McArdle, 2009). The cross-lagged model, in contrast, may have overestimated the cross-effects, perhaps since no mean level structure is included (Hamaker, Kuiper, & Grasman, 2015). These ideas receive at least some tentative support by the results of Melby-Lervåg et al. (2012) who found that, in their cross-lagged model, one of the cross-relations approached significance ($p = .079$), while there was no evidence for a significant cross-relationships in their latent growth model based on the same data. However, close comparisons of these models are beyond the topic of the current study, and other studies fitting various models to the same data set as well as simulations studies are needed to clarify the precise differences between types of models (cf. Usami, Hayes, & McArdle, 2016).

Previous studies have shown that nonword repetition and vocabulary are correlated in preschoolers (e.g., Chiat & Roy, 2007; Gathercole & Adams, 1993; Hoff, Core, & Bridges, 2008; Zamuner, 2009), but, to the best of our knowledge, none of these studies have examined the directionality of these relationships in a longitudinal design. Therefore, our study is the first to assess longitudinally how, in this age group, nonword repetition and vocabulary are interrelated. The current findings present clear evidence for the lexical restructuring view according to which growth in vocabulary underlies growth in nonword repetition through better-developed phonological representations (Bowey, 2001; Metsala & Walley, 1998; Metsala, 1999). Our findings also provide suggestive evidence for a storage-based view on the relationship between nonword repetition and vocabulary, which considers nonword repetition as the driving force behind vocabulary growth at early phases of development (Baddeley, Gathercole, Papagno, 1998; Gathercole, 2006). Taken together, these findings provide at least tentative evidence for the third view outlined in the Introduction which assumes that lexical restructuring and storage-based word learning work in parallel in young children, an idea proposed – but not yet

empirically tested in longitudinal research on young children (Hoff, Core, & Bridges, 2008; Rispens & Baker, 2012; Snowling, 2006).

Our results contrast with those of Melby-Lervåg et al. (2012) and Gathercole et al. (1992) who found no clear evidence for significant predictive relationships between nonword repetition and vocabulary (or vice versa) in a sample of four- to seven-year-olds. We proposed that this may be due the age of the participants in these studies. Indeed, in our data, the strength of the predictive relationship between vocabulary and nonword repetition decreased as children grew older, in line with earlier cross-sectional studies showing that lexical restructuring is especially important in early stages of language development.

Another possible explanation of the (near) lack of significant associations between nonword repetition and vocabulary in earlier work, apart from age, is the type of nonword stimuli used. In Gathercole et al. (1992), the Children’s Test of Nonword Repetition (Gathercole & Baddeley, 1996) was used, which contains stimuli that strongly resemble existing (English) words, such as *defermication* and *voltularity*. In their replication study, Melby-Lervåg et al. (2012) used the same stimuli “adapted to the phonetic and *semantic* features of the Norwegian language” (p. 2, italics added). Hence, in both studies, nonwords contained frequent sound sequences, and even existing morphemes and word endings such as ‘-ation’ and ‘-tually’. The question arises whether the use of such items may have led children to bootstrap the nonwords from similar nonwords or even existing words and syllables in their lexicons, in children with large and smaller vocabularies alike. A number of previous, cross-sectional studies have shown that relationships with vocabulary are stronger for nonword repetition of items composed of infrequent phoneme combinations as compared to items containing highly frequent phoneme combinations (Edwards, Beckman, & Munson, 2004; Munson, Edwards, & Beckman, 2005).

Such findings have been taken as evidence that, when repeating low-probability nonwords, children cannot rely on existing language knowledge and thus are more heavily dependent on well-developed phonological representations. However, while this account may explain the lack of clear predictive effects from vocabulary to nonword repetition in earlier work, it cannot explain why no clear predictive effects from nonword repetition to vocabulary were found in Melby-Lervåg et al. (2012) and Gathercole et al. (1992). Future research could address whether the wordlikeness or phonotactic probability of the nonwords impacts on the longitudinal relationships with vocabulary in young children.

The current study used Item Response Theory (IRT)-based linkage of common items across waves to allow growth modeling of different – but partly overlapping – items across waves (McArdle et al. 2009). Through linkage of overlapping items across waves, latent ability scales for the two skills could be constructed, which served as input for the LCS model. The problem we encountered in our study of not being able to use the same test at each wave is a common one in research on cognitive development in young children, because children’s skills often develop rapidly. A common procedure to deal with this problem is to have changing tests over time and to perform regressions to predict later scores from earlier scores. However, such regression models do not allow us to assess growth at the individual level (Grimm et al., 2013; McArdle et al., 2009). The current procedure in which IRT was applied to construct latent ability scales based on overlapping items across waves, which served as input for modeling growth patterns, may be useful for future research investigating (relationships between) children’s rapidly developing skills over time.

Our findings indicate that lexical restructuring and phonological storage work in parallel, with the clearest support found for lexical restructuring. There are several steps that could be

taken to investigate this finding further. For instance, future studies could target even younger children by relating measures of phonological representations suitable for infants and toddlers (e.g., mispronunciation detection, Swingley, 2005) to vocabulary measures, to see whether lexical restructuring takes place at even earlier stages of language development. Relationships between nonword repetition and productive vocabulary could also be investigated, as there is some evidence that relationships with nonword repetition are stronger for productive than receptive vocabulary, at least in children around two years of age, perhaps due to articulatory demands playing a role in both nonword repetition and productive vocabulary (Hoff, Core, & Bridges, 2008; Stokes, Moran, & George, 2013).

