# Individual differences in lexical processing efficiency and vocabulary in toddlers: A longitudinal investigation

Seamus Donnelly [a,b,*], Evan Kidd [a,b,c]

[a] *The Australian National University, Canberra, ACT 0200, Australia*
[b] *ARC Centre of Excellence for Dynamics of Language, Acton, ACT 2601, Australia*
[c] *Max Planck Institute for Psycholinguistics, 6525 XD Nijmegen, the Netherlands*

## ARTICLE INFO

## ABSTRACT

Research on infants' online lexical processing by Fernald, Perfors, and Marchman (2006) revealed substantial individual differences that are related to vocabulary development, such that infants with better *lexical processing efficiency* show greater vocabulary growth across time. Although it is clear that individual differences in lexical processing efficiency exist and are meaningful, the theoretical nature of lexical processing efficiency and its relation to vocabulary size is less clear. In the current study, we asked two questions: (a) Is lexical processing efficiency better conceptualized as a *central processing capacity* or as an *emergent capacity* reflecting a collection of word-specific capacities? and (b) Is there evidence for a causal role for lexical processing efficiency in early vocabulary development? In the study, 120 infants were tested on a measure of lexical processing at 18, 21, and 24 months, and their vocabulary was measured via parent report. Structural equation modeling of the 18-month time point data revealed that both theoretical constructs represented in the first question above (a) fit the data. A set of regression analyses on the longitudinal data revealed little evidence for a causal effect of lexical processing on vocabulary but revealed a significant effect of vocabulary size on lexical processing

---

* Corresponding author at: The Australian National University, Canberra, ACT 0200, Australia.
  *E-mail address:* seamus.donnelly@anu.edu.au (S. Donnelly).

efficiency early in development. Overall, the results suggest that lexical processing efficiency is a stable construct in infancy that may reflect the structure of the developing lexicon.

## Introduction

Language development is characterized by significant yet stable *individual differences* (e.g., Bates, Dale, & Thal, 1995; Bornstein, Hahn, & Putnick, 2016), which are measureable and developmentally significant from early in life (e.g., Brito, Fifer, Myers, Elliott, & Noble, 2016; Cristia, Seidl, Junge, Soderstrom, & Hagoort, 2014). One domain of language that is subject to significant individual differences is the lexicon, with large variability observed in both the growth rate and size of vocabulary across the life span. Explaining this variability is of important theoretical and practical significance. Theoretically, identifying sources of variation can reveal the mechanisms underlying development (Bates, Bretherton, & Snyder, 1988; Kidd, Donnelly, & Christiansen, 2018). In practical terms, early vocabulary is a strong predictor of language development in other domains (e.g., grammar) and of important outcomes (e.g., educational performance and attainment). Thus, understanding of the sources of early variability enables the potential early detection and remediation of developmental delay. In the current article, we report on an investigation of individual differences in infants' online lexical processing efficiency and their longitudinal relationship with vocabulary development from 18 to 24 months, a period in which there is substantial vocabulary growth and reorganization in the linguistic system.

Perhaps the most common measure of children's early *lexical processing efficiency* is the Looking While Listening (LWL) task (Fernald, Zangl, Portillo, & Marchman, 2008). In this task, children see two images side by side (a target and a distracter) and hear sentences such as "Look at the bird." The dependent measure of interest is children's efficiency in recognizing the target word. This can be measured in two ways: the proportion of time spent looking at the target image after hearing the noun and the time it takes children to look toward the target image on the trials in which they were looking at the distracter image while the target word was playing, namely their reaction time (RT).

The LWL task appears to capture developmentally stable and meaningful individual differences. RTs and accuracy across time are significantly correlated (Fernald, Perfors, & Marchman, 2006; Lany, Giglio, & Oswald, 2018) and, as would be expected, RTs are faster for older children than for younger children in both English (Fernald et al., 2006) and Spanish (Hurtado, Marchman, & Fernald, 2008). Importantly, individual differences in LWL are related to other measures of linguistic competence measured concurrently, retrospectively, and prospectively. Several studies have reported significant correlations between LWL measures and vocabulary size concurrently in infants (Fernald et al., 2006; Fernald, Marchman, & Weisleder, 2013), and under some conditions this relationship can be observed by as young as 12 months (Lany et al., 2018). Moreover, in some circumstances, infants with comparatively efficient lexical processing learn more words in a word-learning task than those with less efficient lexical processing (Lany, 2018). Retrospectively, individual differences in LWL at 24 months predict size of vocabulary, grammatical complexity, and rate of growth in vocabulary over the first year (Fernald et al., 2006). Prospectively, individual differences in LWL predict vocabulary growth (Fernald & Marchman, 2012; Lany et al., 2018) and even broader expressive language abilities 6 years later (Marchman & Fernald, 2012).

Although the validity of the LWL task and its relation to vocabulary development are clear, the theoretical nature of the task is less so, as is the direction of the developmental relationship between lexical processing efficiency and vocabulary. In the current study, we addressed these two issues, which we discuss in turn.

*The nature of individual differences in lexical processing efficiency*

Lexical processing efficiency is operationalized either as the mean RT on all trials in which the participant was looking to the distracter image prior to the onset of the target word or as the proportion of time the infant looks as the target compared with the distractor image. Because these measures are calculated over many words, they could conceivably be interpreted in two different ways. It is, for example, possible that average RTs or proportions on the LWL task reflect some central lexical or cognitive processing speed on which individuals vary and that applies to all known words (hereafter, the *central capacity* account). Such a proposal is consistent with arguments that domain-general processing speed plays a fundamental role in the development of higher-order cognitive processes (Kail & Salthouse, 1994), and adult psycholinguistic research that has shown nonlinguistic processing capacity predicts lexical processing (e.g., Hintz et al., 2019), although in the current case the capacity need not be domain general. On the other hand, it is plausible that average RTs reflect a collection of word-specific processing speeds (hereafter, the *emergent capacity* account). This approach captures the fact that many item-level features of words predict speed of lexical access (e.g., frequency, imageability, age of acquisition), which in some cases interacts with vocabulary knowledge (e.g., Mainz, Shao, Brysbaert, & Meyer, 2017). An emergent capacity account is consistent with domain-specific theoretical approaches to automization, where increased experience with specific events automatizes (i.e., speeds up) processing (e.g., Logan, 1988).

These two accounts can in principle be statistically disentangled because they correspond to different classes of structural equation models: effects-indicator and causal-indicator models (also called reflective and formative measurement models). The central capacity account corresponds to an effects-indicator model. Effects-indicator models assume that covariation in observed variables is caused by variation in some latent variable (i.e., in our case a central processing capacity). Observed variables are modeled as a system of linear regression equations, each of which includes the latent variable as a predictor and separate error terms. In the context of the LWL task, mean RTs for individual words would be modeled as shown in Fig. 1A, which illustrates the assumptions of an effects-indicator model. The direction of the paths indicates that variation in the latent variable (e.g., a centralized lexical processing capacity) causes variation in the indicator variables (e.g., knowledge of words). Because of this, any covariation in the indicator variables is assumed to be due to their shared cause, the latent variable. Moreover, adding or removing indicator variables would not change the interpretation of the latent variable; they are exchangeable indicators of a preexisting theoretical entity (thus, lexical processing speed should, all things being equal, be roughly equivalent for different words).

It is common in psychology to conceptualize constructs in this manner. One example is IQ, where it is traditionally assumed that variation in IQ test scores is caused by *g* (as well as test-specific error). Therefore, increasing *g* would increase the scores on individual IQ tests, but increasing the error components of each individual test would not affect *g*. Moreover, adding an additional IQ test to a preexisting battery would not change the interpretation of *g*.

Treating observed variables as effects indicators is common in psychology, but it is not the only possibility. We could also view the observed variables as causing the latent variable. A common example of this is socioeconomic status (SES). Although this construct may be measured by mother's education, father's education, and number of books in the home (to identify only a few examples), it is difficult to conceive of these variables as effects of some underlying SES dimension; rather, those variables define, or cause, SES. This conceptualization would lead to a different latent variable model such as the one in Fig. 1B. This causal-indicator model assumes that RTs for each word combine to create some general processing speed; that is, lexical processing speed is an *emergent* property of knowledge of individual words. This model corresponds to a single regression equation in which the latent variable is an outcome that has its own error term. In this case, the observed variables are called causal indicators as opposed to effect indicators. Because they are causally prior to the latent variable, adding or removing causal-indicator variables would change the definition of the latent variable (thus, different words would result in different processing speeds). Moreover, this conceptualization makes no assumptions about the correlations between the causal indicators. This model corresponds to the emergent capacity account.
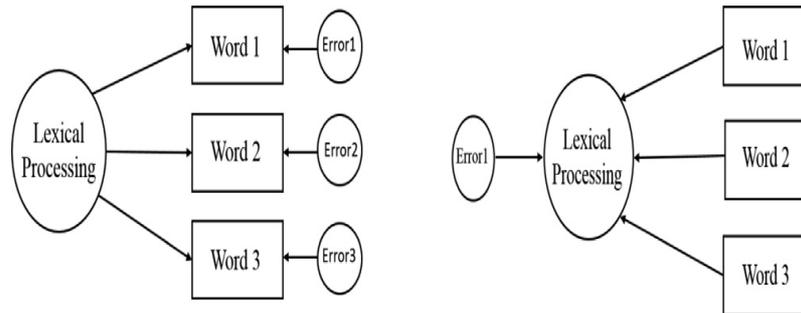
**Fig. 1.** (A) Schematic of effects-indicator model. (B) Schematic of causal-indicator model.

