Applied Psycholinguistics **40** (2019), 137–166
doi:10.1017/S0142716418000504

Context, word, and student predictors in second language vocabulary learning

EVELIEN MULDER, MARCO VAN DE VEN, ELIANE SEGERS, and
LUDO VERHOEVEN
Radboud University

Received: July 21, 2017 Revised: April 25, 2018 Accepted: June 28, 2018

ADDRESS FOR CORRESPONDENCE
Evelien Mulder, Behavioural Science Institute, Radboud University, Montessorilaan 3, P.O. Box 9104, 6500 HE Nijmegen, The Netherlands. E-mail: e.mulder@pwo.ru.nl

ABSTRACT
We examined to what extent the variation in vocabulary learning outcomes (vocabulary knowledge, learning gain, and rate of forgetting) in English as a second language (L2) in context can be predicted from semantic contextual support, word characteristics (cognate status, Levenshtein distance, word frequency, and word length), and student characteristics (prior vocabulary knowledge, reading ability, and exposure to English) in 197 Dutch adolescents. Students were taught cognates, false friends, and control words through judging sentences with varying degrees of semantic contextual support using a pretest/posttest between subjects design. Participants were presented with an English target word and its Dutch translation, followed by an English sentence. They were instructed to judge the plausibility of the sentence. Mixed-efffects models indicated that learning gains were higher for sentences with more semantic contextual support and in students with stronger reading comprehension skills. We were the first to show that Levenshtein distance is an important predictor for L2 vocabulary learning outcomes. Furthermore, more accurate as well as faster learning task performance lead to higher learning outcomes. It can thus be concluded that L2 study materials containing semantically supportive contexts and that focus on words with little L1-L2 overlap are most effective for L2 vocabulary learning.

Keywords: cognate status; context; individual differences; second language; vocabulary learning

Building a rich vocabulary in a second language (L2) is essential to gain a sufficient level of L2 proficiency and, therefore, entails a large part of L2 education. According to the lexical quality hypothesis, language ability is facilitated by detailed semantic, phonological, and orthographic representations of words in the mental lexicon (Perfetti & Hart, 2002). It has been shown that the strength of L2 lexical representations can be fostered by embedding words in context (Nassaji, 2003). Thus far, research on vocabulary learning in context has mostly operationalized the semantic informativeness of the context categorically (i.e., semantically supportive vs. nonsupportive contexts) and not as a continuous measure, reflecting also the effects of more subtle distinctions in the semantic support of a context. Besides semantic contextual support, the strength of the learning of L2 lexical representations has been found to be related to word characteristics, such as the degree of L1-L2 overlap between the L2 items to be learned (e.g., Crossley, Salsbury, & McNamara, 2012; Dijkstra, Grainger, & van Heuven, 1999; Pytlyk, 2017), the frequency and length of words (Bolger, Balass, Landen, & Perfetti, 2008; Elgort, Perfetti, Rickles, & Stafura, 2015), and learner characteristics, such as prior knowledge and reading ability (e.g., Alderson, Nieminen, & Huhta, 2016; Huensch & Ventura, 2017; Zhang, Chin, & Li, 2017). Nevertheless, L2 vocabulary learning has not often been examined as a function of both word and learner characteristics. In the present study, we examined L2 English lexical learning in Dutch adolescents in relation to context, word, and student predictors. English was chosen as the target L2 because it is one of the most commonly taught nonnative languages. The novel contribution of this study is that we used a continuous rather than a categorical measure to reflect the semantic relatedness between a prime and the to-be-learned target word, embedded in the same sentence. Using latent semantic analysis (LSA) scores (Landauer, Foltz, & Laham, 1998) we measured the degree of semantic contextual support of a sentence context. In addition, we operationalized L1-L2 overlap through Levenshtein distance. In this study we related the predictors to three outcomes of L2 vocabulary learning through sentence reading: *vocabulary knowledge* immediately after learning, *learning gain* (difference between pretest and immediate posttest of vocabulary knowledge), and *rate of forgetting* (difference between immediate and delayed posttest of vocabulary knowledge).

CONTEXTUAL SUPPORT IN VOCABULARY LEARNING

One of the key premises of L2 learning is that vocabulary should be acquired in context (Ellis, 2013). Vocabulary learning through a semantic context requires deep processing of words as lexical units, and such learning has been found to result in better storage and retrieval (Nassaji, 2003). Acquiring vocabulary in an instructed learning setting by reading semantic supportive sentence contexts provides abundant clues of the semantic, orthographic, and syntactic information of words (Beck, McKeown, & McCaslin, 1983), thus resulting in robust representations of words in the mental lexicon (e.g., Elgort et al., 2015; Ma, Chen, Lu, & Dunlap, 2015). Providing words in a sentence context can also disambiguate unfamiliar phonological contrasts and thus lead to higher lexical specificity as was, for example, demonstrated in L2 learners of Russian (Chrabaszcz & Gor, 2017).

In addition to phonology, several studies have demonstrated the importance of semantic contextual support on semantic disambiguation, with different context qualities leading to varying results. Rich, often highly constraining contexts, on the one hand, have a high information load, leading to a limited number of possible interpretations of a word and more specific and strong mental lexical representations. Less rich, often low constraining contexts, on the other hand, include less information, leaving more opportunity to infer the meaning of a word, which may lead to less robust word storage (Ma et al., 2015) as the initial interpretation may be erroneous. In proficient and less proficient L2 learners, better performance for vocabulary learning was found when target words were embedded in sentences with highly semantically related words, reflecting a higher degree of semantic contextual support (Elgort et al., 2015). This effect was largest for highly proficient learners, suggesting that such context effects may (partially) depend on learner characteristics. Furthermore, Beck et al. (1983) showed that directive contexts, which were intended to reveal the meaning of a word, provided adult readers with most clues about a word. Daneman and Green (1986) provided skilled adult readers with contexts that included low-frequency words and showed that vocabulary growth was primarily predicted by semantic cues, spread across the seven categories that were provided. Furthermore, Ma et al. (2015) showed that Chinese adults learning English as an L2 had most benefit during vocabulary learning when words were placed in highly constraining sentences.

These studies use word recognition through reading and lexical decision paradigms rather than L2 translation accuracy as an indicator of vocabulary learning and include advanced L2 learners. Furthermore, in experimental settings, L2 vocabulary learning is often assessed only immediately after learning. Long-term effects of L2 vocabulary learning, as operationalized by rate of forgetting in the form of a delayed posttest, are not often addressed. Adlof, Frishkoff, Dandy, and Perfetti (2016) have shown that both adult and novice first language (L1) learners can acquire and retain new L1 words over time when presented in highly constraining (i.e., semantically supportive) contexts, suggesting lower rate of forgetting when words are placed in such contexts. More important, studies that did focus on the effects of semantic contextual support on (L2) vocabulary learning used categorical measures (i.e., semantically supportive vs. non-supportive contexts) or other categorizations, such as different types of semantic cues (Daneman & Green, 1986; Ma et al., 2015). Therefore, these studies do not provide insight into the influence of subtle variations in the semantic contextual support of the context on vocabulary learning. A statistical technique that allows us to gain more insight into the influence of the degree of semantic contextual support on word learning is LSA (Landauer et al., 1998). LSA rests on the assumption that words that often occur in similar contexts are semantically related (the distributional hypothesis), and the LSA score reflects the degree to which this is the case. This computational technique measures the semantic relations between words beyond their direct co-occurrences in the same texts, based on a large corpus of written texts. Previous studies have shown that LSA scores can be used to predict human behavior for example in semantic priming in visual (Landauer & Dumais, 1997) or auditory (van de Ven, Tucker, & Ernestus, 2011) lexical decisions, and may therefore also be used to predict vocabulary learning. In the present study, we used LSA as a continuous measure to operationalize semantic contextual support, assuming that higher LSA scores indicated more semantic contextual support.

