

# Production Practice During Language Learning Improves Comprehension



Elise W. M. Hopman  and Maryellen C. MacDonald

Department of Psychology, University of Wisconsin–Madison

Psychological Science  
2018, Vol. 29(6) 961–971  
© The Author(s) 2018  
Reprints and permissions:  
sagepub.com/journalsPermissions.nav  
DOI: 10.1177/0956797618754486  
www.psychologicalscience.org/PS



## Abstract

Language learners often spend more time comprehending than producing a new language. However, memory research suggests reasons to suspect that production practice might provide a stronger learning experience than comprehension practice. We tested the benefits of production during language learning and the degree to which this learning transfers to comprehension skill. We taught participants an artificial language containing multiple linguistic dependencies. Participants were randomly assigned to either a production- or a comprehension-learning condition, with conditions designed to balance attention demands and other known production–comprehension differences. After training, production-learning participants outperformed comprehension-learning participants on vocabulary comprehension and on comprehension tests of grammatical dependencies, even when we controlled for individual differences in vocabulary learning. This result shows that producing a language during learning can improve subsequent comprehension, which has implications for theories of memory and learning, language representations, and educational practices.

## Keywords

language production, language learning, language comprehension, artificial language learning, learning transfer, open data, open materials

Received 6/14/17; Revision accepted 12/28/17

Imagine the first day of a foreign language course requiring students to speak in the new language immediately. In this curriculum, students do not simply repeat words but must generate whole grammatical sentences within the first hour. Of course this intense production experience should improve production abilities. More surprising is our hypothesis that production practice will yield improved vocabulary and grammatical comprehension abilities compared with an intense comprehension-focused curriculum. If so, such results would have implications for language pedagogy as well as theories of memory, learning transfer, and language representations.

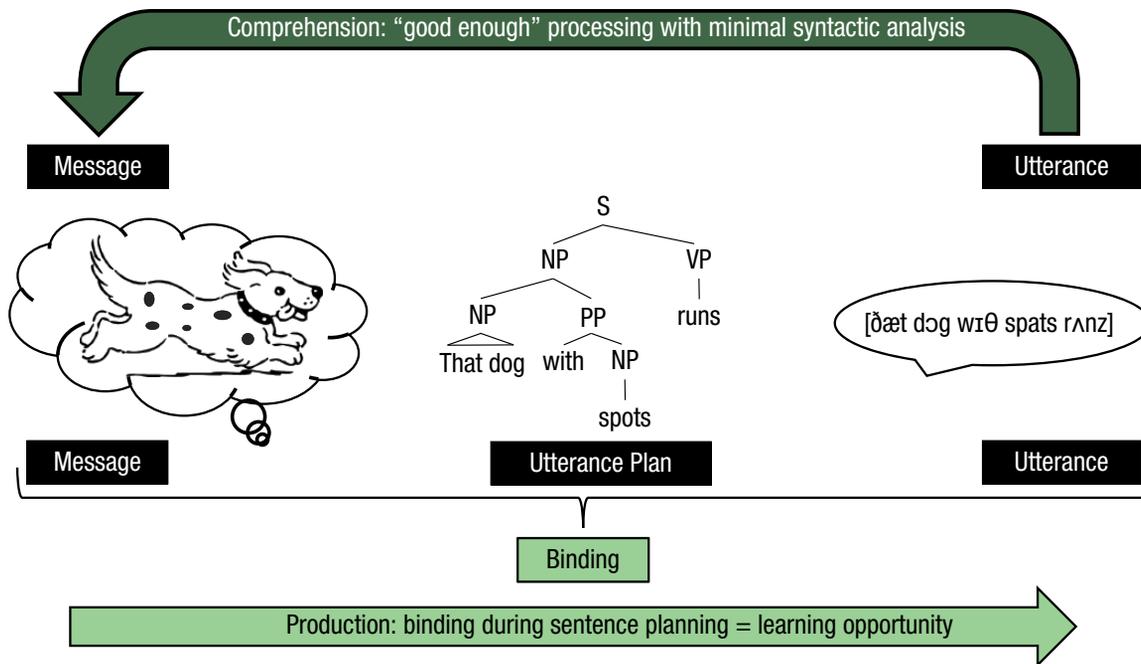
Views of the relationship between production and comprehension, and thus the potential for transfer among tasks, range from full involvement of production in comprehension tasks (Pickering & Garrod, 2013) to minimally overlapping systems (Grodzinsky, 2000). More generally, research on learning transfer in non-linguistic domains shows that some learning does not transfer to new task demands, even with identical materials (Green, Kattner, Siegel, Kersten, & Schrater, 2015). Language comprehension is known to affect native

language production (Bock, Dell, Chang, & Onishi, 2007; Montag & MacDonald, 2015), but evidence for the reverse is scant and conflicting (Branigan, Pickering, & McLean, 2005; Segaert, Menenti, Weber, Petersson, & Hagoort, 2011). The influential input hypothesis in second-language acquisition (Krashen, 2003) claims that language production practice does not benefit language learning, and related research finds that comprehension practice improves second-language production but not vice versa (VanPatten, 2013; VanPatten & Cadierno, 1993). Some studies suggest that speech production practice can impair perception (Baese-Berk & Samuel, 2016; Leach & Samuel, 2007), while other similar studies show benefits (Bixby, 2017). Taken together, these findings suggest that comprehension skill would be best developed with comprehension training, not production training.

---

## Corresponding Author:

Elise W. M. Hopman, Department of Psychology, University of Wisconsin–Madison, 618 Brogden Hall, 1202 West Johnson St., Madison, WI 53706  
E-mail: hopman@wisc.edu



**Fig. 1.** Language processing. Language production is the act of turning an idea into a structured utterance, involving generation of structure from long-term knowledge, which relies on recall from memory. Temporary maintenance of word order, bound to the conceptual representation of the message, could be a route for improved learning of multi-word dependencies. During comprehension, perceivers may settle for a “good enough” interpretation without a detailed analysis of all syntactic dependencies, which is possible with mostly recognition-based processing.