The merits of the distinction between these two sorts of models are still debated in the psychometric community (Bainter & Bollen, 2014; Howell, 2014; Markus, 2014; McCoach & Kenny, 2014; Wang, Engelhard, & Lu, 2014; West & Grimm, 2014; Widaman, 2014). In particular, this debate centers on whether the latent variables in causal-indicator models can be described as a measurement. However, there is no disagreement that effects-indicator models, such as factor analysis, assume latent variables that are causally prior to observed variables or that a set of observed variables can be causally prior to latent variables within a structural equation model. Therefore, because the distinction between these two types of models maps onto a clear conceptual distinction within developmental psycholinguistics, comparing the fit of these models with developmental psycholinguistic data is warranted, regardless of whether they are considered measurement models.[1]

In fact, this distinction between causal-indicator and effects-indicator models has been employed profitably within other areas in developmental psychology. For example, Willoughby and colleagues suggested that batteries of executive functioning (EF) tasks might be better modeled as causal indicators rather than effect indicators (Willoughby, Blair, & Family Life Project Investigators, 2016; Willoughby, Holochwost, Blanton, & Blair, 2014). In a reanalysis of nine influential studies of EF, Willoughby et al. (2014) found that four of the nine studies failed a vanishing tetrad test, a test of model fit often used to determine whether effects indicators are appropriate. In a subsequent study, Willoughby et al. (2016) fit causal-indicator and effects-indicator models to EF data from a longitudinal study from 3 to 5 years of age. At 3 and 4 years a nested vanishing tetrad test preferred the causal-indicator model to the effects-indicator model, and at 5 years there was a marginally significant trend favoring the causal-indicator model. The two models yielded radically different inferences about the developmental stability of EF. The effects-indicator model yielded 2-week and 2-year test–retest reliabilities that were nearly equal ($r = .99$ and $r = .96$, respectively), whereas the causal-indicator model yielded much larger 2-week than 2-year test–retest reliabilities ($r = .76$ and $r = .32$, respectively). In other words, inferences about the developmental stability were greatly influenced by the choice of model; effects indicators suggested very little developmental variability in EF from 3 to 5 years, whereas causal indicators suggested a great deal of variability. However, path coefficients from the latent variable to related variables (intelligence, academic achievement, and attention-deficit/hyperactivity disorder [ADHD] symptoms) were approximately equal in magnitude across the two models, indicating that inferences about the relationship between EF and other variables were not affected by the choice of model.

Interestingly, all the research on individual differences in the LWL task has used mean RTs (or proportions) across all trials as their measure of lexical processing efficiency, which is conceptually similar to a causal-indicator model because it does not have word-specific error terms. Therefore, on top of the theoretically interesting question of whether individual differences in lexical processing efficiency are better explained by the central capacity account or the emergent capacity account, it is also possible that inferences about the developmental stability of lexical processing efficiency and its relation to concurrently measured variables depend on which model is employed. This may explain the

---

[1] We use the terms *causal-indicator* and *effects-indicator* models, rather than formative and reflective measurement models, because they are more neutral with regard to this debate.

seemingly small correlations between the LWL tasks at different time points (e.g., Fernald et al., 2006] reported correlations between average RTs over 3-month intervals ranging from .21 to .39). In the current study, we sought to address these questions by fitting effects-indicator and causal-indicator models (hereafter, *central capacity* and *emergent capacity* models, respectively) to LWL data at 18 months.

Whether to conduct this analysis on RTs or proportions is unclear. On the one hand, RTs seem to more closely reflect what most researchers have in mind when they use the term *efficiency* in lexical processing efficiency. On the other hand, calculation of RTs requires researchers to drop trials on which participants were looking to the distracter image prior to the onset of the target word, thereby reducing the number of valid trials. Analyzing proportions eliminates this problem because they can be calculated regardless of where the participant is looking prior to the onset of the target word. Therefore, we estimated these models on both RTs and proportions.

### Vocabulary size and developing lexical processing efficiency

There is a great deal of evidence that lexical processing efficiency is correlated with vocabulary size. For example, in a seminal longitudinal study from 15 to 25 months, Fernald et al. (2006) found that LWL RT was correlated with concurrent vocabulary by 25 months and an accuracy measure was correlated with vocabulary by 21 months. However, by 25 months, both measures were significantly correlated with vocabulary size at 15, 18, 21, and 25 months and when included in a multilevel model predicted more rapid vocabulary growth over the 10-month window.

Although performance on the LWL task and vocabulary knowledge are clearly coupled, a possible causal relationship between lexical processing efficiency and vocabulary size has been difficult to determine. Fernald et al. (2006) postulated three possible explanations. First, it is possible that young children with more efficient lexical processing are able to free cognitive resources so as to learn new words. Second, it is also possible that growth in vocabulary causes words to be represented in a more efficient manner, thereby facilitating processing. Third, it is possible that there is a bidirectional causal relationship between the two, with vocabulary improving lexical processing efficiency and lexical processing efficiency improving vocabulary.

Several additional longitudinal studies relevant to this question have been conducted, but none offers unambiguous evidence about the direction of the relationship between the variables. For example, Hurtado et al. (2008) studied the relationship among maternal talk, lexical processing efficiency, and vocabulary among Spanish-speaking children from 18 to 24 months. They found that lexical processing efficiency at 24 months mediated the relationship between maternal talk at 18 months, defined as the number of word tokens used during a free play session, and vocabulary at 24 months. In an additional model, they found that vocabulary at 24 months mediated the relationship between maternal talk at 18 months and lexical processing efficiency at 24 months. Similarly, Weisleder and Fernald (2013) found that the relationship between child-directed speech at 19 months and vocabulary size at 24 months was mediated by lexical processing efficiency (measured by proportion of time looking at target rather than RT). Although these studies suggest that variation in input variables is related to shared variance between vocabulary and lexical processing efficiency, they do not directly address the question of whether lexical processing efficiency affects vocabulary over and above prior vocabulary and vice versa.

Two studies provide more direct evidence that lexical processing efficiency predicts changes in vocabulary over and above previous vocabulary. First, Fernald and Marchman (2012) compared the vocabulary growth trajectories of typically developing and late-talking toddlers from 18 to 30 months. They found that late talkers who exhibited more efficient lexical processing at 18 months exhibited faster vocabulary growth from 18 to 30 months. This effect was not found for typically developing toddlers, which they interpreted as a ceiling effect at the older time points. Second, Lany et al. (2018) administered the LWL task and the MacArthur–Bates Communicative Development Inventory (CDI) at two sessions: one in which participants were 12 months old and one in which they were between 15 and 18 months old. They found that lexical processing efficiency, as indexed by proportion of looks to target on trials in which the target word was known during the first session, was correlated with

differences in percentile scores between CDI percentile scores in the first and second sessions. They also found that vocabulary during the first session did not predict either LWL variable during the second session or changes in LWL variables between the second and first sessions. However, as the authors pointed out, unlike the CDI percentile scores, during the second session the LWL variables were not age controlled and the relatively wide range of participant ages may be responsible for this nonsignificant relationship.

In sum, the available evidence suggests a robust relationship between lexical processing efficiency and vocabulary by around 18 months. There is some evidence that lexical processing efficiency predicts subsequent vocabulary size over and above prior vocabulary size. However, whether vocabulary predicts subsequent lexical processing efficiency over and above prior lexical processing efficiency is less clear. Moreover, no attempts have been made to directly test whether there is a bidirectional causal relationship between the two variables.

*The current study*

The current study sought to address these outstanding questions regarding the theoretical nature of lexical processing efficiency and its developmental relationship with vocabulary size. A large sample of infants completed the LWL task, and their vocabulary was measured using the CDI (Fenson et al., 2007) at 18, 21, and 24 months. Our sample size and longitudinal design allowed us to examine the following research questions:

1. Are individual differences in lexical processing efficiency at 18 months better explained by the central capacity or emergent capacity account? If there is strong evidence for one account or the other, does this conceptualization influence inferences about the stability of lexical processing efficiency from 18 to 24 months and the strength of the relationship between lexical processing efficiency and vocabulary at 18 months?
2. What is the nature of the developmental relationship between lexical processing efficiency and vocabulary size? Does variation in lexical processing efficiency predict variation in vocabulary size over and above prior vocabulary size? Does variation in vocabulary size predict variation in lexical processing efficiency over and above prior lexical processing efficiency? Is this direction unidirectional or bidirectional?

**Method**

*Participants*

Data came from a cohort of children who are being followed as part of a larger longitudinal project that is tracking the interaction between language processing and language development from 9 months to 5 years (see Kidd, Junge, Spokes, Morrison, & Cutler, (2018)). Families were recruited from a medium-sized city in Australia. Inclusion criteria for the longitudinal study were (a) full-term babies (at least 37 weeks gestation) born with a typical birth weight (>2.5 kg), (b) a predominantly monolingual language environment (no more than 20% exposure to a language other than English; thus, children were acquiring Australian English as a first language), and (c) no history of medical conditions that would affect typical language development such as repeated ear infections, visual or hearing impairment, or diagnosed developmental disabilities. Consistent with the demographics of the city, the sample was drawn from families high in SES. Approximately 75% of the parents had completed a bachelor's degree or higher.

The LWL task and the CDI were administered at 18, 21, and 24 months. Of the original 130 participants recruited to the study at 9 months, 120 were still in the study at 18 months. Two participants were excluded because they were diagnosed with hearing problems. Because different analyses included different variables, they are based on slightly different sample sizes. Sample sizes for each analysis are noted in the relevant tables.