WORD PREDICTORS OF CONTEXTUAL VOCABULARY LEARNING

A core predictor of the acquisition of lexical representations in contextual L2 vocabulary learning is L1-L2 overlap (Dijkstra et al., 1999). In the present study, we measured L1-L2 overlap in two ways, by means of cognate status and Levenshtein distance. Words that share orthographic, phonological, and meaning similarities are referred to as cognates. Words with phonological and/or orthographic but no semantic overlap are called false cognates or false friends (Carrol, 1992). For instance, the English word FILM and the Dutch word FILM show complete phonological, orthographic, and semantic overlap, and these words are therefore cognates. In contrast, although the English word SPOT (Dutch: VLEK) shares orthographic similarity with the Dutch SPOT (English: MOCKERY), these words have no semantic overlap and are thus false friends. Cognate status has been shown to contribute significantly to translation variance (de Groot, 1992) and to performance on both forward and backward translation (de Groot, Dannenburg, & van Hell, 1994). Furthermore, evidence was found that there is a benefit of cognate status both in learning and in retrieving vocabulary in university students learning a foreign language (e.g., de Groot & Keijzer, 2000; Lotto & de Groot, 1998). However, another study found evidence that lexical items were connected between languages regardless of cognate status in Dutch–English bilinguals (de Groot & Nas, 1991). Different types of priming experiments were conducted to examine word representations in the bilingual lexicon. The results suggest that cognate status does not necessarily benefit or hamper word representations (de Groot & Nas, 1991). The use of cognates was shown to foster morphological awareness in Spanish–English bilinguals in fourth up to eighth grade (Hancin-Bhatt & Nagy, 1994). In their study, Starreveld, de Groot, Rossmark, and van Hall (2014) examined the cognate effect as a marker of activation of a nontarget language during picture naming with varying sentence contexts. They found that the cognate effect was smaller for high-constraint than for low-constraint sentences.

More subtle differences in cross-linguistic overlap can be measured by calculating the Levenshtein distance (Levenshtein, 1966) between two words. This measure reflects the number of insertions, deletions, and substitutions required to edit one word into another. For example, when comparing the English word CLOCK to its Dutch translation KLOK, the Levenshtein distance is 2: the first C is substituted for a K and the second C is deleted. A small Levenshtein distance indicates a large overlap between words, whereas a large distance points to a small overlap. Levenshtein distance has so far only been used as a measure to describe the structure of the mental lexicon and how L1 and L2 words are organized (Dautriche, Mahowald, Gibson, Christophe, & Piantadosi, 2017), and not as a measure to show subtle effects of L1-L2 overlap on L2 vocabulary learning, or in interaction with context characteristics.

Two other, more traditional, predictors that have been used in vocabulary learning research are word frequency (Diependaele, Lemhöfer, & Brysbaert, 2013) and word length (Whaley, 1978). In lexical decision, which can be seen as a measure of a word’s familiarity (i.e., with acoustic as well as semantic features) or word learning (e.g., Reichle & Perfetti, 2003), processing accuracy and speed are higher for words with relatively high frequencies (Hauk & Pulvermüller, 2004).

Frequent words in monolingual settings are, to some extent, used relatively frequently in bilingual translation settings as well (de Groot, 1992). Therefore, it is assumed that L2 learners are more likely to learn words that are relatively frequent in the L1 (Lotto & de Groot, 1998). Furthermore, short response times in lexical decision indicate that representations of highly frequent words are more easily accessible from the mental lexicon than low-frequency words (e.g., Adelman, Brown, & Quesada, 2006; Forster, 1976; Seidenberg & McClelland, 1989). However, after controlling for cognate status, de Groot and Keijzer (2000) found marginal remaining effects of word frequency in experienced foreign language learners. Finally, semantic priming effects were shown to be moderated by word frequency; priming effects are stronger for low- than for high-frequency words (Rayner, Ashby, Pollatsek, & Reichle, 2004). Moreover, word length has been shown to be a predictor of lexical decision. Shorter words tend to be processed more quickly and more accurately than longer words, which suggests that these words have stronger lexical representations (Whaley, 1978) and may, hence, be easier to learn. On the one hand, Hauk and Pulvermüller (2004) found that, when looking at the amplitude of neurophysiological responses, relatively long words evoked stronger responses in the early stages (~100 ms after stimulus onset) than did relatively short words during lexical decision. On the other hand, they found stronger responses to short words at later stages (150–360 ms after stimulus onset). These findings suggest that long and short words may be processed in different ways, which may be related to the way these lexical items are stored in the mental lexicon.

STUDENT PREDICTORS OF CONTEXTUAL VOCABULARY LEARNING

Vocabulary learning outcomes vary in students with different characteristics. Prior vocabulary knowledge has been demonstrated to influence word recognition in 8.5- to 13-year-old L1 learners (Nation & Snowling, 2004). Children with more vocabulary knowledge were better at recognizing structure in novel words (Nation & Snowling, 1998), but this was not studied specifically in a vocabulary learning setting. Cain, Oakhill, and Lemmon (2004) did find that L1 learners with less vocabulary knowledge had more difficulty acquiring new vocabulary than those with more prior vocabulary knowledge.

Reading comprehension skills have been widely shown to foster vocabulary learning within a semantically supportive context. For example, Ouellete (2006) demonstrated relationships between reading comprehension and both vocabulary breadth and depth in Grade 4 children. Furthermore, 9- to 10-year-old L1 learners with weak reading comprehension skills were found to have more difficulty with vocabulary learning than learners with good reading comprehension skills (Cain et al., 2004). In addition, university students who were proficient text comprehenders made larger vocabulary learning gains than poor comprehenders in a study by Elgort and Warren (2014).

Exposure to English media outside school has been found to contribute to L2 learning. Kuppens (2010) performed a study to examine the influence of selfreported media use on incidental language acquisition in Flanders' students, in their final year of primary school. This showed a significant influence of the use of subtitled English television on translation accuracy. Students' media exposure is not limited to television but also comprises listening to English music, reading English texts online, and watching videos (Lindgren & Muñoz, 2013). Exposure to these English media was shown to be a strong predictor for L2 reading and listening comprehension skills in Dutch 10- to 11-year-olds. Highly proficient adult L2 learners were shown to benefit more from semantic contextual support; however, interactions between learner characteristics, such as prior knowledge or exposure, of novice L2 learners and semantic contextual support remain to be examined.

PRESENT STUDY

From the research so far, it can be concluded that context, word, and learner factors may predict the learning of L2 vocabulary. Although relations between these factors have been examined (e.g., de Groot & Keijzer, 2000; Elgort et al., 2015; Starreveld et al., 2014), an attempt to integrate these measures into a single design has not yet been made. Further, both semantic contextual support and cognate status tend to be operationalized categorically rather than continuously. Therefore, the present study aimed to examine context, word, and student predictors of L2 English vocabulary learning in 197 Dutch secondary school students. All students had received English education during primary school and now attended different educational tracks within secondary school: lower and intermediate prevocational education (VMBO-t/Havo), intermediate education (Havo), or higher level and preuniversity education (Havo/VWO). Students were asked to perform a computerized task, consisting of a pretest, learning trials, an immediate posttest, and a delayed posttest. On both pretest and posttests, participants translated cognates, false friends, and control words with different degrees of Dutch–English overlap and varying word frequencies. During learning trials, the students read sentences with these target words. They were instructed to judge the plausibility of these sentences. For each target word, we selected one prime that was strongly related to the target word (i.e., creating a context with relatively high semantic support) and one that was weakly or unrelated to the target word (i.e., creating a context with less semantic support).