Furthermore, the association could be assessed in different populations. Specifically, comparisons among typically developing children, bilingual children and children with language impairment could be made, to see whether the same developmental relationships are attested across these groups. With respect to bilingual children, vocabulary knowledge in one language is generally more limited than that of monolingual children (Bialystok, Luk, Peets & Yang, 2010; Pearson, Fernandez, Lewedeg & Oller, 1997), and phonological memory skills typically higher in their dominant language (Messer, Leseman, Boom, & Mayo, 2010), presenting an interesting case for studying how nonword repetition and vocabulary are related within bilingual children's two languages, to test language-specificity of the processes driving word learning.

Regarding children with language impairment, one would predict even closer associations between nonword repetition and vocabulary, since language-impaired children are at a less advanced stage of language development and, thus, expected to rely even more on processes of lexical restructuring and phonological storage (e.g., de Bree, Rispen & Gerrits, 2007; Rispen & Baker, 2012). Different outcomes cannot be ruled out, however. Previous

(cross-sectional) research has demonstrated that some language-impaired preschoolers may show very poor nonword repetition ability but normal to good vocabulary skill, or vice versa, despite significant associations between nonword repetition and receptive vocabulary at the group level (Chiat & Roy, 2007). What is needed, therefore, are longitudinal studies that address how nonword repetition and vocabulary are interrelated in language-impaired children as compared to typically developing children, to reveal whether the same processes drive word learning in these groups.

There are a number of limitations to our study. First, the number of items used per wave, especially for nonword repetition, was limited. This was deemed necessary, given toddlers' and preschoolers' relatively short attention spans, as well as the fact that the current tasks were part of a larger test battery, and thus had to be very short. Second, participants for this study involved a sub-sample of a larger group of participants, which had been selected, in part, on task completion at ages two and three. Since individual differences in task completion – especially at this young age – may be correlated with participant characteristics such as general intelligence and attention span, the current sample – even though still relatively large, may have been biased. In particular, the fact that children had developed sufficient language at age two years to be assessed raises the possibility that the current sample is not fully representative of the child population at large.

Notwithstanding these limitations, the current study clearly adds to our knowledge on the developmental relationships between nonword repetition and vocabulary, as it presents longitudinal data from younger children than earlier work and provides converging evidence from several statistical techniques. Specifically, the current findings suggest that phonological storage, as tapped by nonword repetition, drives vocabulary growth. They also indicate that

RECIPROCAL RELATIONSHIPS BETWEEN NONWORD REPETITION AND VOCABULARY

increases in vocabulary, in turn, result in better nonword repetition ability. As such, they present at least tentative evidence that there are positive and reciprocal relationships between the two skills, at least in early stages of linguistic development, with lexical restructuring being the more important process.

References

- Asparouhov, T., & Muthén, B., (2015). *IRT in MPLUS*, Webnote obtained from Statmodel.com
- Baddeley, A.D., Gathercole, S.E., & Papagno, C. (1998). The phonological loop as a language learning device. *Psychological Review*, *105*, 158–173.
- Berry, D., & Willoughby, M.T. (2017). On the practical interpretability of of cross-lagged panel models: Rethinking a developmental workhorse. *Child Development*, *88*, 1186-1206.
- Bialystok, E., Luk, G., Peets, K. F., & Yang, S. (2010). Receptive vocabulary differences in monolingual and bilingual children. *Bilingualism: Language and Cognition*, *13*, 525-531.
doi:[10.1017/S1366728909990423](https://doi.org/10.1017/S1366728909990423)
- de Bree, E.H., Rispens, J., & Gerrits, E. (2007). Nonword repetition in Dutch children with (a risk of) dyslexia and SLI. *Clinical Linguistics & Phonetics*, *21*, 935–944.
doi:[10.1080/02699200701576892](https://doi.org/10.1080/02699200701576892)
- Bowey, J.A. (2001). Nonword repetition and young children’s receptive vocabulary: A longitudinal study. *Applied Psycholinguistics*, *22*, 441–469.
- Butner, J.E., Berg, C.A., Baucom, B.R., & Wiebe, D.J. (2014). Modeling coordination in multiple simultaneous latent change scores. *Multivariate Behavioral Research*, *49*, 554-570. doi:[10.1080/00273171.2014.934321](https://doi.org/10.1080/00273171.2014.934321)
- Cheung, H. (1996). Nonword span as a unique predictor of second-language vocabulary learning. *Developmental Psychology*, *32*, 867–873.
- Daelemans, W., Gillis, S., & Durieux, G. (1994). The acquisition of stress: A data-oriented approach. *Computational Linguistics*, *20*, 421–451.