In contrast, memory researchers have found that production can boost some types of learning. Production may improve learning in several different ways. First, language production provides the opportunity to both produce and then hear one’s own speech, providing both an additional presentation of the material and an alternative modality for encoding it (MacLeod & Bodner, 2017). Second, language production is more attention-demanding than comprehension (Boiteau, Malone, Peters, & Almor, 2014), potentially yielding greater depth of processing, with consequent memory benefits ( Craik & Tulving, 1975). Relatedly, production necessitates choices of what to say, and making task-relevant choices improves learning (Carter & Ste-Marie, 2017). Third, comprehension involves recognizing a stimulus, whereas production involves recall from memory, which benefits information retention—the testing effect (Roediger & Karpicke, 2006). In fact, there is even some evidence for learning transfer from recall (production) to recognition (comprehension; Wenger, Thompson, & Bartling, 1980). Furthermore, retrieval practice can guide learning by changing subsequent long-term memory representations (Fan & Turk-Browne, 2013).

These inherent differences between production and comprehension suggest that production experience should improve language learning. Karpicke and Roediger (2008) found that participants learning Swahili–English word pairs benefited significantly more from

repeated retrieval (recall) practice than repeated studying (recognition). However, it is unknown whether production benefits extend beyond vocabulary learning. For example, the sentence “That dog with spots runs” conveys its meaning via grammatical elements, word order, and number agreement. Both *dog* and *that* are singular, and *dog* agrees with *runs*, even though this noun and verb are nonadjacent. These features reflect both the hierarchical and sequential structure of languages, and we predict that language production is likely a strong learning tool here (Fig. 1). Beyond the learning benefits described above, production involves planning the serial order of words, which engages serial ordering mechanisms well known in working memory tasks (MacDonald, 2016). Comprehension is more variable: It may sometimes include careful syntactic analysis but often is “good enough,” giving limited attention to syntax and relying on other cues to meaning (Ferreira & Patson, 2007). This shallower processing may provide a poor learning environment for syntactic dependencies compared with the sequential processing inherent in language production.

The current study investigated the potential benefits of production in a between-subjects manipulation of learning task (comprehension or production), followed by comprehension tests. Our tasks minimize some well-known production–comprehension learning differences, allowing a focus on inherent processing differences

(Fig. 1). Participants learned an artificial language incorporating a strict word order, complex word morphology, and grammatical dependencies across words. We hypothesized that production experience, compared with comprehension experience, would yield improved vocabulary comprehension. Moreover, we hypothesized that the production-learning group would have improved comprehension of grammatical dependencies, even when we controlled for vocabulary comprehension.

## Method

### Participants

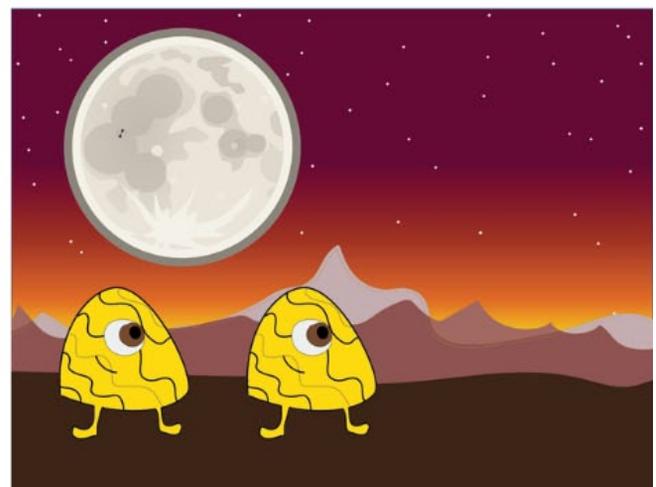
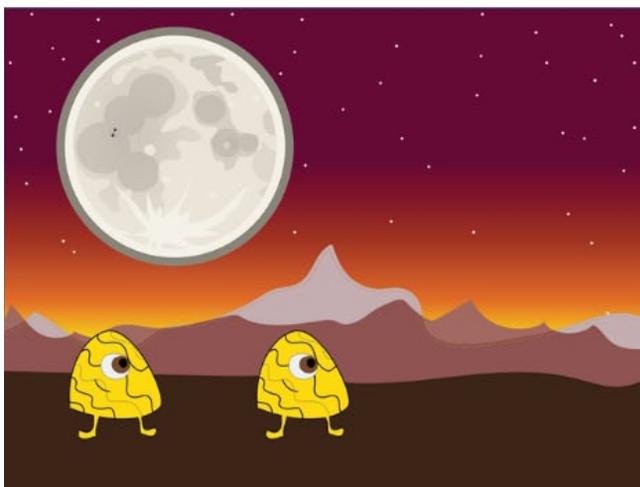
A total of 125 native speakers of English from the University of Wisconsin–Madison received course credit or payment for participating. Based on a power analysis and pilot testing, the goal was 100 participants (50 per learning condition) who scored above threshold on a vocabulary test. To reach this threshold with leeway to remove nonperforming participants, we assigned 62 participants to the comprehension-learning condition and 63 to the production-learning condition. Three participants (two comprehension, one production) were unable to finish the experiment.

### Materials

**Language and visual world.** A cartoon world of monsters situated on alien landscapes was created, including both still pictures and short videos. A language describing the entities, locations, and actions in the world contained 20 root words and four suffixes. All sentences had the structure shown in Figure 2. See the Supplemental Material for more detail on the materials. All materials, including the code to run the experiment in PsychoPy (Peirce, 2007), are available online (<https://osf.io/74kqe/#>).

**Dependencies.** The language contained two deterministic agreement dependencies. Suffixes on four word types (determiner, adjective, monster, and verb) varied with the noun number (singular, plural) and monster type (kind-looking, scary-looking), a semantic category similar to gender or classifier morphology in some natural languages. The *usu* suffix in Figure 2 indicates a kind-looking monster (the *us* part of the suffix) and plural with the final *u* (Table 1).