We examined context, word, and student predictors of three different L2 vocabulary learning outcomes: *vocabulary knowledge* at immediate posttest after the vocabulary learning trials; *learning gain*, operationalized as the difference between pretest and immediate posttest vocabulary knowledge; and *rate of forgetting*, as reflected by the decrease in vocabulary knowledge between immediate posttest and delayed posttest looking at the prediction of time (immediate or delayed). We were the first to examine the continuous effects of semantic contextual support, as measured by the LSA score of semantic similarity between the prime and the target word that needed to be learned. Another novel contribution was that we examined the influence of L1-L2 form overlap, as reflected by Levenshtein distance on L2 vocabulary learning. The vocabulary knowledge, learning gain, and rate of forgetting were related to word characteristics (L1-L2 overlap, word frequency, and word length) and student characteristics (prior knowledge, reading ability, and exposure to English). The research questions were to what extent the three different L2 vocabulary learning outcomes are explained by

1. semantic contextual support;
2. word characteristics (i.e., cognate status, Levenshtein distance between L1 and L2, word frequency, and word length);
3. student characteristics (i.e., prior vocabulary knowledge via pretest accuracy, reading comprehension, exposure to English media outside school, and learning trials performance via accuracy and reaction times during sentence judgment); and
4. interactions between the aforementioned.

We expected that semantically more supportive contexts (i.e., higher LSA scores for prime and target) would result in a larger knowledge and learning gain (e.g., Elgort et al., 2015; Ma et al., 2015), and in a lower rate of forgetting compared to contexts with less semantically supportive contexts (i.e., lower LSA scores; Adlof et al., 2016), shown by an interaction between time and semantic support. We expected several word characteristics to predict vocabulary learning. We hypothesized that cognates (e.g., de Groot, 1992; de Groot et al., 1994), words with smaller Levenshtein distances (e.g., Dautriche et al., 2017), and highly frequent (e.g., Diependaele et al., 2013) and shorter words (e.g., Whaley, 1978) would be known best after learning in context, followed by control words, and finally false friends. We also expected larger gains could be made for false cognates, as compared to cognates (e.g., Starreveld et al., 2014) and control words. With reference to student characteristics, we hypothesized positive effects of prior vocabulary knowledge (e.g., Nation & Snowling, 1998), reading comprehension (e.g., Cain et al., 2004), task performance, and exposure to English media (Kuppens, 2010) on the learning outcomes. We examined interactions between context, word, and student characteristics. We expected highly proficient learners (i.e., students with larger prior vocabulary knowledge, and better reading comprehension and/or task performance; e.g., Elgort et al., 2015) to benefit more from semantic contextual support (context characteristics). We hypothesized semantic contextual support (context characteristic) to be more important for low-frequency words than for high-frequency words (word characteristics; Rayner et al., 2004). Further interactions were explored.

METHOD

Participants

Participants included in this study were 197 Dutch students learning English as an L2 from the seventh grade of three secondary schools in the Netherlands. The sample consisted of 56 students who were in their first year of lower and intermediate prevocational education (VMBO-t/Havo), 30 students in intermediate education (Havo), and 111 students of higher level and preuniversity education (Havo/VWO). The sample comprised 104 boys and 93 girls (mean age = 12 years and 8 months, *SD* = 5.04 months).

At the time of testing, the participants had received 5 months of English instruction at secondary school. Most students had also received English instruction at primary school from the fourth grade onward (*n* = 160); the others had received English education before the fourth grade. The majority of the participants spoke Dutch at home with their parents (*n* = 175) and siblings (*n* = 167); some participants spoke Dutch and English (parents *n* = 6; siblings *n* = 5), or Dutch and another language, such as Turkish, German, or Papiamento (parents *n* = 16; siblings *n* = 20).

Materials

Vocabulary learning in context. A computer-based experiment was constructed that consisted of a pretest, learning phase, immediate posttest, and delayed posttest. In all four parts of the task, participants were presented with 96 target words with varying degrees of English–Dutch overlap: cognates (e.g., *apple–appel*), false friends (e.g., *note* [Dutch translation: briefje]–*noot* [English translation: nut]), and control words (e.g., *fibre–vezel*), along with 20 filler items.

During the pretest, immediate posttest, and delayed posttest, participants translated all targets and fillers from English into Dutch. The English target words appeared on the screen and the students had to type the translations within 8000 ms. In the learning phase, students had to learn the words through reading sentences with different degrees of semantic contextual support and judging their plausibility, after being presented with a Dutch translation of the target word. For each item, we created one semantically supporting context (with a relatively high LSA score) and one semantically less supporting context (with a low LSA score). The LSA scores were retrieved using the pairwise comparison tool from the LSA website at the University of Colorado at Boulder (2003). We used the General Reading Up to First Year of College corpus with 300 factors. The experimental procedure is illustrated in Figure 1. In this figure, prime words were underlined and target words were printed in bold. More information about the selection of primes and sentence construction can be found in the Sentence Contexts section.

All participants were randomly assigned to a list with either of the two versions of the items, all combined with 20 filler items; the condition of each item had been randomly generated. Thus, the presentation of the sentences in either a semantically supportive or less supportive context was counterbalanced.

To intrinsically motivate participants (Martens, Gulikers, & Bastiaens, 2004), they were told they would be playing a game in which they had to judge sentences on their plausibility. Thus, filler items that were implausible sentences were included in the task, such as: “The flowers were having a fight.” All sentences with target words were plausible. Learning trials were presented as follows for both target and filler items: participants were presented with a fixation cross for 200 ms, followed by an English word and its corresponding Dutch

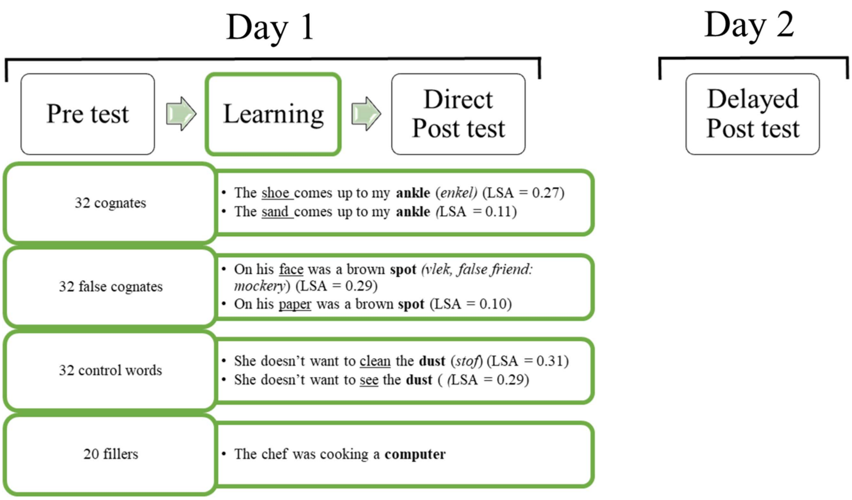

Figure 1. Graphic overview of experiment. Primes are underlined, targets are printed in bold, and Dutch translations of the targets are printed in italics.

translation (e.g., *box–doos*), which was shown for 2500 ms. Subsequently, participants were presented with a sentence. Participants were then to decide whether the sentence was plausible or implausible, pressing "A" on the keyboard for a plausible and "L" for an implausible sentence. The sentence was shown for 8000 ms, unless participants responded earlier, in which case the next sentence would appear. The English word and its Dutch translation were presented prior to the context, which allowed learners to construct a lexical entry (if none was available).

Stimuli were presented in a pseudorandom order with the restriction that semantically or phonologically related target words were never presented successively. To facilitate storage of new representations (Perfetti & Hart, 2002), all learning trials were audio supported: students heard the English target word, its Dutch translation, and the sentence in which the English target word was embedded. The audio recordings were made using Audacity® version 2.1.2 (Audacity Team, 2015). Stimuli were then extracted by means of Praat (Boersma, 2001). The task was programmed using Delphi XE 5 update 2. A detailed overview of all stimulus characteristics can be found in Appendix A.