*Materials*

The LWL task procedure was largely the same at the three time points except that the number of target words, and thus the number of trials, differed. At 18 months, the task contained 8 target words (*ball, bird, book, car, cat, dog, fish,* and *shoe*), each repeated six times. At 21 months, the task contained 10 target words (all prior words plus *apple* and *flower*), each repeated four times. At 24 months, the task contained 12 target words (all prior words plus *frog* and *teddy*), each repeated four times. Words were added at each time point to ensure that infants would maintain interest in the task if they had memories of the previous targets from their last visit. Previous longitudinal studies have also taken this approach (e.g., Fernald et al., 2006). Words at each time point were selected so as to be familiar to a majority of participants during the previous session, which was determined by identifying words that a majority of the children knew as indicated by their CDI measures from their previous visit 3 months earlier.

On each trial, two images were presented on a 1920 × 1200-pixel screen for 7000 ms. The images were of approximately equal size and enclosed in 470 × 450-pixel boxes equal distances from the center of the screen. After approximately 2000 ms, an audio file, recorded by a female native speaker of Australian English, directed the child to the target image in child-friendly natural speech,. The audio was timed so that the target word began playing at 2500 ms. The target word was introduced using one of three carrier phrases ("Look at the ____," "Where is the ____?" or "Find me the ____").

Across trials, each image occurred equally often as a target and a distracter and also occurred equally often on the left and right sides of the screen. To ensure that responses were not due to the visual salience of one target (or distracter) image, several images were used for each word across trials. At 18 months, three images were chosen for each word (meaning that each image occurred four times: twice as the target and twice as a distracter). At 21 and 24 months, two images were chosen for each word (again each image occurred four times: twice as the target and twice as the distracter).

Four pseudorandomized lists were created so that no target word was repeated within 3 trials and the target image appeared on the same side of the screen on no more than 2 consecutive trials. Attention-getting fillers were played after every 6 trials. These were dynamic cartoons with encouraging audio (e.g., "Did you see it?") that were meant to keep children engaged. Eye-tracking data were captured using a Tobii T60XL eye tracker, sampling at a rate of 60 Hz.

The CDI: Words and Sentences form (Fenson et al., 2007) was administered at 18, 21, and 24 months. Following Reilly et al. (2007,2009), some minor changes were made to a small number of words to better capture the Australian dialect. This instrument contains 678 items. Throughout these analyses, we used total productive vocabulary score as our relevant vocabulary measure.

**Results**

All data and code for these analyses are freely available online (https://osf.io/3mxps/).

*The nature of lexical processing efficiency measurements at 18 months*

To explore the differences between the central and emergent capacity accounts, causal-indicator and effects-indicator models were fit to both RTs and proportions for each word in *lavaan* (Rosseel, 2012), using full information maximum likelihood to account for missing data where necessary. Structural equation models assume that all exogenous variables follow a multivariate normal distribution. To accommodate this assumption, we followed two steps. First, for each variable we calculated skewness, kurtosis, and the Shapiro–Wilks test and visually inspected Q-Q plots to determine violations of normality. Where necessary, variables were transformed and outliers were removed (details are discussed for each analysis separately). Second, all models were estimated using the robust maximum likelihood function.

For each operationalization of lexical processing efficiency (RT and proportions), indicator variables were created for all target words at 18 months. Because the effects-indicator models can be estimated only when the latent variable predicts at least two outcome variables, 21-month vocabulary and 21-

month lexical processing efficiency were treated as outcome variables. In addition to being necessary for estimation of the models, including these models allowed us to examine whether inferences about the relationship between lexical processing efficiency and other variables was affected by the decision to use a causal-indicator or effects-indicator variable. That is, do the different conceptualizations of lexical processing efficiency result in different statistical relationships to the outcome variables? Following Willoughby et al. (2016), we created variables that are conceptually similar to causal-indicator and effects-indicator variables at 18, 21, and 24 months and calculated intraclass correlation coefficients across the three time points.

Models were fit to data including all words regardless of whether parents reported that children knew a given target word. Although there are conflicting findings on whether item-level data from the CDI predicts performance on preferential looking tasks (Houston-Price, Mather, & Sakkalou, 2007; Styles & Plunkett, 2009), removing the target words greatly increased the amount of missing data and resulting models fit poorly. Moreover, given that lexical processing efficiency is conceptualized as a graded measure, and that CDI measures at 18 months ask only about production, it is possible that available CDI measures underestimate children's processing of unreported words. Consistent with this argument, previous reported studies (Fernald et al., 2006) have found that inferences about the relationship between lexical processing efficiency and other variables are generally unaffected by removing unknown words.

### Log RTs

Prior to calculating RTs, we removed trials in which participants were looking to the screen for less than 50% of the 3000-ms window between the onset of the target word and the offset of the image ($M$ = 7.14 trials, range = 0–28). Then, following Fernald and Marchman (2012), we calculated the duration to the first look to the target image on trials in which participants were (a) looking to the distracter image prior to the target word and (b) shifted to the target image between 300 and 1800 ms after the onset of the target word. The first look to the target image was defined as the first fixation of at least 100 ms to the target image. This resulted in an average of 14.20 trials per participant (range = 4–26). The RTs were then log transformed.[2] At this point, our data preparation deviated from that of standard LWL studies. Prior studies calculated the average RT across all words. However, to compare the central capacity and emergent capacity accounts, we calculated average log RTs at the level of individual words for each participant. Because this task necessarily entails a large amount of missing data, most participants did not have values for all eight words. Of the 115 participants for whom 18-month eye-tracking data were available, 40 participants had log RTs for all eight words, 31 had log RTs for seven words, 23 had log RTs for six words, 14 had log RTs for five words, 4 had log RTs for four words, and 3 had log RTs for three words. To strike a balance of including as many participants as possible but minimizing the number of missing patterns from the dataset (because models with large amounts of missing data fit poorly), we excluded the 21 participants with log RTs for fewer than six words. The median number (and range) of trials available for each word are presented in Table 1.

Descriptive statistics for all eight words and the three outcome variables are presented in Table 2. As can be seen, 21-month vocabulary was not normally distributed, thereby violating the assumption of multivariate normally distributed exogenous variables. Therefore, we transformed vocabulary using the square root. The resulting distribution was closer to normal according to both test statistics and visual inspection of Q-Q plots.[3] All other variables were approximately normally distributed.

Model fit statistics for each of the two models are presented in Table 3. Both models exhibited excellent fit according to several common fit indices. Table 4 presents standardized path coefficients to and from the latent variables in the two models. In the central capacity model, all path coefficients from the latent lexical processing efficiency variable to the word-specific RTs were in the same direction, and all except *shoe* were statistically significant. In the emergent capacity model, all coefficients were of the same direction, but no coefficient was statistically significant. However, given the large

---

[2] We used log RT rather than raw RT because we found that log RTs were within the normal range for both skewness and kurtosis, thereby satisfying the assumptions of normal structural equation models.

[3] Results did not substantively differ when raw vocabulary was included in the model.

**Table 1**
Number of valid trials per participant for log RTs and proportions.

| | Log RTs | | | Proportions | | |
|---|---|---|---|---|---|---|
| | Median | *SD* | Range | Median | *SD* | Range |
| Ball | 2 | 1.19 | 0–6 | 5 | 1.33 | 0–6 |
| Bird | 2 | 1.26 | 0–6 | 6 | 1.08 | 2–6 |
| Book | 2 | 1.10 | 0–6 | 6 | 1.02 | 2–6 |
| Car | 2 | 1.29 | 0–6 | 6 | 1.07 | 1–6 |
| Cat | 1.5 | 1.11 | 0–4 | 5 | 1.10 | 2–6 |
| Dog | 2 | 0.98 | 0–3 | 6 | 1.06 | 1–6 |
| Fish | 2 | 1.21 | 0–3 | 5 | 1.09 | 1–6 |
| Shoe | 2 | 1.20 | 0–5 | 5 | 0.98 | 3–6 |

*Note.* RT, reaction time.

**Table 2**
Descriptive statistics.

| | Mean | *SD* | Skewness | Kurtosis | Missing | Shapiro–Wilks (*p* value) |
|---|---|---|---|---|---|---|
| *Log RTs* | | | | | | |
| Ball | 6.60 | 0.30 | .16 | 2.62 | 4 | .747 |
| Bird | 6.64 | 0.34 | .29 | 2.59 | 9 | .126 |
| Book | 6.63 | 0.35 | −.05 | 2.68 | 5 | .892 |
| Car | 6.67 | 0.28 | .48 | 3.14 | 7 | .124 |
| Cat | 6.57 | 0.34 | .02 | 2.12 | 24 | .165 |
| Dog | 6.56 | 0.29 | .14 | 2.53 | 12 | .699 |
| Fish | 6.73 | 0.31 | −.07 | 2.63 | 9 | .624 |
| Shoe | 6.72 | 0.33 | −.46 | 3.59 | 7 | .078 |
| Log RT 21 months | 6.54 | 0.19 | .33 | 2.68 | 3 | .289 |
| Vocab 21 months | 220.90 | 140.90 | .64 | 3.05 | 2 | .007 |
| SQRT Vocab 21 months | 13.99 | 5.06 | −.16 | 2.52 | 2 | .645 |
| *E-logit proportions* | | | | | | |
| Ball | 0.46 | 1.24 | .01 | 3.11 | 1 | .491 |
| Bird | 1.15 | 1.32 | .08 | 3.29 | 0 | .294 |
| Book | 1.16 | 1.28 | −.23 | 3.28 | 2 | .040 |
| Car | 1.19 | 1.69 | −.20 | 3.84 | 2 | .137 |
| Cat | 1.55 | 1.35 | .06 | 2.83 | 0 | .213 |
| Dog | 1.78 | 1.36 | .19 | 3.49 | 1 | .234 |
| Fish | 0.55 | 1.32 | −.48 | 2.89 | 1 | .069 |
| Shoe | 0.72 | 1.48 | .10 | 3.54 | 0 | .647 |
| E-logit 21 months | 1.45 | 0.55 | −.21 | 3.25 | 5 | .344 |
| Vocab 21 months | 221.15 | 143.33 | .58 | 2.80 | 4 | .280 |
| SQRT Vocab 21 months | 13.96 | 5.16 | −.15 | 2.38 | 4 | .359 |

*Note.* RT, reaction time; Vocab, vocabulary; SQRT, square root; E-logit, empirical logit.

number of parameters in this model, the standard errors for these coefficients were very large and this model may have been underpowered to detect these effects.