- Duff, F. J., Reen, G., Plunkett, K., & Nation, K. (2015). Do infant vocabulary skills predict school-age language and literacy outcomes? *Journal of Child Psychology and Psychiatry*, *56*, 848–856. doi:10.1111/jcpp.12378.
- Dunn, L., & Dunn, L.M. (2005). *Peabody Picture Vocabulary Test-III-NL. Nederlandse versie door Liesbeth Schlichting* [Dutch version by Liesbeth Schlichting]. Amsterdam: Harcourt Assessment B.V.
- Embretson, E.E., & Reise, S.P. (2000). *Item response theory for psychologists*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Edwards, J., Beckman, M.E., & Munson, B. (2004). The interaction between vocabulary size and phonotactic probability effects on children's production accuracy and fluency in nonword repetition. *Journal of Speech, Language, and Hearing Research*, *47*, 421–436. doi:10.1017/S0305000910000450.
- Ferrer, E., & McArdle, J.J. (2004). An experimental basis of dynamic hypotheses about cognitive abilities and achievement from childhood to early adulthood. *Developmental Psychology*, *40*, 935–952. doi:10.1037/0012-1649.40.6.935
- Ferrer, E., McArdle, J.J., Shaywitz, B.A., Holahan, J.M., Marchione, K., & Shaywitz, S.E. (2007). Longitudinal models of developmental dynamics between reading and cognition from childhood to adolescence. *Developmental Psychology*, *43*, 1460–1473. doi:10.1037/0012-1649.43.6.1460
- Ferrer, E., & McArdle, J.J. (2010). Longitudinal modeling of developmental changes in psychological research. *Current Directions in Psychological Science*, *19*, 149–154. doi:https://doi.org/10.1177/0963721410370300

- Fricke, S., Bowyer-Crane, C., Haley, A.J., Hulme, C., & Snowling, M. (2013). Efficacy of language intervention in the early years. *Journal of Child Psychology and Psychiatry*, *54*, 280–290. doi: 10.1111/jcpp.12010.
- 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*, 336–360.
- Gathercole, S.E., Willis, C.S., Emslie, H.C, & Baddeley, A.D. (1992). Phonological memory and vocabulary development during the early school years: A longitudinal study. *Developmental Psychology*, *28*, 887–898.
- Gathercole, S.E. (1995). Is nonword repetition a test of phonological memory or long-term knowledge? It all depends on the nonwords. *Memory and Cognition*, *23*, 83–94.
- Gathercole, S.E., & Baddeley, A.D. (1996). *The children's test of nonword repetition*. London: The Psychological Corporation.
- Gathercole, S.E. (2006). Nonword repetition and word learning: The nature of the relationship. *Applied Psycholinguistics*, *27*, 513–543. doi: <https://doi.org/10.1017/S0142716406060383>
- Gathercole, S.E., Willis, C., Emslie, H., & Baddeley, A.D. (1992). Phonological memory and vocabulary development during the early school years: A longitudinal study. *Developmental Psychology*, *28*, 887–898.
- Gathercole, S.E., Tiffany, C., Briscoe, J., Thorn, A.S.C., & The ALSPAC Team. (2005). Developmental consequences of phonological loop deficits during early childhood: A longitudinal study. *Journal of Child Psychology and Psychiatry*. *46*, 598–611. doi:<https://doi.org/10.1111/j.1469-7610.2004.00379.x>

- Grimm, K.J. (2008). Longitudinal associations between reading and mathematics achievement. *Developmental Neuropsychology*, *33*, 410–426. doi:10.1080/87565640801982486.
- Hamaker, E.L., Kuiper, R.M., & Grasman, R.P. (2015). A critique of the cross-lagged panel model. *Psychological Methods*, *20*, 102–116. doi:10.1037/a0038889.
- Hoff, E., Core, C., & Bridges, K. (2008). Nonword repetition assesses phonological memory and is related to vocabulary development in one-year-olds. *Journal of Child Language*, *35*, 903–916. doi:10.1017/S0305000908008751.
- Huttenlocher, J., Haight, W., Bryk, A., Seltzer, M., & Lyons, T. (1991). Early vocabulary growth: relation to language input and gender. *Developmental Psychology*, *27*, 236–248.
- Jarrold, C., Baddeley, A.D., Hewes, A.K., Leeke, T.C., & Phillips, C.E. (2004). What links verbal short-term memory performance and vocabulary level? Evidence of changing relationships among individuals with learning disability. *Journal of Memory and Language*, *50*, 134–148. doi:<https://doi.org/10.1016/j.jml.2003.10.004>
- Jusczyk, P.W., & Hohne, E.A. (1997). Infants' memory for spoken words. *Science*, *277*, 1984–1986.
- Lee, J. (2011). Size matters: Early vocabulary as a predictor of language and literacy competence. *Applied Psycholinguistics*, *32*, 69–92. doi:10.1017/S0142716410000299.
- Little, T.D. (2013). *Longitudinal structural equation modeling*. London/New York: The Guilford Press.
- Marulis, L.M., & Neuman, S.B. (2010). The effects of vocabulary intervention on young children's word learning: A meta-analysis. *Review of Educational Research*, *3*, 300–355. doi: 10.3102/0034654310377087