TARGET WORDS. Target words were selected comparing words from four studies (Brenders, van Hell, & Dijkstra, 2011; Dijkstra et al., 1999; Lotto & de Groot, 1998; van Hell & Dijkstra, 2002) with vocabulary lists from the English as a Foreign Language (EFL) method "Stepping Stones." Sets of words consisting of a cognate, a false friend, and a noncognate were constructed and controlled for word type (e.g., adjectives: *wild–glad–near*; nouns: *apple–note–fibre*), length, number of syllables, singular or plural form, and word frequency, based on the

English CELEX database (Baayen, Piepenbrock, & van Rijn, 1993). The frequency of the target words ranged from 11 to 22,071 ($M = 1{,}224.88$, $SD = 2{,}557.62$) and was kept constant across the different word types. Thus, a list of 96 three- to seven-letter nouns, verbs, and adjectives was composed. English–Dutch cognate pairs were constructed based on cross-linguistic similarity in terms of orthography, phonology, and/or semantics. False friend pairs were matched on orthography and/or phonology, but not on semantics. Noncognates were selected if there was no matching orthography and/or phonology and possibly no matching semantics (Dijkstra et al., 1999). For the cognates, Levenshtein distance (Levenshtein, 1966) ranged from 0 to 5 ($M = 1.44$, $SD = 1.32$); for false friends, it ranged from 1 to 7 ($M = 4.03$, $SD = 1.55$), and the range was 2 to 8 for control words ($M = 4.53$, $SD = 1.45$). Cognate status and Levenshtein distance were highly correlated, and therefore these variables were orthogonalized. For the orthogonalization, a linear regression model was fitted with Levenshtein distance as the dependent and cognate status as the independent variable (e.g., see Wurm & Fisicaro, 2014). The residuals of this model (Levenshtein distance$_{resid}$) were used to replace Levenshtein distance as a predictor in the mixed-effects models.

SENTENCE CONTEXTS. Sentences were constructed taking several aspects into account. Sentences consisted of a semantic prime preceding the target word to be learned. Prime words were selected based on semantic relatedness with the target word using LSA (University of Colorado at Boulder, 2003), thus indicating the degree of semantic contextual support. LSA computes a score ranging from –1 to +1, where a higher score indicates that words are more likely to occur in similar texts (i.e., measured beyond first-order co-occurrences). On the basis of the distribution of words across different texts, words are placed in a vector space. The LSA score was computed by taking the cosine of the angle between the vectors for the primes and the targets. Similar to the target words, we verified that the primes were familiar to the participants on the basis of the Dutch EFL method "Stepping Stones." For each sentence, a prime was selected that was closer to 0 and a prime that was closer to 1 compared to the target. LSA scores differed significantly for the highly related prime ($M = 0.46$) and the less related prime ($M = 0.099$), $t\ (112.21) = 21.68$, $p < .0001$, $d = 3.21$. The frequencies of the primes ranged from 33 to 111,471 ($M = 3{,}157.84$, $SD = 8{,}860.41$), based on the English CELEX database (Baayen et al., 1993). We aimed to keep prime frequencies comparable to target frequencies. The primes were always placed in or near sentence-initial position, while the targets were always placed in sentence-final position (similar to Elgort et al., 2015). The distance between primes and targets ranged from 1 to 5 words ($M = 2.57$, $SD = 0.93$) and was similar across word categories and conditions, as was sentence length, which ranged from 4 to 10 words ($M = 6.55$, $SD = 1.07$).

Reading comprehension. A measure often used for reading comprehension is a cloze task or a gap text (e.g., Gellert & Elbro, 2013; Keenan & Meenan, 2014). This measure required completing a text in which words had been omitted. An exam text for lower and intermediate prevocational education was selected (van Gelderen et al., 2004), and every seventh word was omitted and replaced by a blank line. The omitted words were listed below the text, and children were instructed to write down each word on the correct line. Reliability of this task was $\alpha = 0.943$.

Questionnaire. Students were asked to complete a questionnaire in order to measure their linguistic background and exposure to English. Participants were asked to indicate their exposure to English media outside school, answering the questionnaire with a 7-point Likert scale. They were asked how often (1 = *never* to 7 = *daily*) and how long (1 = *never* to 7 = *5 hours or more*) they played English video games, read English books or texts, watched English television programs or films, watched online videos in English, and listened to English music.

Procedure

For this study, a convenience sample was used, consisting of three schools that were contacted by the first author and agreed to participate. The students' parents or guardians received an information letter and provided passive consent, with active consent being received from the students.

Participants completed two 50-min sessions. In the first session, the pretest, learning phase, and immediate posttest were carried out in a classroom setting. In the second session, the delayed posttest was performed, followed by the questionnaire to assess student predictors and the cloze test. Stimuli were presented on a white screen printed in black lowercase letters. The questionnaire and cloze test were paper-and-pencil tasks.

During the first session, students were told that they were testing an English video game called "It's raining rabbits all day." They received an instruction about the game. First, they were told that they had to make plausibility judgments during the game. During a familiarization phase, participants were asked to provide plausibility judgments for several clear examples of plausible and implausible sentences. Second, we explained to the participants that, to help them with the plausibility judgments, they would first see the translation of one English word (the target) from the sentence and its Dutch translation. Thus, for each trial, participants would first see an English target word with its Dutch translation, followed by a full English sentence in which this word was embedded, and they were instructed to provide a plausibility judgment for this sentence. Third, participants were instructed to work individually, and as quietly, quickly, and accurately as possible. After the instruction, they were allowed to ask questions. The delayed posttest was administered a day after Session 1.

Analyses

The data were analysed in R (version 3.3.1) by means of generalized linear mixed-effects models in lme4 (Bates, Maechler, Bolker, & Walker, 2015) using contrast coding for factors (Jaeger, 2008), and with the logit link function (e.g., Breslow & Clayton, 1993; Jaeger, 2008). To control for multicollinearity and possible normality distribution violations, all continuous variables were standardized and centered (Belsley, Kuh, & Welsh, 1980). One control item needed to be excluded from the analyses, because of a mismatch between the audio recording and the displayed sentence, resulting in 32 cognates, 32 false cognates, and 31 control word items. We created three models that all had binomial dependent variables per participant for each item: one for translation accuracy, that is, *vocabulary knowledge* at immediate posttest (correct/incorrect); a second model for *learning gain* (learning gain/no learning gain), as operationalized by the difference between pretest and posttest vocabulary knowledge; and a third model for *rate of forgetting* in which we measured translation accuracy at the immediate and delayed posttest (correct/incorrect) and forgetting could be measure by testing for main effects of and interactions with the variable time. We determined the final mixed-effects models by means of model selection, in which predictors were removed if they did not attain significance at the 5% level. Model selection took place in three separate steps. We determined the significant fixed effects, followed by the random effects (student and word), and the random slopes (i.e., interactions between the fixed and random effects). Variables and interactions were added successively to lead to a converging model with increased model fit. Chi-square tests were used to examine whether inclusion of a variable led to a significantly better model fit. We also ensured that these models then contained lower Akaike information criterion values. To construct the fixed-effects section of the mixed model, variables were added successively, based on preliminary considerations. Once the fixed-effects section was complete, the inclusion of random slopes for the fixed effects was tested using chi-square tests (Baayen, 2008). We report one-tailed significance values for directed hypotheses and two-tailed values for explorative analyses. Effect size is indicated by beta coefficients and their corresponding confidence intervals: large betas indicate a large effect size, and narrow confidence intervals point to more precision as compared to broad confidence intervals.

RESULTS

Descriptives

The descriptive statistics are presented in Table 1, including means and standard deviations of student characteristics and both pretest and posttest accuracy, tabulated by cognate status across the different educational tracks.