We also introduced a probabilistic dependency, in which monsters tended to be marked with either striped or dotted markings more frequently based on their semantic type. We explored the possibility that production



<b><u>V</u>usu</b>	<b>F<u>u</u>mu<u>s</u>u</b>	<b>Teep<u>u</u>s<u>u</u></b>	<b>Traw</b>	<b>Ot</b>	<b><u>D</u>iv<u>u</u>s<u>u</u></b>	<b>Kredel</b>
Determiner	Adjective	Noun	Markings	Preposition	Verb	Location
The-k-pl	Yellow-k-pl	Teep-k-pl	Curves	With	Grow-k-pl	Mountain

The two kind yellow teeps with curved lines grow bigger at the location with the mountain.

**Fig. 2.** Artificial world: the first and last frames of a video, the sentence describing it, the grammatical category of each word, the English translation of each word (k-pl = kind-looking, plural) and the sentence in English. Agreeing suffixes are underlined. In the experiment, the assignment of words to meanings was different for each participant, and the language was always auditorily presented, never written.

**Table 1.** The Four Possible Suffixes and Their Meaning

	Singular	Plural
Kind	-us	-usu
Scary	-ok	-oko

experience would boost such learning, but there was no evidence of any learning of this dependency in our study. While we cannot interpret these null results, it is noteworthy that Amato and MacDonald (2010) found learning of a similar probabilistic dependency with sensitive reading measures but not with measures similar to ones used in the present study. Details about the probabilistic dependency manipulation and results for it can be found in the Supplemental Material.

### Training procedure

Training (Fig. 3) consisted of blocks of passive-exposure trials, interleaved with blocks of either active-comprehension trials (comprehension-learning condition) or active-production trials (production-learning condition). All participants received 78 passive-exposure trials divided into 14 blocks of 2 to 6 trials each in which a picture or video was paired with two auditory presentations of a word, phrase, or sentence in the language that matched the image (Fig. 3d; see Table S1 in the Supplemental Material for more details). For all participants, language training began with a block of still pictures of uncolored, unmarked monsters described by single words, and new vocabulary was gradually added in each block until all elements were combined into full sentences, as in Figure 2. All participants received 96 active learning trials, each divided into 17 blocks of 2 to 6 trials of the same type.

**Active-comprehension task.** In the *active-comprehension* task (Fig. 3e), participants saw a picture and heard a phrase in the novel language, and they indicated with a keypress whether the phrase matched the picture. Half of the trials in each block were mismatches. Feedback on response accuracy was presented onscreen. Regardless of accuracy, feedback was followed by a repetition of the auditory phrase, accompanied by its matching picture.

**Active-production task.** The *active-production* (Fig. 3f) task prompted participants to describe a picture aloud in the artificial language. Responses were recorded. Participants pressed a key after speaking, then heard the phrase that correctly described the picture. The picture remained onscreen throughout the trial, and the correct phrase was presented independent of the accuracy of their production.

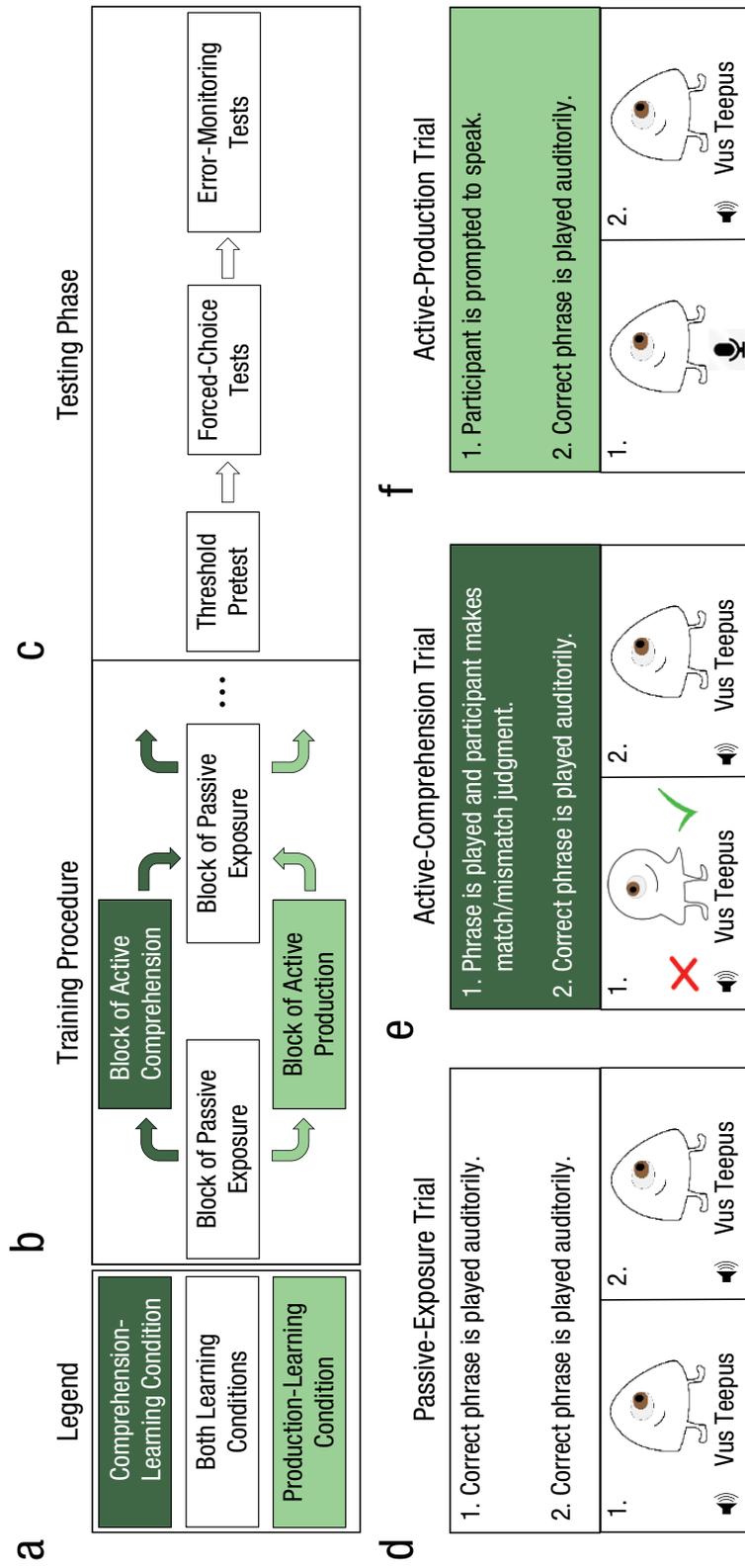
Language production and comprehension differ in many dimensions, but our procedure reduced some of these differences. First, the amount of listening experience (Factor 1, Table 2) was more balanced than in typical comprehension/production comparisons: Comprehension participants heard a phrase that sometimes matched the picture, and production participants heard their own production, which also often was not correct. Furthermore, all participants heard the correct phrase after they made a judgment or said a phrase, providing them with a correct pairing of language and picture. The tasks were also designed to minimize differences in attention and task-relevant choices (Factor 2, Table 2), as both tasks required an overt response to a picture. Comprehension trials involved a match/mismatch choice, whereas production trials involved more open-ended production choices. Both tasks thus substantially differed from the passive-exposure trials, which required no response.