- Masoura[EdB1], E.V., & Gathercole, S.E. (2005). Phonological short-term memory skills and new word learning in young Greek children. *Memory*, *13*(3-4), 422–429.
- McArdle, J.J. (2001). A latent difference score approach to longitudinal dynamic structural analysis. In R. Cudeck, S., du Toit, & D. Sorbom (Eds.), *Structural equation modeling: Present and future* (pp. 342–380). Lincolnwood, IL: Scientific Software International.
- McArdle, J.J. (2009). Latent variable modeling of differences and changes with longitudinal data. *Annual Review of Psychology*, *60*, 577–605.
doi:10.1146/annurev.psych.60.110707.163612.
- McArdle, J.J., Grimm, J.K., Hamagami, F., Bowles, R.P. & Meredith, W. (2009). Modeling life-span growth curves of cognition using longitudinal data with multiple samples and changing scales of measurement. *Psychological Methods*, *14*, 126–149.
doi:10.1037/a0015857.
- Melby-Lervåg, M., Lervåg, A., Halaas Lyster, S-A., Klem, M., Hagtvet, B., & Hulme, C. (2012). Nonword repetition ability does not have a causal influence on children’s vocabulary development. *Psychological Science*, *24*, 1–7. doi:10.1073/pnas.93.24.13468.
- Messer, M.H., Leseman, P.P.M., Boom, J., & Mayo, A.Y. (2010). Phonotactic probability effect in nonword recall and its relationship with vocabulary in monolingual and bilingual preschoolers. *Journal of Experimental Child Psychology*, *105*, 306-323.
doi:10.1016/j.jecp.2009.12.006.
- Metsala, J.L., & Walley, A.C. (1998). Spoken vocabulary growth and the segmental restructuring of lexical representations: Precursors to phonemic awareness and early reading ability. In J.L. Metsala & L.C. Ehri (Eds.), *Word recognition in beginning literacy* (pp. 89–120). Mahwah, New Jersey: Erlbaum.

- Metsala, J.L. (1999). Young children's phonological awareness and nonword repetition as a function of vocabulary development. *Journal of Educational Psychology, 91*, 3–19.
- Mulder, H., Hoofs, H., Verhagen, J., Van der Veen, I., & Leseman, P.P.M. (2014). Psychometric properties and convergent validity of an executive function test battery for 2-year-olds using item response theory. *Frontiers in Psychology, section Developmental Psychology, 5*, 733. doi: 10.3389/fpsyg.2014.00733.
- Mulder, H., Verhagen, J., van der Ven, S., Slot, P.L., & Leseman, P.P.M. (2017). Early executive function at age two predicts emergent mathematics and literacy at age five. *Frontiers in Psychology, 8*: 1706.
- 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 language impairments. *Journal of Speech, Language, and Hearing Research, 48*, 1033–1047. doi:[10.1044/1092-4388\(2005/072\)](https://doi.org/10.1044/1092-4388(2005/072))
- Muthén, L.K., & Muthén, B.O. (1998-2012). *Mplus User's Guide. Seventh Edition*. Los Angeles, CA: Muthén & Muthén.
- Pearson, B.Z., Fernández, S.C., Lewedeg, V., & Oller, D.K. (1997). The relation of input factors to lexical learning by bilingual infants. *Applied Psycholinguistics, 18*, 41–58.
- Quinn, J.M., Wagner, R.K., Petscher, Y., & Lopez, D. (2015). Developmental relations between vocabulary knowledge and reading comprehension: A latent change score modeling study. *Child Development, 86*, 159-175. doi:10.1111/cdev.12292
- Rispens, J., & Baker, A. (2012). Nonword repetition: the relative contributions of phonological short-term memory and phonological representations in children with language and

- hearing impairment. *Journal of Speech, Language, and Hearing Research*, 55, 683–694.
doi:10.1044/1092-4388(2011/10-0263
- Rowe, M.L., Raudenbusch, S.W., & Goldin-Meadow, S. (2012). The pace of vocabulary growth helps predict later vocabulary skill. *Child Development*, 83, 508–525. doi:10.1111/j.1467-8624.2011.01710.x.
- Snowling, M.J. (2006). Nonword repetition and language learning disorders: A developmental contingency framework. *Applied Psycholinguistics*, 27, 588–91.
doi:10.1017/S0142716406060504.
- Statline, (2018). *Bevolking: onderwijsniveau; geslacht, leeftijd en migratieachtergrond* [Population: educational level, gender, age, and migration background]. Retrieved 20 February 2018, from
<http://statline.cbs.nl/StatWeb/publication/?VW=T&DM=SLnl&PA=82275NED&LA=nl>
- Stokes, S.F., & Klee, T. (2009). Factors influencing vocabulary development in two-year-old children. *Journal of Child Psychology and Psychiatry*. 50, 498–505. doi:10.1111/j.1469-7610.2008.01991.x.
- Stokes, S.F., Moran, C., & George, A. (2013). Nonword repetition and vocabulary use in toddlers. *Topics in Language Disorders*, 33, 224–237. doi:10.1097/TLD.0b013e31829d038c
- Storch, S.A., & Whitehurst, G.J., (2002). Oral language and code-related precursors to reading: Evidence from a longitudinal structural model. *Developmental Psychology*, 38, 934-947.
- Storkel, H.L. (2002). Restructuring similarity neighborhoods in the developing mental lexicon. *Journal of Child Language*, 29, 251–274.