We first assessed whether actual vocabulary learning took place. This was the case, as a significant difference between *pretest and posttest accuracy* was found, χ^2 (1, $N = 142$) = 42.55, $p < .0001$. The effect size (Φ coefficient) for this effect was 0.55, which can be considered a large effect (Ellis & Steyn, 2003). This indicated that, in general, students' posttest accuracy was higher than their pretest accuracy, and a learning effect had occurred. Although there was a significant overall difference between pretest and immediate posttest accuracy, a ceiling

Table 1. *Means and standard deviations (in parentheses) of reading comprehension (RC), sentence judging accuracy (SJA), reaction time (SJRT), and proportion of words correct on pretest (T1) and posttest (T2) tabulated by cognate status, and across different tracks*

	Track 1*	Track 2*	Track 3*	Total
RC	8.85 (7.23)	19.97 (7.03)	11.85 (7.92)	11.32 (7.96)
SJA	71.5 (12.3)	75.29 (9)	73.65 (10.21)	73.28 (10.64)
SJRT	2744.69 (1445.64)	3097.18 (1315.51)	2822.29 (1366.5)	2842.39 (1381.33)
Cognate (T1)	0.82 (0.39)	0.82 (0.39)	0.85 (0.36)	0.83 (0.37)
Cognate (T2)	0.86 (0.34)	0.89 (0.32)	0.90 (0.30)	0.89 (0.31)
False friend (T1)	0.36 (0.48)	0.36 (0.48)	0.35 (0.48)	0.35 (0.48)
False friend (T2)	0.52 (0.50)	0.55 (0.50)	0.52 (0.50)	0.52 (0.50)
Control (T1)	0.28 (0.45)	0.27 (0.44)	0.28 (0.45)	0.28 (0.45)
Control (T2)	0.38 (0.48)	0.46 (0.50)	0.41 (0.49)	0.41 (0.49)
N =	56	30	111	197

Note: *Track 1 = prevocational, track 2 = intermediate, track 3 = preuniversity.

effect was present for cognates. Furthermore, we examined the possible presence of a speed–accuracy trade-off during the sentence reading trials. There was a significant, yet small, correlation between sentence judgment accuracy and sentence judgment reaction time, $r = .10$, $p < .0001$. This indicates students were less accurate in their sentence judgment when they had faster reaction times.

After this, we fitted three different models, using mixed-effects regression, to examine the effect of context, word, and student predictors on the three learning outcomes: *vocabulary knowledge* at immediate posttest, *learning gain* (difference between pretest and immediate posttest vocabulary knowledge), and *rate of forgetting* (vocabulary knowledge at immediate posttest and delayed posttest). For all models, the model intercept indicates the model prediction holds when all variables have the intercept value; that is, the intercept levels for factors, and the standardized means for numeric variables.

Predictors of vocabulary knowledge

We created a model for *vocabulary knowledge* at immediate posttest, to examine the vocabulary knowledge after the sentence judgment during the learning trials. To analyze the results, in the *vocabulary knowledge* model we used *immediate posttest accuracy* as the dependent variable. The following independent variables were included: *semantic contextual support* (as indicated by LSA scores; higher LSA scores pointed to a larger semantic overlap between prime and target and thus larger semantic contextual support), *prior vocabulary knowledge* (i.e., pre–test accuracy), *reading comprehension* (scores on a written cloze task), *sentence judgment accuracy* (during the learning phase), *sentence judgment reaction time* (reaction times during the learning phase), *cognate status* (cognates on

Table 2. *Summary of a generalized linear mixed-effects model predicting vocabulary knowledge at pretest*

Predictor fixed effects	β	*Z*	*p*
Intercept	3.399	6.867	<.0001
Word predictors			
Cognate status (false friends)	–4.022	–5.645	<.0001[a]
Cognate status (control words)	–5.423	–7.703	<.0001[a]
Levenshtein distance	–1.609	–4.162	<.0001[a]
Target frequency	1.262	4.311	<.0001[a]
Student predictors			
Reading comprehension	0.589	5.552	<.0001
Predictor random effects	Variance explained	χ^2	*p*
Word	7.079	7072.7	<.0001
Word: Reading comprehension	0.350	301.74	<.0001
Student	0.892	796.71	<.0001
Student: Cognate status	0.855	290.64	<.0001
Student: Levenshtein distance	0.137	75.493	<.0001
Student: Target frequency	0.508	164.08	<.0001

Note: [a]Tested one-tailed.

the intercept), *Levenshtein distance*$_{resid}$ (after controlling for cognate status), and *target frequency*. An overview of the influence of the relevant variables on vocabulary knowledge at pretest can be found in Table 2. A summary of the final model for *vocabulary knowledge at immediate posttest* is presented in Table 3.

There were main effects of *semantic contextual support*, *pretest accuracy*, *reading comprehension*, *sentence judgment accuracy and reaction time*, *cognate status*, *Levenshtein distance*, and *target frequency* on vocabulary knowledge, as reflected by *immediate posttest accuracy*. Main effects of variables that were also included in an interaction (i.e., *pretest accuracy*, *sentence judgment accuracy and reaction time*, and *target frequency*) are discussed below, together with the interactions.

Semantic contextual support and vocabulary knowledge. We examined whether *vocabulary knowledge* at immediate posttest could be explained by *semantic contextual support*. As can be seen in Table 3, there was a significant positive main effect of *semantic contextual support*, $b = 0.078$, 95% confidence interval (CI) [0.024, 0.131]. This indicates that the *vocabulary knowledge* was higher for target words placed in a sentence context that included a prime that had a strong semantic relation with the target (higher LSA scores and thus higher semantic contextual support). When the target word was placed in a sentence context with a less strongly related semantic prime (lower LSA scores and thus lower semantic contextual support), *vocabulary knowledge* was lower.

Table 3. *Summary of a generalized linear mixed-effects model predicting vocabulary knowledge at immediate posttest*

Predictor fixed effects	β	Z	p
Intercept	1.420	5.00	<.0001
Semantic contextual support	0.078	2.86	<.001[a]
Word predictors			
Cognate status (false friends)	–1.787	–5.04	<.0001[a]
Cognate status (control words)	–2.748	–7.68	<.0001[a]
Levenshtein distance	–0.806	–4.28	<.0001[a]
Target frequency	0.520	2.64	<.001[a]
Student predictors			
Pretest accuracy	3.414	43.67	<.0001[a]
Reading comprehension	0.472	4.58	<.0001[a]
Task performance: Sentence judgment accuracy	0.356	3.80	<.001
Task performance: Sentence judgment reaction time	0.013	0.41	*ns*
Interactions			
Sentence Judgment Accuracy × Sentence Judgment Reaction Time	–0.093	–3.29	<.01
Pretest Accuracy × Target Frequency	0.542	3.66	<.001
Predictor random effects	**Variance explained**	**χ^2**	***p***
Word	1.444	1807.7	<.0001
Word: Reading comprehension	0.127	76.375	<.0001
Student	1.523	660.14	<.0001
Student: Cognate status	0.992	202.05	<.0001
Student: Levenshtein distance	0.332	154.67	<.0001
Student: Target frequency	0.668	181.97	<.0001

Note: [a]Tested one-tailed.

Word characteristics and vocabulary knowledge. Furthermore, we investigated to what extent *vocabulary knowledge* was influenced by word characteristics, namely, *cognate status*, *Levenshtein distance*, and *word frequency*. There were main effects of all word predictors. Regarding *cognate status*, students knew fewer false friends, $b = –1.787$, 95% CI [–2.482, –1.093] and control words, $b = –2.748$, 95% CI [–3.449, –2.064] compared to cognates at the immediate posttests. Students knew fewer control words than false cognates, $b = –0.960$, 95% CI [–1.629, –0.291] at the immediate posttest. We controlled for multiple comparisons with a

Bonferroni correction. Furthermore, immediate posttest accuracy was larger for words with smaller Levenshtein distances, $b = -0.806$, 95% CI [–1.175, –0.437]. This indicates larger vocabulary knowledge at immediate posttest for English words that were more similar to their Dutch translations.