The two learning tasks still capture important inherent differences between production and comprehension. Production involves recall, whereas comprehenders, especially in naturalistic learning settings where a context is provided, can rely on recognition (Fig. 1), with known consequences for vocabulary learning (Roediger & Karpicke, 2006). In order to investigate the benefits of production beyond vocabulary learning, we controlled for potential differences in vocabulary learning between the two conditions in the two ways described in Factor 3, Table 2.

### Testing phase

**Threshold pretest.** After training was completed, participants were assessed on their learning of the content words of the artificial language, to exclude low-performing participants (Factor 3, Table 1). The test consisted of 18 trials in which one word was presented together with two pictures of the same category (e.g., two monsters), testing all content words of the language. Comprehension participants ( $M = 16.5$ ,  $SD = 2.1$ ) and production participants ( $M = 16.3$ ,  $SD = 2.2$ ) did not differ in accuracy on this test,  $t(120) < 1$ . On the basis of pilot testing, we set a threshold of 15/18 correct (83%) for inclusion in the main analyses. A total of 52 out of 60 comprehension participants and 52 out of 62 production participants met this threshold. All further analyses reported here included data only from these 104 participants scoring above threshold. However, results remained the same when data from the 18 low-scoring participants were included.

**Forced-choice tests.** All participants completed forced-choice comprehension trials, similar in format to the comprehension group's active-comprehension trials during



**Fig. 3.** Flow of experimental procedure. Training consisted of 31 blocks of alternating passive- and active-learning trials (see Table S1 in the Supplemental Material for more details). After training, participants completed three tests of learning. The two learning conditions were identical in passive-exposure blocks and all comprehension tests; the groups' experience differed only in the active-learning blocks.

**Table 2.** Three Factors That Differ Between Comprehension and Production

Factor	Explanation	Methods to reduce factor in current study
1. Double experience	Every production yields perception experience (e.g., hearing oneself talk).	Comprehension and production conditions both involved hearing a phrase that did/did not match a picture, then hearing correct matching phrase.
2. Attention and task-relevant choices	Production is more attention demanding than comprehension and inherently involves making choices about what to say.	Active task involving task-relevant choices in both production and comprehension training.
3. Improved vocabulary learning in production	Comprehension requires recognition of the linguistic signal, while production requires recall, which has been shown to improve vocabulary learning. Any investigation of effects on grammar learning should accommodate potential vocabulary learning differences.	Performance threshold for participant inclusion; vocabulary score as a covariate to test additional benefit of production on grammar learning beyond any benefit for vocabulary learning.

learning. In each trial, participants saw two pictures on the screen and heard a phrase. They were instructed to choose the picture matching the phrase as quickly as possible, using a keypress, which ended the trial. The dependent variables for these tests were accuracy and reaction time (RT), measured from the onset of the first word in the auditory phrase that could be used to identify the correct picture (words marked with arrows in Fig. 4). Because the participant could respond at any point in the trial, responses occurring before this critical word were recorded as negative RTs. Each trial assessed a particular aspect of language knowledge (vocabulary, suffix meaning, etc.) by virtue of the contrast between the target and the foil picture, and trials were randomly intermixed.

*Vocabulary test (18 trials).* Participants heard a phrase and chose between two pictures that differed in only the meaning of one critical content word. In the example in Figure 4a, participants heard a five-word phrase and chose between two pictures differing only in color (Word 2 of the phrase). As in the threshold pretest, all 18 content words of the language were tested as trial-critical words. Unlike in the pretest, these test items were embedded in full sentences, yielding a difficult auditory sentence comprehension task. We nonetheless expected the groups to perform similarly, as low-performing participants had been excluded by the pretest. Our aim with these trials was both to compare vocabulary and auditory comprehension across groups and also to provide a covariate (vocabulary score) that would allow us to examine learning of grammatical features across groups while controlling for vocabulary learning (Factor 3, Table 2). Foils in this test were always within semantic type for monsters and within marking type, so that knowledge of the suffixes and probabilistic monster-marking regularity could not help choose the right picture.