- Swingley, D. (2005). 11-month-olds' knowledge of how familiar words sound. *Developmental Science*, 8, 432–443. doi:10.1111/j.1467-7687.2005.00432.
- Usami, S., Hayes, T., & McArdle, J.J. (2015). On the mathematical relationship between latent change score and autoregressive cross-lagged factor approaches: Cautions for inferring causal relationships between variables. *Multivariate Behavioral Research*, 50, 676–687. doi:10.1080/00273171.2015.1079696.
- Usami, S., Hayes, T., & McArdle, J.J. (2016). Inferring longitudinal relationships between variables: Model selection between the latent change score and autoregressive cross-lagged factor models. *Structural Equation Modeling: A Multidisciplinary Journal*, 23, 331–342. doi:10.1080/00273171.2015.1079696.
- Verhagen, J., de Bree, E.H., Mulder, H., & Leseman, P.P.M. (2017). Effects of vocabulary and phonotactic probability on 2-year-olds' nonword repetition. *Journal of Psycholinguistic Research*, 46, 507-524.
- Walley, A.C., Metsala, J.L., & Garlock, V.M. (2003). Spoken vocabulary growth: Its role in the development of phoneme awareness and reading ability. *Reading and Writing*, 16, 5–20. doi:<https://doi.org/10.1023/A:1021789804977>

RECIPROCAL RELATIONSHIPS BETWEEN NONWORD REPETITION AND VOCABULARY

Table 1

Sample Characteristics and Task Completion Rates per Study Wave

	Age (years; months)	Gender	Vocabulary task ^a	Nonword repetition task ^a
	<i>M (SD)</i>	% girls	<i>N (%)</i>	<i>N (%)</i>
Wave 1	2;4 (0;3)	53.0	409 (86.8%)	401 (85.1%)
Wave 2	3;6 (0;3)	52.6	471(100 %)	470 (99.8%)
Wave 3	4;10 (0;2)	52.9	452 (96.0%)	452 (96.0%)
Wave 4	5;10 (0;2)	53.0	444 (94.3%)	443 (94.1%)

^a Number (percentage) of children out of 471 children who completed at least one item.

RECIPROCAL RELATIONSHIPS BETWEEN NONWORD REPETITION AND VOCABULARY

Table 2

Overview of the Unique and Overlapping Items across Study Waves for Nonword Repetition

Wave 1	Wave 2	Wave 3	Wave 4
Items 1 – 6			
Items 7 – 12	Items 7 – 12		
	Items 13 – 18	Items 13 – 18	
		Items 19 – 24	Items 19 – 24
			Items 25 – 30

RECIPROCAL RELATIONSHIPS BETWEEN NONWORD REPETITION AND VOCABULARY

Table 3

Overview of the Unique and Overlapping Items across Study Waves for Vocabulary

Wave 1	Wave 2	Wave 3	Wave 4
Items 1 – 16			
Items 17 – 24	Items 17 – 24		
	Items 25 – 40	Items 25 – 40	
		Items 41 – 52	Items 42 – 52
			Items 53 – 69

Note. One item (nr. 41) was not repeated from wave 3 to wave 4, because of very low scores at wave 3.

RECIPROCAL RELATIONSHIPS BETWEEN NONWORD REPETITION AND VOCABULARY

Table 4

Descriptive Statistics for Nonword Repetition and Vocabulary for Overlapping Items per Study

Wave (Proportion Correct)

	Wave 1		Wave 2		Wave 3		Wave 4	
Items	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
<i>Nonword repetition</i>								
1-6	.35	.34	-	-	-	-	-	-
7-12	.22	.29	.57	.28	-	-	-	-
13-18	-	-	.37	.27	.57	.27	-	-
19-24	-	-	-	-	.41	.27	.50	.30
25-30	-	-	-	-	-	-	.21	.24
<i>Vocabulary</i>								
1-16	.64	.31	-	-	-	-	-	-
17-24	.40	.29	.83	.18	-	-	-	-
25-40	-	-	.60	.18	.82	.20	-	-
41-52	-	-	-	-	.56	.22	.74	.25
53-69	-	-	-	-	-	-	.46	.18

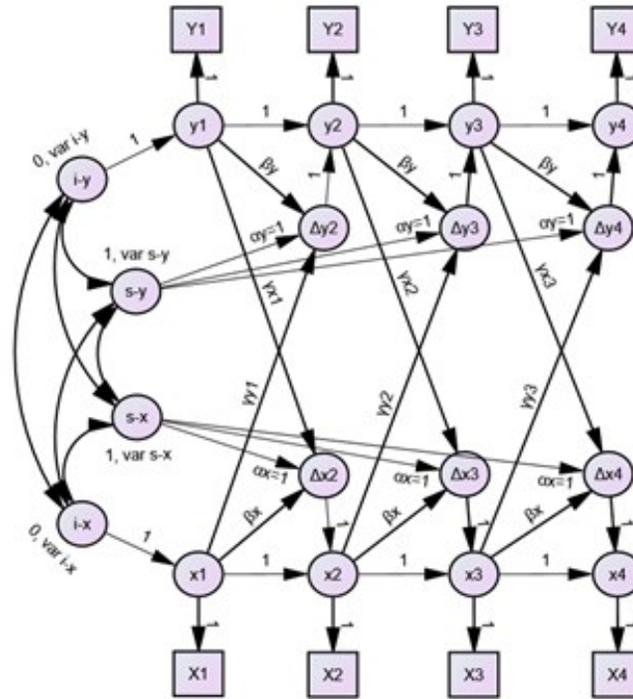


Figure 1. Conceptual model of a bivariate Latent Change Score model.