Student characteristics and vocabulary knowledge. In addition, to examine the effects of student characteristics *pretest accuracy*, *reading comprehension*, *sentence judgment accuracy*, and *reaction time* were included in the model. The addition of the predictors *education level*, *native language*, and *exposure to English media* did not improve model fit. There were main effects of *pretest accuracy*, *reading comprehension*, and *sentence judgment accuracy*, but the main effect of *sentence judgment reaction time* was not significant. The main effect of *reading comprehension*, as is shown in Table 3, indicated that students with stronger reading comprehension skills had larger vocabulary knowledge at immediate posttest than those with lower reading comprehension skills, $b = 0.472$, 95% CI [0.270, 0.674].

Interactions. We explored whether there were any two-way interactions between context, word, and student characteristics associated with *vocabulary knowledge*. There was a two-way interaction between *sentence judgment accuracy* (student characteristic) and *sentence judgment reaction time* (student characteristic), and there was a two-way interaction between *pretest accuracy* (student characteristic) and *target frequency* (word characteristic). The two-way interaction between sentence judgment accuracy and sentence judgment reaction time, $b = -0.093$, 99% CI [–0.15, –0.038] showed that higher *sentence judgment accuracy* scores lead to higher *posttest accuracy*. This effect was smaller when *sentence judgment reaction time*s were also slower. This means that students with higher *sentence judgment accuracy* overall had higher *posttest accuracy* than those with lower *sentence judgment accuracy*, but only if they were quick enough at performing the sentence judgment. The two-way interaction between *pretest accuracy* and *target frequency* indicated that higher word frequencies lead to higher *vocabulary knowledge*, and this relationship was even stronger for students with high *pretest accuracy*, $b = 0.542$, 99% CI [0.251, 0.833]. No other interactions were found.

Predictors of learning gain

In addition to vocabulary knowledge, we wanted to examine the *learning gain* between pretest and posttest. We created an additional model for which we recoded the accuracy scores across time (from pretest to posttest). Learning gain was coded as "0" when either the response on both pretest and posttest was incorrect or when the response was correct at pretest but incorrect at the posttest. When the response on the pretest was incorrect but correct on the posttest, the learning gain was coded as "1." Responses that were correct at both pretest and posttest were excluded from the analyses. We included the following independent variables: *semantic contextual support*, *reading comprehension*, *sentence judgment accuracy*, *sentence judgment reaction time*, *cognate status*, and *Levenshtein*

Table 4. *Summary of a generalized linear mixed-effects model predicting learning gain*

Predictor fixed effects	β	*Z*	*p*
Intercept	0.108	0.342	*ns*
Semantic contextual support	0.086	2.851	<.01[a]
Word predictors			
Cognate status (false friends)	–0.798	–2.029	<.05[a]
Cognate status (control words)	–1.736	–4.369	<.0001[a]
Levenshtein distance	–0.585	–2.871	<.01[a]
Student predictors			
Reading comprehension	0.430	3.872	<.0001[a]
Task performance - Sentence judgment accuracy	0.331	3.248	<.01
Task performance - Sentence judgment reaction time	0.033	0.931	*ns*
Interactions			
Sentence judgment accuracy: sentence judgment reaction time	–0.110	–3.517	<.001

Predictor random effects	Variance explained	χ^2	*P*
Word	1.555	1611.5	<.0001
Word: Reading comprehension	0.134	60.542	<.0001
Student	2.245	513.82	<.0001
Student: cognate status	1.538	193.29	<.0001
Student: Levenshtein distance	0.359	126.29	<.0001

Note: [a]Tested one-tailed.

distance$_{resid}$. The variable *target word frequency* was no longer significant. A summary of the final model is presented in Table 4.

There were main effects of *semantic contextual support*, *reading comprehension*, *sentence judgment accuracy*, *cognate status*, and *Levenshtein distance.* The main effect of *semantic contextual support* was similar to the effect in the vocabulary knowledge model, $b = 0.086$, 95% CI [0.027, 0.145]. This provides evidence that larger semantic contextual support results in larger learning gains. The influence of the included word and student predictors in this learning gain model was also similar to the influence of these predictors in the vocabulary knowledge model. Finally, there was an interaction between the *sentence judgment accuracy* and *reaction time*, $b = –0.110$, 95% CI [–0.171, –0.049].

Predictors of rate of forgetting

We created a model to look at *rate of forgetting* to see how well the newly learnt vocabulary was retained. In this model, the dependent variable was accuracy on immediate posttest and delayed posttest. The independent variables included in

Table 5. *Summary of a generalized linear mixed-effects model predicting rate of forgetting*

Predictor fixed effects	β	*Z*	*p*
Intercept	–3.548	–11.088	<.0001
Time	0.122	2.369	<.05[a]
Semantic contextual support	–0.019	–0.692	*ns*
Word predictors			
Cognate status (false friends)	2.765	6.247	<.0001
Cognate status (control words)	4.059	9.278	<.0001
Levenshtein distance	1.219	5.196	<.0001
Target frequency	–0.732	–3.919	<.0001
Student predictors			
Reading comprehension	–0.552	–5.884	<.0001
Task performance - Sentence judgment accuracy	0.015	0.458	*ns*
Task performance - Sentence judgment reaction time	–0.452	–5.043	<.0001
Interactions			
Sentence judgment accuracy: sentence judgment reaction time	<.01	2.573	<.01
Predictor random effects	**Variance explained**	χ^2	*p*
Word	2.618	3294.4	<.0001
Word: Reading comprehension	0.087	41.583	<.0001
Student	0.854	465.75	<.0001
Student: cognate status	0.576	95.183	<.0001
Student: Levenshtein distance	0.102	32.252	<.0001
Student: target frequency	0.152	25.812	<.0001

Note: [a]Tested one-tailed.

this model were *time* (immediate posttest vs. delayed posttest), *semantic contextual support*, *reading comprehension*, *sentence judgment accuracy*, *sentence judgment reaction time*, *cognate status*, *Levenshtein distance*$_{resid}$, and *target frequency*. A summary of the final model can be found in Table 5.

There were main effects of *time*, *reading comprehension*, *sentence judgment accuracy, cognate status*, *Levenshtein distance*, and *target frequency*. However, as opposed to vocabulary knowledge at immediate posttest, there was no longer a main effect of *contextual support*. Hence, rate of forgetting does not appear to be influenced by *contextual support*. As there was no significant negative effect of *contextual support* either, this suggests that there are long-term effects of LSA on word learning (similar to the short-term effects established in models above). The main effect of time indicated that students knew fewer words at the delayed posttest than at the immediate posttest, $b = 0.122$, 95% CI [0.021, 0.223] and forgetting took place. We explored the interactions between time and the other variables. However, including these interactions did not improve model fit.

DISCUSSION

The present study investigated the effects of semantic contextual support and various word and student characteristics, and their interactions on L2 vocabulary learning outcomes (vocabulary knowledge, learning gain, and rate of forgetting) obtained by means of a computerized L2 vocabulary learning task in context in Dutch seventh-grade students. We addressed the effects of semantic contextual support on L2 vocabulary learning outcomes using a continuous measure (LSA; Landauer et al., 1998) for the first time, rather than a categorical measure of semantic relatedness in a vocabulary learning study. We found stronger learning gains for more supportive contexts, in line with previous research (Beck et al., 1983; Chrabaszcz & Gor, 2017; Daneman & Green, 1986; Elgort et al., 2015; Ellis, 2013; Howard & Kahana, 2002; Ma et al., 2015; Nassaji, 2003). Whereas these studies had more exposure trials (e.g., Elgort et al., 2015), we demonstrated the effects of semantic contextual support even in vocabulary learning with merely a single exposure in a sentence context. In addition, we showed that the rate of forgetting was not influenced by semantic contextual support. This suggests that there are long-term effects of LSA on word learning (similar to the short-term effects established for vocabulary knowledge and learning gain). Previous studies already showed an effect of contrasting contexts (e.g., Bolger et al., 2008). Here, we provided evidence that even subtle semantic variations can make a difference for L2 vocabulary learning outcomes, in an understudied group of L2 learners: adolescents in secondary school.