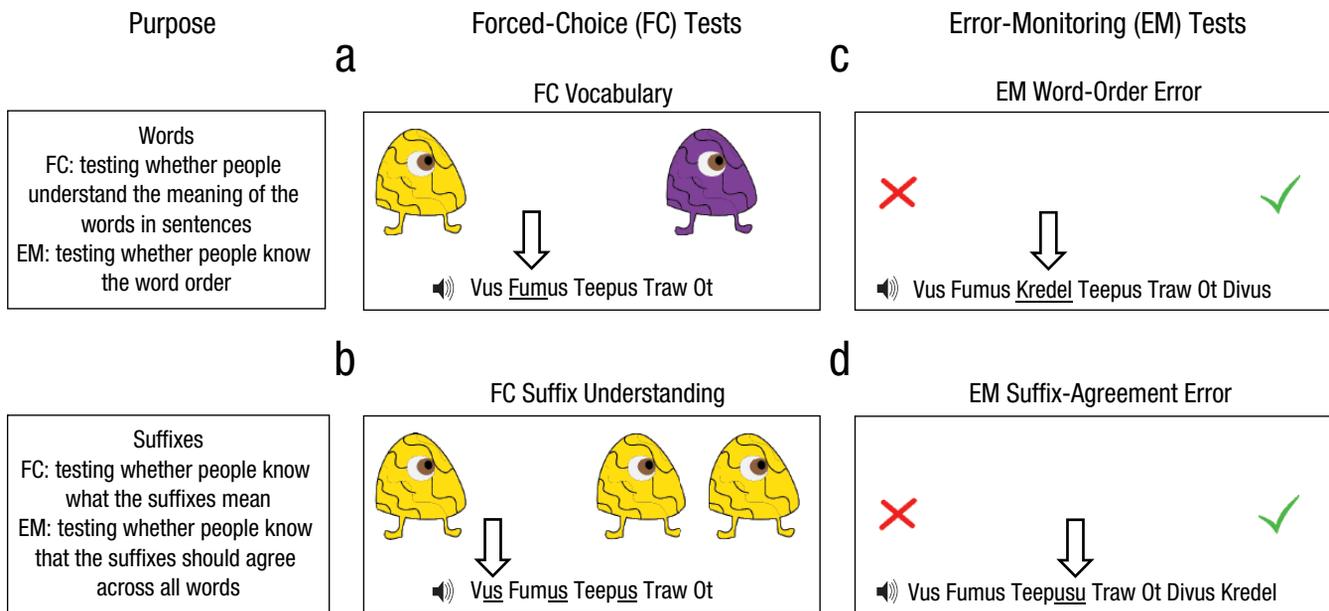
*Suffix understanding test (24 trials).* Participants heard a phrase and had to choose between two pictures that

differed either in monster number (12 items; example in Fig. 4b) or in semantic monster type (12 items). Because the monster word is preceded by two suffixed words (determiner and color) that carry number and semantic information, it is possible to identify the correct picture before hearing the monster word (which also conveys the correct response). The resulting within-subject predictor for number/semantic items did not interact with our main learning condition predictor, and is thus further discussed only in the Supplemental Material.

*Error-monitoring tests.* An error-monitoring task assessed participants' sensitivity to violations of language patterns. This test differed from both learning conditions in that there was no picture presented, but the participant's task, judging the correctness of a sentence, was similar to the judgment task in the active-comprehension condition. Trials assessing word order and suffix agreement (Fig. 4) were randomly intermixed with 44 grammatically correct sentences. None of these sentences had been presented during training, nor had the correct version for sentences with errors been presented in training. Participants heard a sentence and pressed a key as quickly as possible to indicate whether the sentence contained an error or was grammatical. The dependent variables were accuracy and RT. For each trial, the critical word was defined as the first word that was incorrect (words marked with an arrow in Fig. 4c and Fig. 4d); in fully correct sentences, the critical word was the last word. Participants occasionally responded before hearing the critical word, leading to some negative RTs.

*Word-order error test (32 trials).* We included trials with four different ungrammatical word orders, each of which had one word in an ungrammatical position (example in Fig. 4c).

*Suffix-agreement error test (48 trials).* Participants heard sentences with different kinds of agreement errors, with



**Fig. 4.** Overview of main tests. Participants never saw the language written, they heard only auditory phrases. In the forced-choice tests (a, b), participants heard a phrase and chose between the two pictures. In the error-monitoring tests (c, d), participants heard a sentence and made a grammaticality judgment. Underlining and arrows indicate the critical word (or words) for the participant's response.

one suffix that did not match the other three in the sentence. In the example in Figure 4d, the mismatching suffix *usu* is plural, whereas the other suffixes are *us*, which is singular. The within-subject predictors for location of the mismatching suffix never interacted with our main learning condition predictor, and so error type is discussed only in the Supplemental Material.

**Predictions.** Because of the enhanced serial processing requirements of production, we expected the production group to outperform the comprehension group on tasks with a serial dependency: both word order and suffix agreement across words. If transfer does not occur, the comprehension group should outperform the production group, both because all tests assessed comprehension and because the testing procedures were more similar to the tasks performed in the comprehension-learning group than the production group.