Our next step was to compare model fit. The two models are not formally nested (i.e., there is no set of parameters in the emergent capacity model that can be set to 0 to get the central capacity model), and so they cannot be compared via a likelihood ratio test. However, the models are nested with respect to their vanishing tetrads. This allows a formal test of whether the central capacity model fits significantly worse than the emergent capacity model (Hipp, Bauer, & Bollen, 2005). In particular, because the central capacity model assumes that all correlations between indicator variables are due to their shared latent variable, it implies that the differences in the products of certain pairs of covariances should be 0, so-called vanishing tetrads. On the other hand, because the emergent capacity model indicator model includes correlations between every pair of indicator variables, it assumes that fewer should be 0. The two models can, therefore, be compared with a nested vanishing tetrad test. The central capacity model places more restrictions on the covariance structure than the emer-

**Table 3**
Fit statistics for structural equation models.

| | Log RT ($n$ = 94) | | Proportions ($n$ = 115) | |
|---|---|---|---|---|
| | Centralized model | Emergent model | Centralized model | Emergent model |
| Chi-square | $\chi^2(34)$ = 33.47 $p$ = .493 | $\chi^2(7)$ = 5.25, $p$ = .630 | $\chi^2(34)$ = 31.99, $p$ = .567 | $\chi^2(7)$ = 5.03, $p$ = .656 |
| CFI | 1.00 | 1.00 | 1.00 | 1.00 |
| TLI | 1.01 | 1.09 | 1.05 | 1.12 |
| RMSEA | .00 | .00 | .00 | .00 |
| VTT | $\chi^2(34)$ = 13.59, $p$ = .999 | $\chi^2(7)$ = 3.98, $p$ = .782 | $\chi^2(34)$ = 9.80, $p$ = .999 | $\chi^2(7)$ = 8.32, $p$ = .832 |
| Nested VTT | $\chi^2(27)$ = 9.97, $p$ = .998 | | $\chi^2(27)$ = 6.27, $p$ = .999 | |

*Note.* RT, reaction time; CFI, comparative fit index; TLI, Tucker–Lewis index; RMSEA, root mean square error of approximation; VTT, vanishing tetrad test.

**Table 4**
Indicator variables from structural equation models.

| | Log RT ($n$ = 94) | | E-logit proportions ($n$ = 115) | |
|---|---|---|---|---|
| | Centralized model | Emergent model | Centralized model | Emergent model |
| Ball | .718 (fixed) | .231 (fixed) | .360[*] | .188 |
| Bird | .562[***] | .189 | .429[*] | .352[a] |
| Book | .472[**] | .082 | .246[*] | .195 |
| Car | .433[*] | .184 | .184 | .231 |
| Cat | .554[**] | .166 | .098 | .098 |
| Dog | .378[**] | .039 | −.018 | −.076 |
| Fish | .478[***] | .021 | .548 (fixed) | .556 (fixed) |
| Shoe | .173 | .115 | .341[*] | .114 |

*Note.* RT, reaction time; E-logit, empirical logit.
[a] 05 < $p$ < .10.
[*] .01 < $p$ < .05.
[**] .001 < .$p$ < .01.
[***] $p$ < .001.

**Table 5**
Regressions from latent variable in structural equation models.

| | Log RT ($n$ = 94) | | E-logit proportions ($n$ = 115) | |
|---|---|---|---|---|
| | Central model | Emergent model | Central model | Emergent model |
| Vocab | −.516[***] | −.697[*] | .673[***] | .545** |
| LPE | .621[***] | .927[*] | .372[*] | .291[*] |

*Note.* RT, reaction time; E-logit, empirical logit; Vocab, vocabulary; LPE, lexical processing efficiency.
[*] .01 < $p$ < .05.
[***] $p$ < .001.

gent capacity model, and this allows model comparison via a nested vanishing tetrad test. A statistically significant nested vanishing tetrad test would imply that the central capacity model, which has fewer parameters, fits significantly worse than the emergent capacity model, which has more parameters. We conducted the nested vanishing tetrad test using the CTANEST SAS macro (Hipp et al., 2005). The test was nonsignificant, indicating that the central capacity model did not fit significantly worse than the emergent capacity model, $\chi^2(27)$ = 9.97, $p$ = .998.

As can also be seen in Table 5, both models yielded significant relationships between the 18-month lexical processing efficiency variable and the 21-month lexical processing efficiency and vocabulary. However, the magnitudes of the coefficients were larger for the emergent capacity model. A related

question is whether these two approaches yield different conclusions about the developmental stability of lexical processing efficiency. To address this question, we constructed two indices that were conceptually similar to the central capacity and emergent capacity models using the data from the LWL task at 18, 21, and 24 months. To build emergent capacity indices, we took an average of the *z*-scored log RTs for the eight words given in the 18-month sessions at each of the three time points. To build central capacity indices, we fit linear mixed-effects models with crossed random effects by items and words at 18, 21, and 24 months separately and then extracted the by-participant random intercepts.[4] We then calculated the intraclass correlation coefficient (ICC) on each of these estimates. We found that the ICC for the emergent capacity indices was .33 (confidence interval [CI] = .21–.46) and that the ICC for central capacity indices was .24 (CI = .11–.36). Whereas the ICC was slightly higher for the central capacity indices, the confidence interval overlapped with the value of the ICC for the emergent capacity indices, suggesting that the two procedures do not differ.

*Proportions*

As was the case prior to calculating RTs, we removed trials in which participants were looking to the screen for less than 50% of the 3000 ms. Trials were not excluded for any other reason. The proportion of looks to the target word was, then, calculated for each trial. We applied the empirical logit transformation to these proportions because we found that this reduced the skew. We then averaged these empirical logit transformed proportions across words for each participant. Unlike the case with RTs, all but 1 participant had proportions for all eight words. The distribution of several words still deviated from normal. Examination of the Q-Q plot revealed that this may be due to outliers. Therefore, we removed data points that exceeded 3 standard deviations from the mean of each word (*n* = 5). The median number (and range) of trials available for each word are presented in Table 1.

Descriptive statistics for all eight words and the three outcome variables are presented in Table 2. As was the case with RTs, vocabulary was not normally distributed, and so we included the square root of vocabulary in both models.

Model fit statistics for each of the two models are presented in Table 3. Both models exhibited excellent fit according to several common fit indices. Table 4 presents standardized path coefficients to and from the latent variables in the two models. In the central capacity model, all path coefficients from the latent lexical processing efficiency variable to the word-specific RTs were in the same direction except that for the word *dog,* and all except *car, cat,* and *dog* were statistically significant. In the emergent capacity model, all coefficients except that for *dog* were in the same direction, and only that for *bird* was marginally statistically significant.

To formally compare model fit for the two models, we again ran the nested vanishing tetrad test, which was nonsignificant, indicating that the central capacity model did not fit worse than the emergent capacity model, $\chi^2(27) = 6.28$, $p = .999$. To determine whether the models yielded different inferences, we again examined the implied relationships between 18-month lexical processing efficiency and 21-month variables and the developmental stability of lexical processing efficiency from 18 to 24 months. As can be seen in Table 5, the path coefficients from 18-month lexical processing efficiency to 21-month lexical processing efficiency and vocabulary were significant in both models but slightly larger in magnitude in the central capacity model than in the emergent capacity model. ICCs for both conceptualizations were small with overlapping confidence intervals (central capacity: ICC = .14, CI = .03–27; emergent capacity: ICC = .18, CI = .06–.31).

*The longitudinal relationship between lexical processing efficiency and vocabulary*

To determine whether lexical processing efficiency predicted subsequent growth in vocabulary and whether vocabulary predicted subsequent growth in lexical processing efficiency, we estimated a set

---

[4] The causal-indicator indices were estimated using the same eight words at all three time points. However, the effects-indicator indices were estimated using all eight words at 18 months, all 10 words at 21 months, and all 12 words at 24 months. This is because in the causal-indicator model the meaning of the latent variable changes, depending on which indicator variables are used. However, in the effects-indicator model the latent variable exists independently from the indicator variables, and its meaning should be the same regardless of which indicators are used.

of regressions.[5] To examine the relationship between lexical processing efficiency and vocabulary, we predicted vocabulary at both 21 and 24 months using the prior time points' vocabulary and lexical processing efficiency. We followed the same method for lexical processing efficiency. For each set of analyses, we considered lexical processing efficiency defined using both RTs and proportions. Models with RTs as the dependent variable produced skewed residuals, whereas those with log transformed RTs did not. Therefore, we used log transformed RTs in all analyses.[6]

One possible explanation for any observed relationship between lexical processing efficiency and vocabulary is that variation in measured lexical processing efficiency is affected by variation in the number of target words that participants know. Although results of previous research were generally unaffected by including unreported words (Fernald et al., 2006), to control for this possibility we conducted all analyses using lexical processing efficiency calculated on all target words and lexical processing efficiency calculated using only the trials for which participants knew the target word according to their CDI measures. Because results were qualitatively similar across these two approaches, we report on analyses of all words in the main text but report parameter estimates from both analyses in tables.