X_t and Y_t = observed scores at time t ; x_t and y_t = latent true scores at time t ; Δx_t and Δy_t = latent changes at time t ; $i-x$ and $i-y$ = initial intercepts (with averages fixed to zero, variances free); $s-x$ and $s-y$ = initial slopes (with averages fixed to one, variances free). Covariances (double arrows), β_x and β_y = self-feedback parameters and γ_{x_t} and γ_{y_t} = coupling parameters are freely estimated; α_x and α_y = loading parameters regressing change on the initial slopes and all other paths are fixed to one.

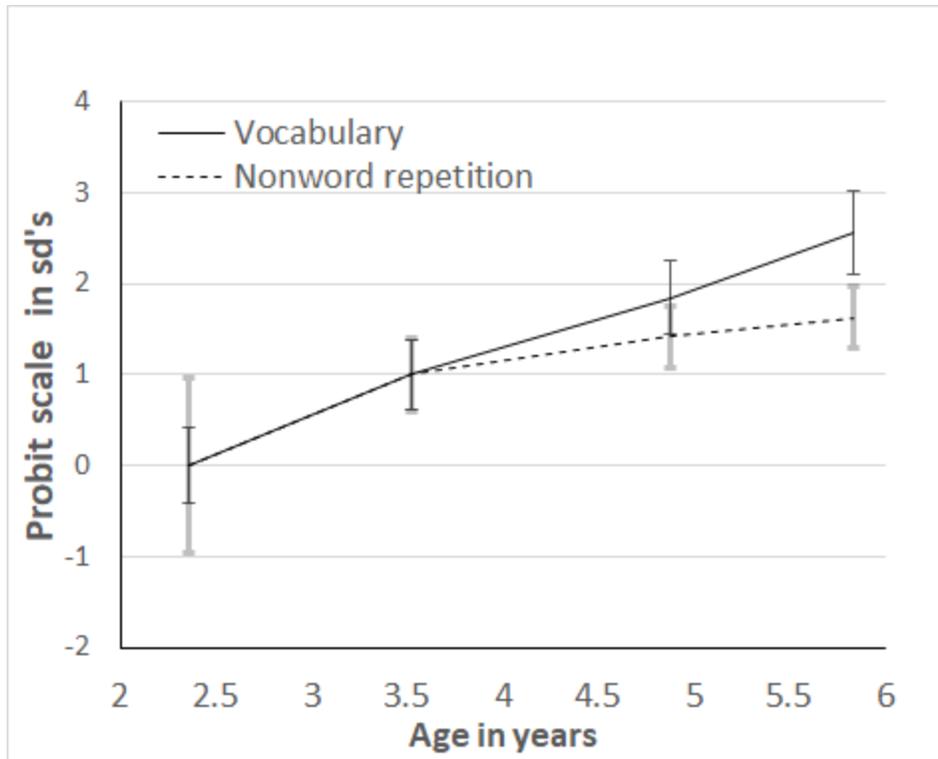


Figure 2. Average growth of nonword repetition and vocabulary abilities from two to five years.

Error bars indicate an interval of one standard deviation above or below the estimated person abilities. (These standard deviations for the person abilities should not be confused with the standard deviations that refer to the differences between person ability and item difficulty, indicated on the Y-axis).