## Results

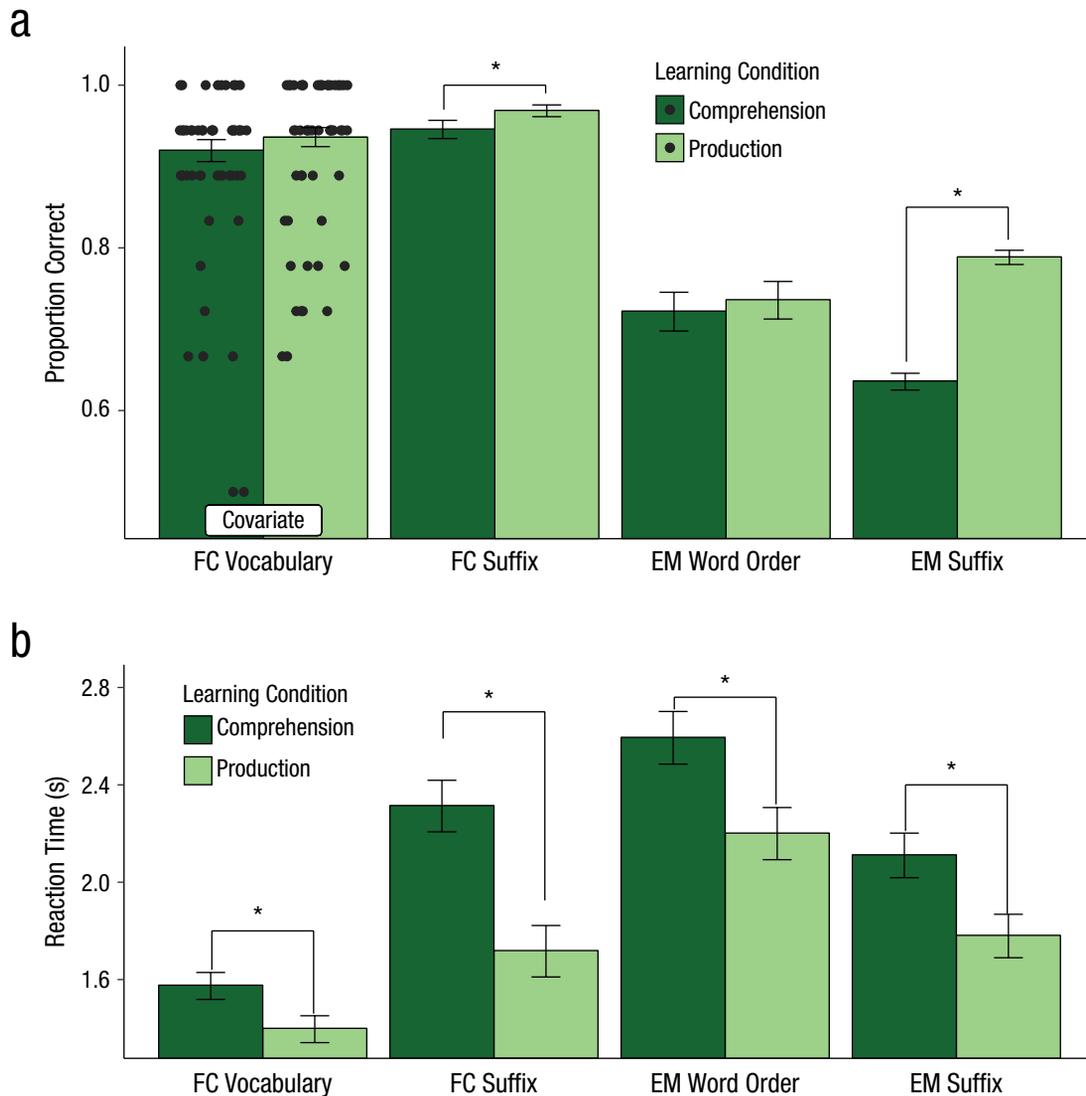
### Data processing

RTs were analyzed with mixed effects regression analyses in R (R Core Team, 2016). Accuracy data were analyzed with mixed effects logistic regression analyses using the lme4 package (Bates, Mächler, Bolker, & Walker, 2015). No trials were removed for the accuracy analysis. In the RT analysis, trials in which the participant responded incorrectly or before the critical word

(negative RTs) and RTs more than 3 *SDs* above a participant's own mean were removed, leaving 78% of trials for the RT analysis. Following Barr, Levy, Scheepers, and Tily (2013), we initially included a random intercept by participants as well as random slopes by participants for all within-subjects predictors in the regression models. The model for accuracy in the agreement error test did not converge, so we gradually simplified it (Barr et al., 2013), leading to a model without an intercept but with random slopes by participants; see the Supplemental Material for all statistical models and their outputs. Model predictions for the learning condition predictor for each test are plotted in Figure 5. They are based on the full model but collapsed over other predictors by taking the average for all within-subject predictors, because our main manipulation of learning condition (production versus comprehension) never interacted with any within-subjects predictors; those predictors are not discussed further. All data and analyses are freely available online (<https://osf.io/bbf3c>).

### Forced-choice tests

Comprehension participants and production participants did not significantly differ in accuracy on the forced-choice vocabulary test. However, in relation to proportion correct, there was a range of individual differences (Fig. 5a, leftmost bars). Not surprisingly, performance on this task (vocabulary score) was a reliable



**Fig. 5.** Overview of comprehension-test results: proportion of correct responses (a) and mean reaction time (b) as a function of test type and learning condition. For the forced-choice (FC) vocabulary test, the dots represent the proportion of correct responses for individual participants, which was used as a covariate in all other regression analyses. Asterisks indicate significant differences between learning conditions ( $p < .05$ ). Error bars shown 95% confidence intervals. Table S2 in the Supplemental Material shows the regression models for all tests. EM = error monitoring.

predictor of accuracy and RT on almost all other tests, indicating that word comprehension in auditory phrases is associated with higher accuracy and shorter RTs on other auditory comprehension tests. Specific results for each test can be found in the results table (Table S2 in the Supplemental Material). Importantly, because we included each participant's score as a covariate in all subsequent analyses, all further regression results hold true over and above potential vocabulary score differences between participants.

Production participants were significantly faster than comprehension participants on the vocabulary test items they answered correctly. Production participants were also significantly more accurate and had shorter RTs

than comprehension participants on suffix understanding items, which is evidence that production participants had a better understanding of what the suffixes mean.

### ***Error-monitoring tests***

We calculated a  $d'$  score for each participant, reflecting their sensitivity to grammar (discriminating between correct vs. incorrect word order and agreement), and compared  $d'$  scores between learning conditions. Production participants ( $M = 2.4$ ,  $SD = 1.1$ ) were overall significantly more sensitive than comprehension participants ( $M = 1.8$ ,  $SD = 1.1$ ) across the two error types  $t(102) = 2.36$ ,  $p = .020$ .

Separate word order and suffix-agreement tests yielded similar results, with production participants outperforming comprehension participants in both accuracy and speed, with the exception of no reliable differences in accuracy for the word-order test.

## Discussion

Production-focused training yielded superior learning and comprehension of a novel language, across a variety of language features and task demands, compared with training focusing on comprehension itself. Importantly, production's learning advantage went beyond the word level: even after we controlled for vocabulary knowledge, production participants were both faster and more accurate on tests of grammar comprehension.

The balancing of production and comprehension conditions allows us to take steps in identifying possible mechanisms underlying production's beneficial effects. Although lexical retrieval (the recall of words from long-term memory) is likely to be a powerful component of production's learning benefits, other aspects of utterance planning may also be important contributors. Language production begins with a conceptual message that the producer aims to communicate. This message, fully known to the producer, is activated throughout utterance planning and execution, promoting binding over all parts of the utterance (Savill et al., 2017; Fig. 1). This situation should afford a stronger learning opportunity than in comprehension, where the input signal and the message that the comprehender gleans from it unfold over time. Language production also requires the generation of an utterance plan, and because planning precedes execution by some time, planning entails maintaining information in working memory (Brown-Schmidt & Konopka, 2015). Indeed, MacDonald (2016) argued that the utterance plan is the maintenance portion of verbal working memory. The temporary maintenance, serial ordering, and binding across the different linguistic levels that occur during utterance planning provide benefits for learning interword grammatical, conceptual, and phonological relationships. These relationships may underlie our finding that production benefits grammatical learning beyond vocabulary knowledge. Our control for vocabulary knowledge in grammar learning is a first step to exploring the different kinds of learning opportunities that production processes afford.