Table 6 contains descriptive statistics for the lexical processing efficiency variables, vocabulary, and demographic variables from relevant subsamples across the three time points. To put our sample within context, we also report median vocabulary and interquartile range of vocabulary from the American sample of Wordbank (Frank, Braginsky, Yurovsky, & Marchman, 2017) because the Australian sample available on Wordbank completed a separate form. Consistent with its relatively high SES, our sample was above the median but within the interquartile range at all three time points. Tables 7 and 8 contain correlations between log RT and vocabulary at the three time points. As can be seen, when lexical processing efficiency was operationalized as log RT, it was correlated with vocabulary at all time points when measured at 18 and 21 months, but 24-month RTs were correlated with 21- and 24-month vocabulary only. When operationalized as proportions, 18- and 21-month lexical processing efficiency was correlated with all vocabulary measurements and 24-month lexical processing efficiency was correlated with concurrent vocabulary.

*Changes in vocabulary*

One complication in predicting changes in vocabulary size from lexical processing efficiency is that in the current sample the relationship between vocabulary sizes at consecutive time points is markedly nonlinear (see Appendix A). Determining whether lexical processing efficiency predicts vocabulary over and above prior vocabulary requires modeling this nonlinearity. Therefore, prior to fitting models with lexical processing efficiency as a predictor, we fit a series of linear regressions predicting vocabulary (at both 21 and 24 months) with increasingly higher-order orthogonal polynomial terms for prior vocabulary. As can be seen in Table 9, at 21 months the model with linear and quadratic terms of prior vocabulary fit nearly as well as the model with a cubic term. Thus, we selected the model with linear and quadratic terms for the sake of parsimony. At 24 months, the model with linear, quadratic, and cubic terms for prior vocabulary was comparable to the model with the quartic term according to adjusted $R^2$ and was slightly worse according to Akaike information criterion (AIC). For the sake of parsimony, we decided to choose the model with linear, quadratic, and cubic terms. To examine the effect of lexical processing efficiency, we added the lexical processing efficiency variables to the models identified above.

Parameter estimates for models of 21-month vocabulary are presented in Table 10. When indexed with log RT, lexical processing efficiency at 18 months did not significantly predict 21-month vocabulary when controlling for 18-month vocabulary, $b = -29.19$, $t(105) = -0.77$, $p = .441$. However, when indexed by proportion of looks, lexical processing efficiency marginally significantly predicted

---

[5] The ideal approach to addressing this question would be a random-intercept cross-lagged path analysis. However, path analysis would not be able to easily accommodate the nonlinear relationships between vocabulary at consecutive time points (described in more detail below).

[6] One participant was excluded from all regressions because across all models this participant had a very large Cook's distance relative to other cases, suggesting that the participant was greatly affecting regression lines. Examination of raw data revealed that this participant had an extremely high productive vocabulary at 18 months (412 words).

**Table 6**
Descriptive statistics for variables at the three time points.

|  | 18 months | 21 months | 24 months |
|---|---|---|---|
| RT | 848 (156) | 769 (168) | 566 (117) |
| Log RT | 6.67 (0.18) | 6.56 (0.19) | 6.26 (0.18) |
| Prop | .63 (.07) | .68 (.06) | .70 (.07) |
| Vocab current sample | 102 (94) | 217 (136) | 340 (146) |
| Vocab Wordbank | 76 (32–173) | 178 (81–320) | 316 (165–454) |
| Age (in days) | 568 (7) | 656 (6) | 749 (8) |
| Percentage female | 46 | 46 | 48 |
| Caregiver 1 Education | 3.96 (1.14) | 3.96 (1.14) | 3.93 (1.15) |
| Caregiver 2 education | 3.57 (1.55) | 3.53 (1.54) | 3.51 (1.54) |

*Note.* Standard deviations are in parentheses except for Vocab Wordbank (where ranges are in parentheses). RT, reaction time; Prop, proportion; Vocab, vocabulary. Vocab Wordbank is the median vocabulary size (and interquartile range) from the American sample on Wordbank for the relevant age group.

**Table 7**
Correlation matrix for log RTs.

|  | RT 18 months | RT 21 months | RT 24 months | Vocab 18 months | Vocab 21 months | Vocab 24 months |
|---|---|---|---|---|---|---|
| RT 18 months | 1.00 |  |  |  |  |  |
| RT 21 months | .45*** | 1.00 |  |  |  |  |
| RT 24 months | .26** | .27** | 1.00 |  |  |  |
| Vocab 18 months | −.25** | −.33*** | −.18[a] | 1.00 |  |  |
| Vocab 21 months | −.32*** | −.40*** | −.20* | .85*** | 1.00 |  |
| Vocab 24 months | −.35*** | −.41** | −.24* | .70*** | .90*** | 1.00 |

*Note.* RT, reaction time; Vocab, vocabulary.
[a] $.05 < p < .10$.
* $.01 < p < .05$.
** $.001 < p < .01$.
*** $p < .001$.

**Table 8**
Correlation matrix for proportions.

|  | Prop 18 months | Prop 21 months | Prop 24 months | Vocab 18 months | Vocab 21 months | Vocab 24 months |
|---|---|---|---|---|---|---|
| Prop 18 months | 1.00 |  |  |  |  |  |
| Prop 21 months | .42*** | 1.00 |  |  |  |  |
| Prop 24 months | .23* | .21* | 1.00 |  |  |  |
| Vocab 18 months | .37*** | .31** | .15 | 1.00 |  |  |
| Vocab 21 months | .43*** | .38*** | 17[a] | .85*** | 1.00 |  |
| Vocab 24 months | .43*** | .37*** | .20* | .70*** | .90*** | 1.00 |

*Note.* Prop, proportion; Vocab, vocabulary.
[a] $.05 < p < .10$.
* $.01 < p < .05$.
** $.001 < p < .01$.
*** $p < .001$.

increases in vocabulary size, $b = 181.15$, $t(105) = 1.85$, $p = .067$. Parameter estimates for models of 24-month vocabulary are presented in Table 11. When indexed with log RT, lexical processing efficiency at 21 months was nonsignificant after controlling for prior vocabulary, $b = 40.61$, $t(102) = 1.34$, $p = .183$. The result was the same when 21-month lexical processing efficiency was operationalized as proportion of looks, $b = −40.86$, $t(103) = −0.44$, $p = .664$.

**Table 9**
Fit statistics for vocabulary models with nonlinear terms.

| | 21-Month Vocab | | 24-Month Vocab | |
|---|---|---|---|---|
| | AIC | Adjusted $R^2$ | AIC | Adjusted $R^2$ |
| Linear | 1250.50 | .71 | 1208.60 | .80 |
| Linear + quadratic | 1229.57 | .76 | 1188.94 | .84 |
| Linear + quadratic + cubic | 1229.45 | .77 | 1182.07 | .85 |
| Linear + quadratic + cubic + quartic | 1231.44 | .76 | 1180.48 | .85 |

*Note.* Vocab, vocabulary; AIC, Akaike information criterion.

**Table 10**
Regression models predicting 21-month vocabulary.

| | LWL RT | | LWL Prop | |
|---|---|---|---|---|
| | All trials | No unknown | All trials | No unknown |
| Intercept | 412.64 | 390.64 | 104.31[a] | 114.32[a] |
| Vocab 18 months | 1192.73*** | 1162.49*** | 1158.19*** | 1152.26*** |
| Vocab 18 months$^2$ | −318.69*** | −307.32*** | −305.96*** | −311.87*** |
| LWL | −29.19 | −25.19 | 181.15[a] | 167.96[a] |
| Adjusted $R^2$ | .76 | .75 | .77 | .76 |
| N | 109 | 106 | 109 | 107 |

*Note.* LWL, Looking While Listening; RT, reaction time; Prop, proportion; Vocab, vocabulary; Vocab 18 months $^2$, squared 18 month vocabulary.
  [a] $.05 < p < .10$.
  *** $p < .001$.

**Table 11**
Regression models predicting 24-month vocabulary.

| | LWL RT | | LWL Prop | |
|---|---|---|---|---|
| | All trials | No unknown | All trials | No unknown |
| Intercept | 78.55 | 71.87 | 371.60*** | 341.08*** |
| Vocab 21 months | 1375.73*** | 1374.55*** | 1361.66*** | 1350.27*** |
| Vocab 21 months$^2$ | −280.22*** | −278.45*** | −284.31*** | −280.36*** |
| Vocab 21 months$^3$ | 168.90** | 168.44** | 170.81** | 166.92** |
| LWL | 40.61 | 41.65 | −40.86 | 4.14 |
| Adjusted $R^2$ | .85 | .85 | .85 | .85 |
| N | 107 | 107 | 108 | 108 |

*Note.* LWL, Looking While Listening; RT, reaction time; Prop, proportion; Vocab, vocabulary; Vocab 21 months$^2$, squared 21 month vocabulary; Vocab 21 months$^3$, cubed 21 month vocabulary.
  ** $.001 < p < .01$.
  *** $p < .001$.