RECIPROCAL RELATIONSHIPS BETWEEN NONWORD REPETITION AND VOCABULARY

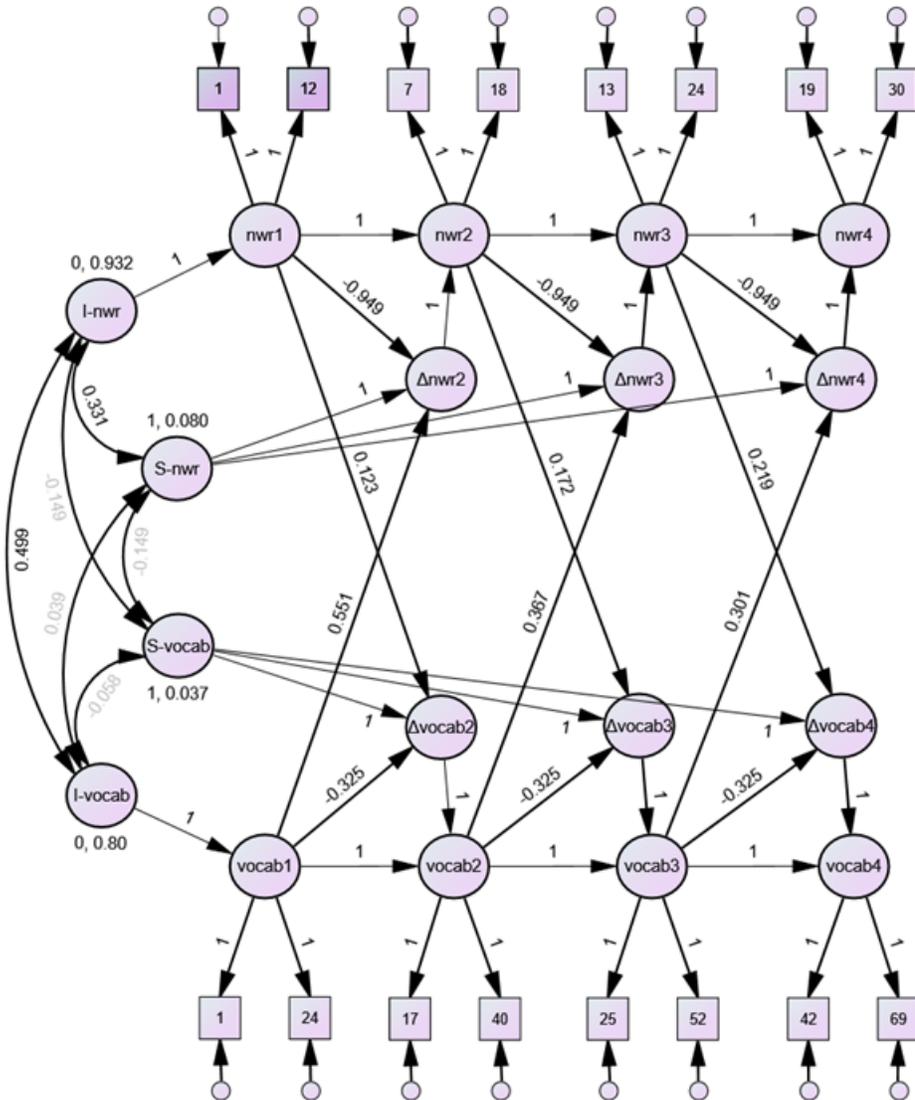


Figure 3. Results of bivariate Latent Change Score Model for nonword repetition and vocabulary ability from two to five years. For latent variables (large circles) unstandardized means and variances are presented, separated by a comma. For covariances (double-sided arrows), standardized estimates (correlations) are presented; grey values are not significant. For regressions (single-sided arrows), unstandardized estimates are presented. For observed variables (squares) only the first and the last item per factor are shown; all item loadings were fixed to one; all item thresholds and item residual variances (small circles) were estimated.

Supplemental Materials

I. Cross-Lagged and Latent Growth Models

To obtain a cross-lagged model that resembled the cross-lagged models used by Melby-Lervåg et al. (2012), our LCS model described above was adapted in two ways. First, slopes were deactivated in the model (i.e., fixed to zero) and the paths between an ability and this same ability at the previous wave were freely estimated. Second, all paths from nwr_t to Δnwr_{t+1} were removed (i.e., fixed to zero). The resulting model, depicted in Figure 4, fitted our data well ($\chi^2(9812) = 10325.570$, RMSEA = .011, CFI = .926, TLI = .926), but not as well as the LCS model.