Our results constrain theoretical positions on verbal memory and learning in several ways. First, they show that the benefit of production on language learning need not depend on an additional potential-learning experience in the form of hearing oneself speak (see MacLeod & Bodner, 2017, for other comprehension–production differences in word lists). The current study

balanced listening experience across learning conditions and still found benefits for production over comprehension-learning tasks. Second, our results expand the reach of the testing effect (Roediger & Karpicke, 2006): We posit that language production inherently has important learning benefits that have been associated with testing. Full language production involves recall of words from long-term memory and assembly of sentence structure, whereas comprehension relies more heavily on recognition. A more limited production task, repetition of another's utterance, does not require full generation of language from memory and appears to have reduced learning benefits in vocabulary learning compared with full, generative production (Kang, Gollan, & Pashler, 2013; Middleton, Schwartz, Rawson, & Garvey, 2015). Third, we extend for the first time a production-based testing effect beyond single words, and show that language production is superior to comprehension training in learning and comprehension of grammar, even after imposing controls for differences in word knowledge. Consistent with our result with spoken language, there is evidence that retrieval practice improves conceptual learning from texts (Karpicke & Blunt, 2011), also suggesting an important role for retrieval/production practice in relational learning, whether it is making inferences about concepts or learning grammatical regularities. Fourth, our results corroborate findings that spelling practice on difficult written words improves reading speed on these words (Ouellette, Martin-Chang, & Rossi, 2017); though not explicitly phrased as such, these findings provide another example of production practice improving comprehension in a different but related modality.

The finding that production training improves subsequent comprehension performance more than comprehension practice itself provides a clear example of learning transfer, where experience with one task yields subsequent benefits on a different task (Fan, Turk-Browne, & Taylor, 2016; Green et al., 2015). This transfer effect is most readily understood as reflecting shared representations between comprehension and production. Future work should examine the extent to which benefits for production extend to other levels of language perception and comprehension beyond the lexical and grammatical aspects studied here, because evidence is mixed concerning benefits and costs to production at the level of speech sound perception (Baese-Berk & Samuel, 2016; Bixby, 2017; Leach & Samuel, 2007).

Our findings also have implications for language instruction, including Krashen's (2003) input hypothesis, which holds that language learning is driven by comprehension practice, not production. Studies of classroom language learning have supported this claim, showing that comprehension practice leads to better

production performance, but not vice versa (VanPatten & Cadierno, 1993). These results initially seem to directly contradict our own, but a key difference is in how “production” is instantiated. Whereas Krashen and colleagues associate production with repetition of teacher input and spoken grammar drills, the production learning in our experiment involved generation of meaningful language, and we showed that such practice is beneficial. Given the widespread recurrence of the mantra that comprehension is better than production practice in second-language instruction (Krashen, 2003; VanPatten, 2013; VanPatten & Cadierno, 1993), it will be important to distinguish repetition from more generative production in both future research and in recommendations to instructors.

This work may also illuminate effects of child production and comprehension in first language acquisition. Children from economically disadvantaged households tend to have reduced language experience compared with those in more affluent households, with consequences for vocabulary development and educational outcomes (Hart & Risley, 1995). While differences are commonly framed in terms of comprehension—the “30-million-word gap” in the amount of language the child hears—other studies suggest a key role for the child’s own production experience. Zimmerman et al. (2009) found that the number of turns in adult–child conversations was a better predictor of language development than language input (comprehension experience). In conversational exchanges, the child not only hears adult input but also produces language. Beyond other stimulating and engaging aspects of conversational turns, the present results suggest that affording the child opportunities to produce language may provide the learning benefits inherent in language production.

### Action Editor

Rebecca Treiman served as action editor for this article.

### Author Contributions

E. W. M. Hopman and M. C. MacDonald designed the study. E. W. M. Hopman conducted the study and analyzed the data under the supervision of M. C. MacDonald. E. W. M. Hopman and M. C. MacDonald interpreted the findings and wrote and revised the manuscript.

### ORCID iD

E. W. M. Hopman  <https://orcid.org/0000-0002-6162-3311>

### Acknowledgments

We thank Teresa Turco for creating stimuli. We thank the Language and Cognitive Neuroscience Lab, Jenny Saffran, and Tim Rogers for helpful discussion.

### Declaration of Conflicting Interests

The author(s) declared that there were no conflicts of interest with respect to the authorship or the publication of this article.

### Funding

Support for this research was provided by the Graduate School and the Office of the Vice Chancellor for Research and Graduate Education at the University of Wisconsin–Madison with funding from the Wisconsin Alumni Research Foundation.

### Supplemental Material

Additional supporting information can be found at <http://journals.sagepub.com/doi/suppl/10.1177/0956797618754486>

### Open Practices



All data and materials have been made publicly available via the Open Science Framework and can be accessed at <https://osf.io/bbf3c/> and <https://osf.io/74kqe/#>, respectively. The data and analysis plans for the experiment were not preregistered. The complete Open Practices Disclosure for this article can be found at <http://journals.sagepub.com/doi/suppl/10.1177/0956797618754486>. This article has received badges for Open Data and Open Materials. More information about the Open Practices badges can be found at <http://www.psychologicalscience.org/publications/badges>.