*Changes in lexical processing efficiency*

There was no evidence for a nonlinear relationship between measures of lexical processing efficiency across time periods. Adding polynomial terms to the models did not improve fit, so we do not report those analyses here. When lexical processing efficiency was operationalized using RT, examination of residuals indicated that linear models on log RT fit better than linear models on raw RT. When lexical processing efficiency was operationalized as a proportion, examination of residuals indicated that beta regressions (Smithson & Verkuilen, 2006) fit better than linear regressions.[7] Vocabulary was converted to *z* scores to facilitate interpretation of model coefficients.

---

[7] Beta regression assumes a beta likelihood function, which is defined between 0 and 1, thereby accounting for the floor and ceiling effects, and the consequent heteroscedasticity, inherent in proportions.

**Table 12**
Regression models predicting 21-month lexical processing efficiency.

| | LWL RT | | LWL Prop | |
|---|---|---|---|---|
| | All trials | No unknown | All trials | No unknown |
| Intercept | 3.86*** | 4.23*** | −0.18 | 0.05 |
| LWL 18 months | 0.40*** | 0.35** | 1.47*** | 1.12** |
| Vocab 18 months | −0.04* | −0.05* | 0.05[a] | 0.06* |
| Adjusted $R^2$ | .24 | .17 | .21 | .14 |
| N | 107 | 104 | 108 | 105 |

*Note.* LWL, Looking While Listening; RT, reaction time; Prop, proportion; Vocab, vocabulary.
[a] $.05 < p < .10$.
* $.01 < p < .05$.
** $.001 < p < .01$.
*** $p < .001$.

**Table 13**
Regression models predicting 24-month lexical processing efficiency.

| | LWL RT | | LWL Prop | |
|---|---|---|---|---|
| | All trials | No unknown | All trials | No unknown |
| Intercept | 4.89*** | 5.23 | 0.33 | 0.35 |
| LWL 21 months | 0.21* | 0.16 | 0.79 | 0.76 |
| Vocab 21 months | −0.02 | −0.02 | 0.04 | 0.03 |
| Adjusted $R^2$ | .08 | .04 | .06 | .05 |
| N | 103 | 103 | 104 | 103 |

*Note.* LWL, Looking While Listening; RT, reaction time; Prop, proportion; Vocab, vocabulary.
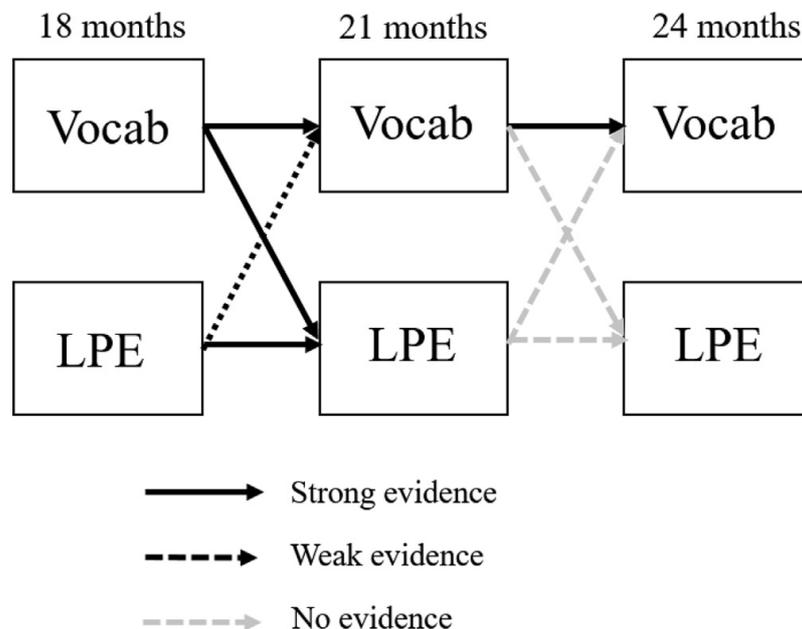* $.01 < p < .05$.
*** $p < .001$.



**Fig. 2.** Schematic relationship of the results of the longitudinal study. Vocab, vocabulary; LPE, lexical processing efficiency.

As can be seen in Table 12, 18-month vocabulary significantly predicted 21-month log RTs over and above 18-month log RTs, $b = -0.04$, $t(104) = -2.65$, $p = .009$, and marginally significantly predicted proportions, $b = 0.05$, $z = 1.82$, $p = .069$. Table 13 shows that 21-month vocabulary did not significantly

predict 24-month log RTs over and above prior log RTs, $b = -0.02$, $t(98) = -0.808$, $p = .421$, or 24-month proportions over and above prior proportions, $b = 0.04$, $z = 1.14$, $p = .256$.

Fig. 2 depicts a schematic overview of the results of the longitudinal analyses. There was strong evidence for an effect of 18-month vocabulary on 21-month lexical processing efficiency over and above 18-month lexical processing efficiency, an effect that was either significant or marginally significant over all four analyses. There was some evidence for an effect of 18-month lexical processing efficiency on 21-month vocabulary over and above 18-month vocabulary, an effect that was marginally significant when lexical processing efficiency was measured by proportions. There was no evidence for an effect of 21-month lexical processing efficiency on 24-month vocabulary or of 21-month vocabulary on 24-month lexical processing efficiency.

## Discussion

The current study had two goals. The first was to determine whether lexical processing efficiency is better conceptualized as a central or emergent processing capacity. To this end, we modeled 18-month LWL RTs using effects-indicator and causal-indicator models, which we labeled the *central capacity* and *emergent capacity* models, respectively. Both models displayed excellent fit according to conventional structural equation modeling fit indices. Moreover, inferences about the relationship between lexical processing efficiency and other constructs and inferences about the stability of lexical processing efficiency over time were similar in the two models. The second goal was to determine the relationship between vocabulary and lexical processing efficiency over time. We found clear evidence that 18-month vocabulary predicted 21-month lexical processing efficiency over and above 18-month LWL; however, this relationship was nonsignificant between 21 and 24 months. We found weaker evidence of an effect of 18-month lexical processing efficiency on 21-month vocabulary over and above 18-month vocabulary and found no effect of 21-month LWL on 24-month vocabulary.

### The nature of individual differences in lexical processing efficiency

Two points about the comparative fits of the central capacity and emergent capacity models warrant discussion. First, although we could not statistically distinguish between the two models, all the coefficients from the emergent capacity model to the latent variable were nonsignificant. If the emergent capacity model were correct, one would expect these coefficients to be significant because they cause variation in lexical processing efficiency. This pattern of nonsignificant coefficients could be taken as evidence for the central capacity model. However, for log RTs all the coefficients were in the same (positive) direction, and for proportions all but one coefficient were in the same direction. Moreover, in both cases standard errors were quite large, which is unsurprising given the large number of parameters in the causal-indicator model. If lexical processing efficiency reflects a constellation of word-specific processing speeds, one would expect the contribution of any one particular word to be quite small. As such, the data may have been underpowered to estimate these coefficients. Second, the nested vanishing tetrad test indicated that the central capacity model did not fit significantly worse than the emergent capacity model. Following the logic of a likelihood ratio test, one is tempted to view this as evidence for the superiority of the central capacity model; the model with fewer parameters did not fit significantly worse and, therefore, is the more parsimonious option. However, we are uncomfortable with this logic. Because the models are not nested in terms of their parameters, it is not clear that the central capacity model is more parsimonious than the emergent capacity model. Therefore, we conclude that there is no strong reason to prefer one model over the other.

One encouraging result of our comparison of the two models is that inferences about the stability of lexical processing efficiency and its relationship to other constructs did not vary substantially. Standardized path coefficients to 21-month vocabulary and lexical processing efficiency were significant in both models, and whereas they were larger for the emergent capacity model than for the central capacity model when log RTs were used, they were larger for the central capacity model than for the emergent capacity model when proportions were used. Moreover, the ICCs for the two models were of similar magnitudes with overlapping confidence intervals. This differs from the results of

Willoughby et al. (2016), who found that inferences about the developmental stability of executive functions depended critically on whether it was modeled using effects indicators or causal indicators. Willoughby et al.'s finding creates a challenge for developmental researchers in that their conclusions will be strongly affected by the theoretical assumptions implicit in their analytical strategy. Our results suggest that this might not be the case for researchers studying lexical processing efficiency. Given the similarity of results in these two models, researchers can ask questions about the stability of lexical processing efficiency and its relation to other variables without making strong assumptions about the theoretical nature of individual differences in lexical processing efficiency. However, it remains possible that, with more trials per item, inferences from these two approaches might differ more substantially.

### The longitudinal relationship between lexical processing efficiency and vocabulary development

Fernald et al. (2006) identified three possible causal explanations for the relationship between vocabulary and lexical processing efficiency during the second year of life, namely that increases in lexical processing efficiency could lead to increases in vocabulary, increases in vocabulary could lead to increases in lexical processing efficiency, or there could be a bidirectional relationship between the two. Despite the fact that lexical processing efficiency and vocabulary were correlated at all three time points, our longitudinal design allowed us to disentangle these three possibilities. The current data rule out the possibility of a mono-causal relationship from lexical processing efficiency to vocabulary; there was strong evidence of an effect of 18-month vocabulary on 21-month vocabulary and only weak evidence of an effect of 18-month lexical processing efficiency on 21-month vocabulary. We consider the possible explanations for an effect of vocabulary on lexical processing efficiency and a bidirectional relationship in turn.