Further, we found partial evidence of the influence of word characteristics, including L1-L2 overlap, on L2 vocabulary knowledge and learning gain. Regarding L1-L2 overlap, we found that L2 vocabulary learning outcomes differed across words with varying cognate status. Cognates were easier to translate and retain, compared to both false friends and control words, which is in line with previous studies (de Groot, 1992; de Groot et al., 1994; Dijkstra et al., 1999; Hancin-Bhatt & Nagy, 1994). However, previous studies also demonstrated that false cognates were harder to recognize or learn than control words. Possibly, the control words were too difficult for the inexperienced L2 learners in the present study after all. It has previously been shown that the cognate facilitation effect is reduced for highly proficient L2 learners (Bultena, Dijkstra, & van Hell, 2014). However, we did not replicate this finding, which may be explained by the fact that we used a reading comprehension task instead of a fluency or standardized vocabulary task as a measure of English proficiency. These measures may also reflect English proficiency in a different way than in the study by Bultena et al. (2014). Furthermore, cognate facilitation effects have been shown to depend on task demands; Bultena et al. (2014) used a self-paced reading task and eye movements, whereas we used sentence verification and word typing. Thus, the difference in task demands between self-paced reading on the one hand (Bultena et al., 2014) and sentence verification, as was used in our study, where the entire sentence was presented at once, on the other, may explain this difference in results. The second measure for L1-L2 overlap was Levenshtein distance, which also predicted L2 vocabulary learning outcomes, after controlling for cognate status. Words with smaller Levenshtein distances were easier to learn compared to words with larger Levenshtein distances. This measure has been used to reveal the structure of the mental lexicon (Dautriche et al., 2017), but has not before been examined as a predictor for L2 vocabulary learning outcomes. Apparently, Levenshtein distance can be used to predict variability in L2 vocabulary learning outcomes due to lexical cross-linguistic influences. We are the first to show the effect of Levenshtein distance on L2 vocabulary learning outcomes. We found no effect of word length, whereas previous studies do point toward the predictive effect of word length on L2 vocabulary learning (Hauk & Pulvermüller, 2004; Whaley, 1978). It has been argued that this effect is hard to disentangle from word frequency (e.g., Hauk & Pulvermüller, 2004). On the one hand, short words are possibly easier to remember, but on the other hand, short words are also more easily confused with other words than longer words. Whereas we found an effect of target frequency on vocabulary knowledge, this did not persist in the learning gain model. Possibly, this is because words that were translated correctly during the pretest and posttest were omitted, and these words typically had high frequencies (and were often cognates). Hence, there was not only a reduction of statistical power in the learning gain model but also a reduction in terms of the variability in word frequency in this model. We did not find any interactions between time and the other variables for rate of forgetting. This suggests that although forgetting takes place, this effect is not mediated by any of the other variables.

With respect to the role of student characteristics, our hypothesis was partly confirmed; we found that prior vocabulary knowledge contributed to vocabulary knowledge. This corroborates the previous finding that students with a larger prior vocabulary were better at recognizing novel word structures (Nation & Snowling, 1998). Reading comprehension was found to be a predictor of all three L2 vocabulary learning outcomes. Students with good reading comprehension skills also made larger gains. This is in line with previous studies that showed that poor comprehenders performed weakly on making inferences from text (e.g., Cain et al., 2004). Exposure to English media did not contribute to L2 vocabulary learning outcomes. English exposure may have already been reflected in reading comprehension skills (Kuppens, 2010), and media exposure does not appear to have an additional influence on L2 vocabulary learning outcomes on top of reading comprehension. Our findings were similar across different educational tracks and in students with varying linguistic backgrounds.

There was evidence of interactions between the aforementioned variables. We found that L2 vocabulary learning outcomes were higher for students with higher sentence judgment accuracies, but only when they also responded relatively quickly. The influence of task performance in general had been shown previously (Sense, Meijer, & van Rijn, 2016), and here we specifically demonstrated the influence of within-task behavior on later vocabulary learning outcomes. A speed–accuracy trade-off may have induced students to make more mistakes during sentence judgment when they had low reaction times. Furthermore, the relationship between sentence verification and reading comprehension in Dutch EFL learners was previously demonstrated (van Gelderen et al., 2004). However, we demonstrated the unique contribution of sentence judgment speed combined with accuracy to L2 vocabulary learning outcomes. Regarding word frequency, Monaghan, Chang, Welbourne, and Brysbaert (2017) found that vocabulary size can reduce frequency effects in lexical processes. We found that higher word frequency resulted in higher vocabulary knowledge at immediate posttest, especially when prior vocabulary knowledge was relatively high. This could be explained by the fact that we had already found a ceiling effect on the pretest for cognates. Students had already correctly translated many of the cognates before the vocabulary learning task, leaving less room for improvement. We did not find any interactions between contextual support and any of the other predictors, in contrast with previous studies. This could be due to the fact that single exposures during vocabulary learning were utilized in the present study, whereas other studies (e.g., Elgort et al., 2015) used repeated exposure trials. This could also be due to the preexposure to the target words, which was included to allow learners to create a lexical entry (if there was none). The fact that we did not find any two-way interactions between context and word predictors or student predictors suggests that students benefit from contextual support, regardless of their proficiency and word predictors.

There are several matters that may have limited the study. First, we used pretest accuracy as a measure for prior vocabulary, and no standardized vocabulary measure was taken into account. This may have affected the results, as students showed ceiling effects for cognates on prior vocabulary knowledge. A standardized vocabulary test could be used in future research as a predictor for vocabulary learning. Second, we used a cloze task to measure reading comprehension. It may be argued that reading text passages and answering comprehension question is more representative for vocabulary learning through sentence verification. Third and finally, though we contributed to the literature of vocabulary learning through a single exposure, it would be useful to know what the effect is of several exposure trials on L2 vocabulary learning and the rate of forgetting in adolescents. This has been shown to be effective in other age groups as well (e.g., Bolger et al., 2008).

Based on this study we have suggestions for further research. First, it would be interesting to look at the influence of repeated exposure to a word and subtle semantic contextual support differences to get closer to a natural situation in which L2 learners encounter words repeatedly and possibly in different ways. Second, it might be interesting to administer another delayed posttest. In the present study, the delayed posttest was administered a day after the vocabulary learning task and the immediate posttest. Differences between educational tracks might also be explained by this delayed posttest. It is possible that short-term results appear similar across educational tracks, but differences emerge a longer period of time after the intervention, as words are harder to consolidate for students in the lower tracks. The addition of another test may show more individual differences in vocabulary learning consolidation. Third and finally, concentration (Bialystok, 2015) or L1 fluency (Alderson et al., 2016; Huensch &

Ventura, 2017) could be relevant to take into account in L2 vocabulary learning as student characteristics.

The practical implications of this study are that a semantic supportive context benefits L2 vocabulary acquisition. This can be useful for teachers in secondary education. In addition, L2 vocabulary learning methods should focus on words with different degrees of Dutch–English overlap. Finally, individual learners' characteristics can be taken into account to signal possible difficulties and utilize strengths in L2 vocabulary learning.

In conclusion, we were the first to show that L2 study materials containing more semantic supportive contexts and materials with a focus on words with small L1-L2 overlap are most effective for L2 vocabulary learning outcomes.

APPENDIX A

Stimuli used in the word learning experiment. For each item, the English target word, its Dutch translation, the phonetic transcription of the Dutch cognate or false friend, Levenshtein distance (LD), target word frequency, target word length, and two primes, and corresponding sentences with varying semantic distances (LSA) between primes and targets are displayed.