### References

- Amato, M. S., & MacDonald, M. C. (2010). Sentence processing in an artificial language: Learning and using combinatorial constraints. *Cognition*, *116*, 143–148.
- Baese-Berk, M. M., & Samuel, A. G. (2016). Listeners beware: Speech production may be bad for learning speech sounds. *Journal of Memory and Language*, *89*, 23–36.
- Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*, 255–279.
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1). doi:10.18637/jss.v067.i01.
- Bixby, K. N. (2017). *Production effects on perception: How learning to produce sound changes auditory perception*. Unpublished doctoral dissertation, Department of Brain and Cognitive Sciences, University of Rochester, New York.
- Bock, K., Dell, G. S., Chang, F., & Onishi, K. H. (2007). Persistent structural priming from language comprehension to language production. *Cognition*, *104*, 437–458.
- Boiteau, T. W., Malone, P. S., Peters, S. A., & Almor, A. (2014). Interference between conversation and a concurrent visuomotor task. *Journal of Experimental Psychology: General*, *143*, 295–311.

- Branigan, H. P., Pickering, M. J., & McLean, J. F. (2005). Priming prepositional-phrase attachment during comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*, 468–481.
- Brown-Schmidt, S., & Konopka, A. E. (2015). Processes of incremental message planning during conversation. *Psychonomic Bulletin & Review*, *22*, 833–843.
- Carter, M. J., & Ste-Marie, D. M. (2017). Not all choices are created equal: Task-relevant choices enhance motor learning compared to task-irrelevant choices. *Psychonomic Bulletin & Review*, *24*, 1879–1888.
- Craik, F. I. M., & Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General*, *104*, 268–294.
- Fan, J. E., & Turk-Browne, N. B. (2013). Internal attention to features in visual short-term memory guides object learning. *Cognition*, *129*, 292–308.
- Fan, J. E., Turk-Browne, N. B., & Taylor, J. A. (2016). Error-driven learning in statistical summary perception. *Journal of Experimental Psychology: Human Perception and Performance*, *42*, 266–280.
- Ferreira, F., & Patson, N. D. (2007). The ‘good enough’ approach to language comprehension. *Language & Linguistics Compass*, *1*, 71–83.
- Green, C. S., Kattner, F., Siegel, M. H., Kersten, D., & Schrater, P. R. (2015). Differences in perceptual learning transfer as a function of training task. *Journal of Vision*, *15*(10), Article 5. doi:10.1167/15.10.5
- Grodzinsky, Y. (2000). The neurology of syntax: Language use without Broca’s area. *Behavioral & Brain Sciences*, *23*, 1–71.
- Hart, B., & Risley, T. R. (1995). *Meaningful differences in the everyday experience of young American children*. Baltimore, MD: Paul H. Brookes.
- Kang, S. H. K., Gollan, T. H., & Pashler, H. (2013). Don’t just repeat after me: Retrieval practice is better than imitation for foreign vocabulary learning. *Psychonomic Bulletin & Review*, *20*, 1259–1265.
- Karpicke, J. D., & Blunt, J. R. (2011). Retrieval practice produces more learning than elaborative studying with concept mapping. *Science*, *319*, 966–968.
- Karpicke, J. D., & Roediger, H. L., III. (2008). The critical importance of retrieval for learning. *Science*, *331*, 772–775.
- Krashen, S. D. (2003). *Explorations in language acquisition and use*. Portsmouth, NH: Heinemann.
- Leach, L., & Samuel, A. G. (2007). Lexical configuration and lexical engagement: When adults learn new words. *Cognitive Psychology*, *55*, 306–355.
- MacDonald, M. C. (2016). Speak, act, remember: The language-production basis of serial order and maintenance in verbal memory. *Current Directions in Psychological Science*, *25*, 47–53.
- MacLeod, C. M., & Bodner, G. E. (2017). The production effect in memory. *Current Directions in Psychological Science*, *26*, 390–395.
- Middleton, E. L., Schwartz, M. F., Rawson, K. A., & Garvey, K. (2015). Test-enhanced learning versus errorless learning in aphasia rehabilitation: Testing competing psychological principles. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *4*, 1253–1261.
- Montag, J. L., & MacDonald, M. C. (2015). Text exposure predicts spoken production of complex sentences in 8- and 12-year-old children and adults. *Journal of Experimental Psychology: General*, *144*, 447–468.
- Ouellette, G., Martin-Chang, S., & Rossi, M. (2017). Learning from our mistakes: Improvements in spelling lead to gains in reading speed. *Scientific Studies of Reading*, *21*, 350–357.
- Peirce, J. W. (2007). PsychoPy—Psychophysics software in Python. *Journal of Neuroscience Methods*, *162*, 8–13.
- Pickering, M. J., & Garrod, S. (2013). An integrated theory of language production and comprehension. *Behavioral & Brain Sciences*, *36*, 329–347.
- R Core Team. (2016). *R: A language and environment for statistical computing*. Retrieved from <http://www.R-project.org/>
- Roediger, H. L. III, & Karpicke, J. D. (2006). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science*, *17*, 249–255.
- Savill, N., Ellis, R., Brooke, E., Koa, T., Ferguson, S., Rojas-Rodriguez, E., . . . Jefferies, E. (2017). Keeping it together: Semantic coherence stabilizes phonological sequences in short-term memory. *Memory & Cognition*. Advance online publication. doi:10.3758/s13421-017-0775-3
- Segaert, K., Menenti, L., Weber, K., Petersson, K. M., & Hagoort, P. (2011). Shared syntax in language production and language comprehension—An fMRI study. *Cerebral Cortex*, *22*, 1662–1670.
- VanPatten, B. (2013). Input processing. In S. M. Gass & A. Mackey (Eds.), *The Routledge handbook of second language acquisition* (pp. 268–281). New York, NY: Routledge.
- VanPatten, B., & Cadierno, T. (1993). Explicit instruction and input processing. *Studies in Second Language Acquisition*, *15*, 225–243.
- Wenger, S. K., Thompson, C. P., & Bartling, C. A. (1980). Recall facilitates subsequent recognition. *Journal of Experimental Psychology: Human Learning and Memory*, *6*, 135–144.
- Zimmerman, F. J., Gilkerson, J., Richards, J. A., Christakis, D. A., Xu, D., Gray, S., & Yapanel, U. (2009). Teaching by listening: The importance of adult-child conversations to language development. *Pediatrics*, *124*, 342–349.