Why might increases in vocabulary size lead to increases in lexical processing efficiency? Although this question has been considered less frequently than whether increases in lexical processing efficiency lead to increases in vocabulary, Fernald et al. (2006) speculated that increasing vocabulary size results in pressure to create more fine-grained representations of word forms. However, young infants already have fairly detailed phonological representations. For instance, Swingley and Aslin (2002) showed that infants as young as 15 months are sensitive to initial-consonant mispronunciations of words, and recent work by Kidd, Junge, Spokes, Morrison, and Cutler (2018) showed that some 9-month-old infants can rapidly create and access memories for newly presented words (for a review of early lexical knowledge, see Johnson, 2016). Thus, other forces are likely to be at play. Moreover, any account of the effect of vocabulary on lexical processing efficiency must explain why it existed from 18 to 21 months but not from 21 to 24 months.

One possibility is the changing network structure of the lexicon between 18 and 24 months. It is well understood that among adults neighborhood effects in the lexicon can be facilitatory or inhibitory, with semantic neighborhood effects facilitating lexical access and phonological neighborhood effects having an inhibitory effect (Chen & Mirman, 2012). There is increasing evidence of both effects in children as young as 24 months. For example, Borovsky, Ellis, Evans, and Elman (2016) found evidence of a facilitatory effect of semantic neighborhood density in an LWL task at 24 months, and Mani and Plunkett (2011) found evidence of inhibitory effects of phonological neighborhood density in a priming task at 24 months. If lexical development over the second year of life is initially characterized by increased semantic neighborhood density and then becomes influenced by both phonological and semantic neighborhood density, we could expect to see the pattern observed in the current study. Early growth in vocabulary would facilitate lexical processing efficiency increasing semantic neighborhood density; later growth would be unrelated to lexical processing efficiency because of the countervailing effects of both semantic and phonological neighborhood density. There is some indirect evidence supporting this account. Rämä, Sirri, and Serres (2013) observed electrophysiological evidence of semantic priming in high-vocabulary 18-month-olds but not in low-vocabulary 18-month-olds, and Mani and Plunkett (2010, 2011) found no evidence of phonological neighborhood effects at 18 months but did find effects at 24 months. Thus, facilitative semantic neighborhood effects appear developmentally prior to inhibitory phonological neighborhood effects and are related to vocabulary size.

The influence of the structure of the lexicon on the LWL task and its relation to individual differences in vocabulary awaits further research.

Our analyses do not rule out the possibility of a bidirectional relationship between lexical processing efficiency and vocabulary, but evidence for an effect of lexical processing efficiency on vocabulary was relatively weak. At first glance, this finding seems inconsistent with two studies directly testing the effect of lexical processing efficiency on vocabulary, but these findings can be reconciled. First, Lany et al. (2018) reported that 12-month LWL accuracy predicted changes in CDI percentile scores between a first testing session (12 months) and a second testing session (15–18 months), with infants who were more accurate on the LWL task exhibiting bigger gains in their percentile scores on the CDI over the following months. Given the difference of ages between the samples, it is possible that the effects of lexical processing efficiency on vocabulary are larger at earlier ages and slowly diminish. This would be consistent with the 18-month correlation between lexical processing efficiency and vocabulary. However, it is also important to note that the difference scores used by Lany et al. (2018) implicitly assume a linear relationship between vocabulary at consecutive time points; whether this assumption is more plausible for percentile scores than for raw vocabulary scores is unclear.

This finding also seems inconsistent with the results of Lany (2018), who found that participants with more efficient lexical processing as operationalized by RTs learned words more quickly under moderately challenging conditions than those with less efficient lexical processing. However, it may be that after 18 months lexical processing efficiency plays a role in the initial stages of word learning but that the effect is too subtle to be observed in aggregate vocabulary size. For example, Storkel and Lee (2011) pointed out that learning a new word involves several partially separable but dependent cognitive processes and that these differentially affect immediate and long-term retention of new words. Lexical processing efficiency may affect initial cognitive processes in detecting and encoding a novel word. Therefore, they may be more easily detectable in a laboratory-based novel word-learning task but less detectable when the outcome measure is existing vocabulary.

More recent research suggests that the relationship between lexical processing efficiency and vocabulary may be quite complex. In a study conducted concurrently with the current one, Peter et al. (2019) administered the LWL task at 19 months and examined how it predicted growth in lexical knowledge from 19 to 31 months. They found that for children with smaller vocabularies at 19 months 19-month lexical processing efficiency predicted subsequent vocabulary growth, but for children with larger vocabularies at 19 months it did not. Examining such a relationship was not possible in the current study because of the relatively small number of time points.

Finally, we note that although the variables remain correlated later in development, our regression analyses did not detect a relationship between the variables after 21 months, such that they predict themselves (at best) only at 24 months. We should not be surprised at identity relations like this in longitudinal data, but the results suggest a cautionary note: We should be careful about interpreting cross-variable effects (e.g., lexical processing efficiency predicting vocabulary) in this age range when prior variables have not been controlled. This is particularly important for central capacity explanations of development, which predict that development proceeds via domain-general increases in central processing speed that in turn support the learning of new words. Although attractive, such explanations have been criticized across several domains for privileging capacity over knowledge (e.g., Cowan, Rouder, Blume, & Scott Saults, 2012; MacDonald & Christiansen, 2002), which in developmental studies are only controlled using cross-lagged longitudinal designs like ours. Thus, future longitudinal studies, perhaps across a wider age range, should employ similar designs to better understand the theoretical nature of lexical processing efficiency and its role in vocabulary acquisition.

## Limitations

The current research should be considered with two significant limitations in mind. First, the data used in the structural equation models fit on the 18-month data were necessarily noisy. Indeed, RT data could be based on as few as one trial per word for each child. Although more data were available for proportions, future research should consider designs that maximize the number of valid trials per

participants. Increasing the number of trials may allow the models to better distinguish between these two conceptualizations. Second, our sample contains children predominantly from upper-SES backgrounds. It is possible that the longitudinal relationship between vocabulary and lexical processing efficiency would differ in a more diverse sample given that both lexical processing efficiency and vocabulary are related to the quality of input (Weisleder & Fernald, 2013).

## Conclusions

Lexical processing efficiency as measured by the LWL task is an important concept in research investigating individual differences in early language development, but the interpretation of the developmental relationship has been unclear. In the current study, we investigated (a) the nature of individual differences on lexical processing efficiency tasks and (b) the longitudinal relationship between lexical processing efficiency and vocabulary over the second year of life. We found that lexical processing efficiency could be well modeled as either an emergent capacity or a central capacity, and we found that there was a clear effect of 18-month vocabulary on 21-month lexical processing efficiency, possibly due to changes in the structure of the infant lexicon, and at best subtle effects of lexical processing efficiency at 18 months on 21-month vocabulary. Although the results suggest that data fit either central capacity or emergent capacity accounts of infant lexical processing efficiency, the longitudinal analysis provides greater evidence for a developmental account where lexical processing efficiency is closely tied to and predicted by infants' vocabulary knowledge. Further research should examine how lexical network structure changes over the second year of life and beyond.

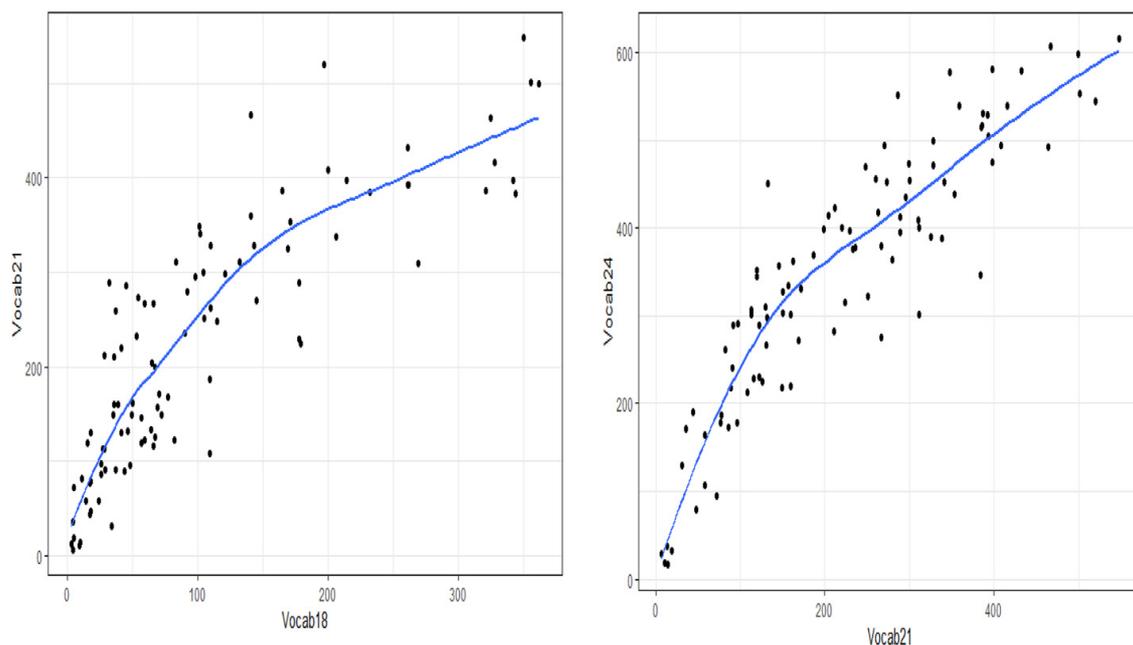## Acknowledgment

## Appendix A

See Fig. A1.



**Fig. A1.** Left: Relationship between 18-month vocabulary and 21-month vocabulary. Right: Relationship between 21-month vocabulary and 24-month vocabulary.