RECIPROCAL RELATIONSHIPS BETWEEN NONWORD REPETITION AND VOCABULARY

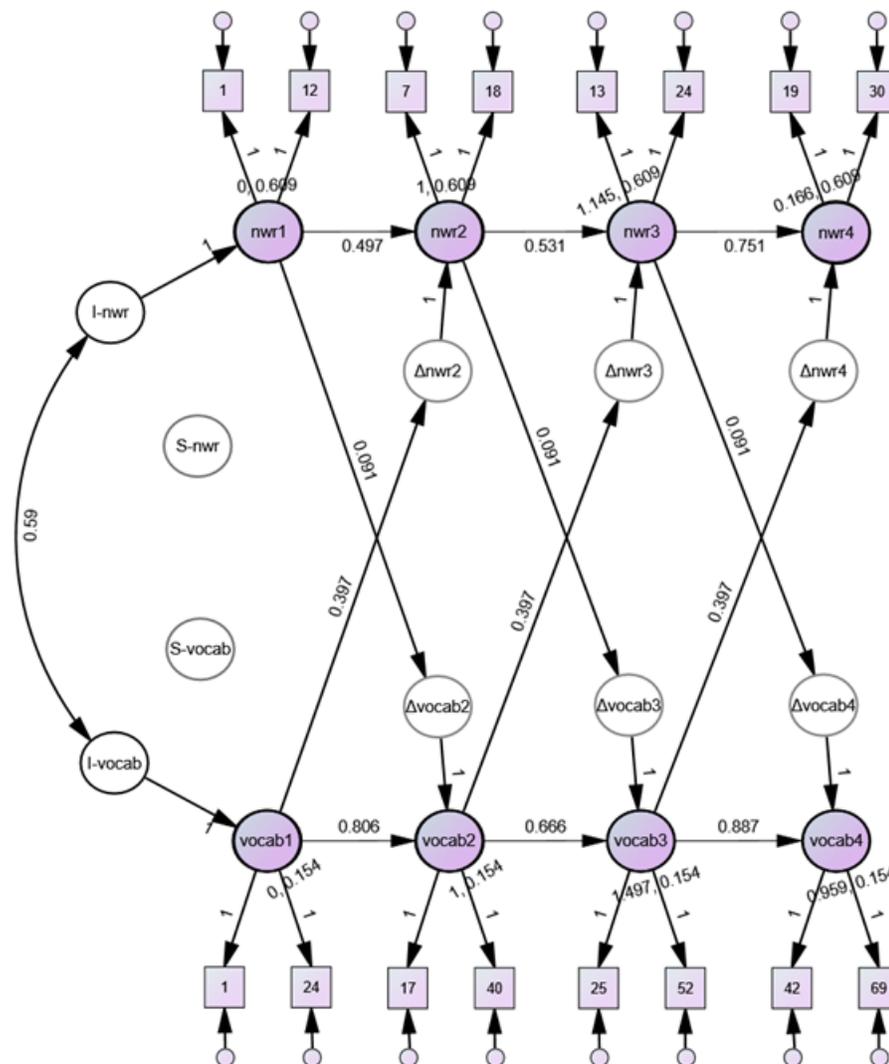


Figure 4. Results of bivariate cross-lagged model for nonword repetition and vocabulary ability from two to five years. For latent variables (grey circles) unstandardized means and variances are presented, separated by a comma. For the covariance (double-sided arrow), a standardized estimate is presented. Dummy variables (transparent circles) are provided to facilitate comparison with Figure 3, but have no effects. For regressions (single-sided arrows), unstandardized estimates are presented. For observed variables (squares) only the first and the last item per factor are shown; all item loadings were fixed to one; all item thresholds and item residual variances (small circles) were estimated.

This model shows positive and significant cross-lagged loadings between both abilities across successive waves, from nonword repetition to vocabulary ($b = 0.091, p < .001$) and from vocabulary to nonword repetition ($b = 0.397, p < .001$).

In our second analysis, a latent growth model was fitted to the data, as in Melby-Lervåg et al. (2012). To this aim, all paths in the original LCS model from x_t to Δx_{t+1} and from y to Δy as well as all coupling parameters from x_t to Δy_{t+1} and from y_t to Δx_{t+1} were fixed to zero. The resulting model, depicted in Figure 5, is formally equivalent to a standard LGM, modeling linear developmental trajectories for each participant, although it was parameterized slightly differently.

The LGM model fitted the data reasonably well ($\chi^2(9817) = 10394.656$, RMSEA = .011, CFI = .917, TLI = .916), but less so than the LCS model, as evidenced by lower CFI/TLI values as well as a significant outcome of a WLSM-based corrected, chi-square difference test for nested models (WLSM $\Delta\chi^2(8) = 73.009$, $p < .001$). Importantly, moreover, as shown in Figure 5, there were significant cross-loadings between both abilities. Specifically, the path from the intercept of nonword repetition to the slope of vocabulary was significant ($b = 0.147$, $p = .015$), as well as the path from the intercept of vocabulary to the slope of nonword repetition ($b = 0.037$, $p = .013$), indicating bidirectional relationships between these abilities.

II. Illustrations of Change Equations in LCS Model

The growth of each variable in a LCS model is a function of three different influences which may work in different directions and can be proportional or linear. To illustrate this in more detail, the average estimated latent abilities from wave one to four, as depicted in Figure 2, can be calculated based on the relevant parameters given in Figure 3.

Specifically, equations can be presented, which are based on a number of values that can be read from Figure 3. The average for nwr_1 is 0 [$i-nwr * 1$], the average for $vocab_1$ is 0, the average for nwr_2 is 1 unit, the average for $vocab_2$ is 1 unit. Then, to illustrate how the average for nwr_3 comes about, the equation is as follows: 1 unit inherited from wave two, plus growth for wave three (i.e., $(1 + -0.949 * 1$ [=self-feedback parameter * nonword repetition score at the previous wave]) + $(-0.367 * 1$ [=coupling parameter * vocabulary score of the previous wave]) = a total positive change of 1.418 units). Similarly, the average for $vocab_3$ is as follows: 1 unit inherited from wave two, plus growth for wave three (i.e., $(1 + -0.325 * 1$ [self-feedback parameter * nonword repetition score at the previous wave] + $(-0.172 * 1$ [=coupling parameter * vocabulary scores of the previous wave]) = a total positive change of 1.847 units). The average for nwr_4 is calculated as follows: 1.418 units inherited from wave two, plus growth for wave four (is 1 [from the initial slope], plus $-0.949 * 1.418$ [=the self-feedback parameter * the nonword repetition score at the previous wave], plus $0.301 * 1.847$ [=the coupling parameter * the vocabulary scores of the previous wave]), yielding a total positive change of 1.63 units. The average for $vocab_4$ is calculated as: 1.847 units inherited from wave two, plus growth for wave four (is 1 [from the initial slope], plus $-0.325 * 1.847$ [=the self-feedback parameter * the nonword repetition score at the previous wave], plus $0.219 * 1.418$ [=the coupling parameter * the vocabulary scores of the

RECIPROCAL RELATIONSHIPS BETWEEN NONWORD REPETITION AND VOCABULARY

previous wave]), yielding a total positive change of 2.55 units. These are simple computations, because the initial averages of slope and intercepts were 0 and 1. Note, however, that, at the individual level, participants typically deviated from these averages, so that computations were actually more complex, taking this variation into account.