# References

Bainter, S. A., & Bollen, K. A. (2014). Interpretational confounding or confounded interpretations of causal indicators?. *Measurement: Interdisciplinary Research & Perspectives, 12*, 125–140.

Bates, E., Bretherton, I., & Snyder, L. (1988). *From first words to grammar: Individual differences and dissociable mechanisms*. New York: Cambridge University Press.

Bates, E., Dale, P., & Thal, D. (1995). Individual differences and their implications for theories of language development. In P. Fletcher & B. MacWhinney (Eds.), *Handbook of child language* (pp. 96–151). Oxford, UK: Basil Blackwell.

Bornstein, M. H., Hahn, C., & Putnick, D. L. (2016). Long-term stability of core language skill in children with contrasting language skills. *Developmental Psychology, 52*, 704–716.

Borovsky, A., Ellis, E. M., Evans, J. L., & Elman, J. L. (2016). Lexical leverage: Category knowledge boosts real-time novel word recognition in 2-year-olds. *Developmental Science, 19*, 918–932.

Brito, N. H., Fifer, W. P., Myers, M. M., Elliott, A. J., & Noble, K. G. (2016). Associations among family socioeconomic status, EEG power at birth, and cognitive skills during infancy. *Developmental Cognitive Neuroscience, 19*, 144–151.

Chen, Q., & Mirman, D. (2012). Competition and cooperation among similar representations: Toward a unified account of facilitative and inhibitory effects of lexical neighbors. *Psychological Review, 119*, 417–430.

Cowan, N., Rouder, J. N., Blume, C. L., & Scott Saults, J. (2012). Models of verbal working memory capacity: What does it take to make them work? *Psychological Review, 119*, 480–499.

Cristia, A., Seidl, A., Junge, C. M. M., Soderstrom, M., & Hagoort, P. (2014). Predicting individual variation in language from infant speech perception measures. *Child Development, 85*, 1330–1345.

Fenson, L., Marchman, V. A., Thal, D. J., Dale, P. S., Reznick, J. S., & Bates, E. (2007). *MacArthur–Bates Communicative Development Inventories: User's guide and technical manual* (2nd ed.). Baltimore, MD: Paul H. Brookes.

Fernald, A., & Marchman, V. A. (2012). Individual differences in lexical processing at 18 months predict vocabulary growth in typically developing and late-talking toddlers. *Child Development, 83*, 203–222.

Fernald, A., Marchman, V. A., & Weisleder, A. (2013). SES differences in language processing skill and vocabulary are evident at 18 months. *Developmental Science, 16*, 234–248.

Fernald, A., Perfors, A., & Marchman, V. A. (2006). Picking up speed in understanding: Speech processing efficiency and vocabulary growth across the 2nd year. *Developmental Psychology, 42*, 98–116.

Fernald, A., Zangl, R., Portillo, A. L., & Marchman, V. A. (2008). Looking while listening: Using eye movements to monitor spoken language comprehension by infants and young children. In I. A. Sekerina, E. M. Fernández, & H. Clahsen (Eds.), Language acquisition and language disorders, Vol. 44: Developmental psycholinguistics: On-line methods in children's language processing (pp. 97–135). Amsterdam: John Benjamins.

Frank, M. C., Braginsky, M., Yurovsky, D., & Marchman, V. A. (2017). Wordbank: An open repository for developmental vocabulary data. *Journal of Child Language, 44*, 677–694.

Hintz, F., Jongman, S. R., Dijkhuis, M., van't Hoff, V., McQueen, J. M., & Meyer, A. S. (2019). Shared lexical access processes in speaking and listening? An individual differences study. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. https://doi.org/10.1037/xlm0000768.

Hipp, J. R., Bauer, D. J., & Bollen, K. A. (2005). Conducting tetrad tests of model fit and contrasts of tetrad-nested models: A new SAS macro. *Structural Equation Modeling, 12*, 76–93.

Houston-Price, C., Mather, E., & Sakkalou, E. (2007). Discrepancy between parental reports of infants' receptive vocabulary and infants' behaviour in a preferential looking task. *Journal of Child Language, 34*, 701–724.

Howell, R. D. (2014). What is the latent variable in causal indicator models?. *Measurement: Interdisciplinary Research & Perspectives, 12*(4), 141–145.

Hurtado, N., Marchman, V. A., & Fernald, A. (2008). Does input influence uptake? Links between maternal talk, processing speed and vocabulary size in Spanish-learning children. *Developmental Science, 11*, F31–F39.

Johnson, E. K. (2016). Constructing a proto-lexicon: An integrative view of infant language development. *Annual Review of Linguistics, 2*, 391–412.

Kail, R., & Salthouse, T. A. (1994). Processing speed as a mental capacity. *Acta Psychologica, 86*, 199–225.

Kidd, E., Donnelly, S., & Christiansen, M. H. (2018). Individual differences in language acquisition and processing. *Trends in Cognitive Sciences, 22*, 154–169.

Kidd, E., Junge, C., Spokes, T., Morrison, L., & Cutler, A. (2018). Individual differences in infant speech segmentation: Achieving the lexical shift. *Infancy, 23*, 770–794.

Lany, J. (2018). Lexical-processing efficiency leverages novel word learning in infants and toddlers. *Developmental Science, 21*, e12569.

Lany, J., Giglio, M., & Oswald, M. (2018). Infants' lexical processing efficiency is related to vocabulary size by one year of age. *Infancy, 23*, 342–366.

Logan, G. D. (1988). Toward an instance theory of automatization. *Psychological Review, 95*, 492–527.

MacDonald, M., & Christiansen, M. (2002). Reassessing working memory: Comment on Just and Carpenter (1992) and Waters and Caplan (1996). *Psychological Review, 109*, 35–54.

Mainz, N., Shao, Z., Brysbaert, M., & Meyer, A. S. (2017). Vocabulary knowledge predicts lexical processing: Evidence from a group of participants with diverse educational backgrounds. *Frontiers in Psychology, 8*. https://doi.org/10.3389/fpsyg.2017.01164.

Mani, N., & Plunkett, K. (2010). In the infant's mind's ear: Evidence for implicit naming in 18-month-olds. *Psychological Science, 21*, 908–913.

Mani, N., & Plunkett, K. (2011). Phonological priming and cohort effects in toddlers. *Cognition, 121*, 196–206.

Markus, K. A. (2014). Unfinished business in clarifying causal measurement: Commentary on Bainter and Bollen. *Measurement: Interdisciplinary Research & Perspectives, 12*, 146–150.

McCoach, D. B., & Kenny, D. A. (2014). A few thoughts on the similarities and the differences between causal or reflective indicators of latent variables. *Measurement: Interdisciplinary Research & Perspectives, 12*, 151–154.

Peter, M. S., Durrant, S., Jessop, A., Bidgood, A., Pine, J. M., & Rowland, C. F. (2019). Does speed of processing or vocabulary size predict later language growth in toddlers?. *Cognitive Psychology, 115*, 101238.

Rämä, P., Sirri, L., & Serres, J. (2013). Development of lexical-semantic language system: N400 priming effect for spoken words in 18- and 24-month-old children. *Brain and Language, 125*, 1–10.

Reilly, S., Bavin, E. L., Bretherton, L., Conway, L., Eadie, P., Cini, E., ... Wake, M. (2009). The Early Language in Victoria Study (ELVS): A prospective, longitudinal study of communication skills and expressive vocabulary development at 8, 12 and 24 months. *International Journal of Speech and Language Pathology, 11*, 344–357.

Reilly, S., Wake, M., Bavin, E. L., Prior, M., Williams, J., Bretherton, L., ... Ukoumunne, O. C. (2007). Predicting language at 2 years of age: A prospective community study. *Pediatrics, 120*, e1441–e1449.

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software, 48*(2), 1–36.

Smithson, M., & Verkuilen, J. (2006). A better lemon squeezer? Maximum-likelihood regression with beta-distributed dependent variables. *Psychological Methods, 11*, 54–71.

Storkel, H. L., & Lee, S. (2011). The independent effects of phonotactic probability and neighbourhood density on lexical acquisition by preschool children. *Language and Cognitive Processes, 26*, 191–211.

Styles, S., & Plunkett, K. (2009). What is "word understanding" for the parent of a one-year-old? Matching the difficulty of a lexical comprehension task to parental CDI report. *Journal of Child Language, 36*, 895–908.

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

Wang, J., Engelhard, G., & Lu, Z. (2014). Clarifying the conceptualization of indicators within different models. *Measurement: Interdisciplinary Research & Perspectives, 12*, 155–159.

Weisleder, A., & Fernald, A. (2013). Talking to children matters: Early language experience strengthens processing and builds vocabulary. *Psychological Science, 24*, 2143–2152.

West, S. G., & Grimm, K. J. (2014). Causal indicator models: Unresolved issues of construction and evaluation. *Measurement: Interdisciplinary Research & Perspectives, 12*, 160–164.

Widaman, K. F. (2014). Much ado about nothing—or at best, very little. *Measurement: Interdisciplinary Research & Perspectives, 12*, 165–168.

Willoughby, M. T., Blair, C. B., & Family Life Project Investigators (2016). Measuring executive function in early childhood: A case for formative measurement. *Psychological Assessment, 28*, 319–330.

Willoughby, M., Holochwost, S. J., Blanton, Z. E., & Blair, C. B. (2014). Executive functions: Formative versus reflective measurement. *Measurement: Interdisciplinary Research & Perspectives, 12*, 69–95.