Item number	English target	Dutch translation	Transcription	LD	Target frequency	Target length	Prime 1	Sentence 1	LSA 1	Prime 2	Sentence 2	LSA 2
Cognates												
1	film	film	/‘fɪləm/	0	1555	4	cinema	He visits the cinema and watches a film	.7	friend	He visits a friend and watches a film	0.01
2	harp	harp	/‘hɑrəp/	0	43	4	play	I would love to play a harp	0.40	have	I would love to have a harp	0.14
3	storm	storm	/‘stɔrəm/	0	454	5	lightning	He saw lightning in the storm	0.69	people	He saw people in the storm	0.14
4	clock	klok	/‘klɔk/	2	637	5	ticking	She listens to the ticking of the clock	0.52	sound	She listens to the sound of the clock	0.06
5	cliff	afgrond	/‘klɪf/	5	294	5	climbed	They climbed close to the cliff	0.50	stayed	They stayed close to the cliff	0.02
6	ankle	enkel	/‘ɛŋkəl/	4	185	5	shoe	The shoe comes up to my ankle	0.27	sand	The sand comes up to my ankle	0.11
7	humour	humor	/‘hymɔr/	1	431	6	joke	Their jokes are full of humour	0.23	story	Their stories are full of humour	0.04
8	boat	boot	/‘bot/	1	1000	4	row	She wants to row the boat	0.74	stare	She wants to stare at the boat	0.1
9	wheel	wiel	/‘wil/	2	499	5	bike	The child’s bike had an extra wheel	0.48	toy	The child’s toy had an extra wheel	0.12
10	wild	wild	/‘wɪlt/	0	1551	4	animals	The animals look wild	0.48	women	The plants look wild	0.09
11	code	code	/‘kodə/	0	447	4	message	The message was written in code	0.33	story	The story was written in code	0.05
12	bride	bruid	/‘brœyt/	2	187	5	husband	The husband looks at the bride	0.68	boy	The boy looks at the bride	0.10
13	apple	appel	/‘ɑpəl/	2	315	5	cherries	He wants his cherries and an apple	0.51	teddy bear	He wants his teddy bear and an apple	0.05
14	bread	brood	/‘brot/	2	1327	5	baker	They go to the baker to buy some bread	0.29	shop	They go to the shop to buy some bread	0.11
15	thunder	donder	/‘dɔndər/	3	187	7	raining	It was raining and there was thunder	0.48	March	It was March and there was thunder	0.02
16	dry	droog	/‘drox/	3	1459	3	desert	The desert is extremely dry	0.65	earth	The earth is extremely dry	0.10

APPENDIX A (*cont.*)

Item number	English target	Dutch translation	Transcription	LD	Target frequency	Target length	Prime 1	Sentence 1	LSA 1	Prime 2	Sentence 2	LSA 2
17	dream	droom	/ʻdrom/	2	42	5	night	Last night I had a dream	0.52	Monday	Last Monday I had a dream	0.14
18	honey	honing	/ʻhonɪŋ/	3	371	5	bees	They use bees to make honey	0.85	sugars	They use sugars to make honey	0.09
19	guitar	gitaar	/giʼtar/	2	102	6	sound	They love the sound of that guitar	0.55	look	They love the look of that guitar	0.11
20	ring	ring	/ʻrɪŋ/	0	804	4	wear	She loves to wear a ring	0.21	buy	She loves to buy a ring	0.07
21	snow	sneeuw	/ʻsneːu/	3	1040	4	mountains	The mountains were covered with snow	0.34	cars	The cars were covered with snow	0.06
22	ball	bal	/ʻbɑl/	1	1664	4	throwing	He is throwing the ball	0.76	using	He is using the ball	0.08
23	bed	bed	/ʻbɛt/	0	4376	3	pillow	The pillow was in his bed	0.81	boy	The boy was in his bed	-0.04
24	book	boek	/ʻbuk/	1	4857	4	pages	There were many pages in this book	0.62	animals	There were many animals in this book	0.02
25	bus	bus	/ʻbʏs/	0	1155	3	seat	She has a seat on the bus	0.50	snack	She has a snack on the bus	0.11
26	concert	concert	/kɔnʼsɛrt/	0	272	7	clapping	He was clapping during the concert	0.80	standing	He was standing during the concert	0.11
27	dance	dans	/ʻdɑns/	2	605	5	theatre	In the theatre she saw girls dance	0.42	park	In the park she saw girls dance	0.09
28	toilet	toilet	/twɑʻlɛt/	0	415	6	bathroom	The bathroom had a toilet	0.48	house	The house had a toilet	0.13
29	better	beter	/ʻbetər/	1	83	6	improved	She improved and now she's better	0.48	food	She had food and now she's better	0.12
30	hope	hoop	/ʻhop/	2	578	4	rescue	His rescue gave them hope	0.85	action	His action gave them hope	0.14
31	tent	tent	/ʻtɛnt/	0	657	4	camp	He likes to camp in a tent	0.63	be	He likes to be in a tent	0.22
32	tone	toon	/ʻton/	2	786	4	voice	His voice has a serious tone	0.61	book	His book has a serious tone	0.08
False Friends												
33	step	stap	/ʻstɛp/	1	952	4	walk	I walk with a quick light step	0.31	run	I run with a quick light step	0.13
34	star	ster	/ʻstɑr/	1	952	4	sky	In the sky he sees a star	0.64	flash	In a flash he sees a star	0.09
35	box	doos	/ʻbɔks/	3	704	3	shoe	She puts the shoe back in the box	0.44	stuff	She puts the stuff back in the box	0.07
36	spot	vlek	/ʻspɔt/	4	755	4	face	On his face was a brown spot	0.29	Paper	On his paper was a brown spot	0.10
37	arts	kunsten	/ʻɑrts/	5	826	4	painter	The painter was a fan of arts	0.54	woman	The woman was a fan of arts	0.04
38	bond	band	/ʻbɔnt/	1	295	4	weak	The brothers have a weak bond	0.47	strong	The brothers have a strange bond	0.08
39	stage	podium	/ʻstaʒə/	5	2453	5	actor	The actor is standing on stage	0.67	girl	The girl is standing on stage	0.07
40	glad	opgelucht	/ʻxlɑt/	5	1146	4	help	Her help makes me feel glad	0.25	music	Her music makes me feel glad	0.07
41	roof	dak	/ʻrof/	4	831	4	windows	There are windows in the roof	0.67	birds	There are birds on the roof	0.10
42	steel	staal	/ʻstel/	2	755	5	iron	It was made of iron and steel	0.80	glass	It was made of glass and steel	0.04

43	boot	laars	/‘bot/	5	159	4	shoe	That shoe looks like a boot	0.30	object	That object looks like a boot	-0.01
44	tube	buis	/‘tybe/	3	264	4	juice	She drinks juice through the tube	0.27	something	She drinks something through the tube	0.06
45	brave	dapper	/‘bravə/	5	346	5	hero	What the hero did was very brave	0.39	child	What the child did was very brave	0.10
46	brand	merk	/‘brɑnt/	5	189	5	advertisement	The advertisement is from a nice brand	0.53	soap	The soap is from a nice brand	0.14
47	note	briefje	/‘not/	6	1257	4	writes	She writes a short note	0.24	hands over	She hands over a short note	0.15
48	aid	hulp	/‘et/	4	986	3	victim	The man gave the victim first aid	0.41	girl	The man gave the girl first aid	0.04
49	lake	meer	/‘lakə/	3	718	4	water	There was water in the lake	0.35	plants	There were plants in the lake	0.14
50	bang	knal	/‘bɑŋ/	4	11	4	exploded	It exploded with a bang	0.39	dropped	It dropped with a bang	0.11
51	back	rug	/‘bɑk/	4	22071	4	bag	He carries a bag on his back	0.50	child	He carries a child on his back	0.12
52	cook	kok	/‘kok/	2	381	4	food	He loves the food made by this cook	0.32	creations	He loves the creations made by this cook	0.04
53	hate	haat	/‘hatə/	2	450	4	wrong	Being wrong is something I hate	0.66	sick	Being sick is something I hate	0.07
54	road	weg	/‘rot/	4	3791	4	car	The car was on the road	0.53	dog	The dog was on the road	0.08
55	safe	veilig	/‘sef/	6	1426	4	helmet	She has a helmet and feels safe	0.25	stress	She has no stress and feels safe	0.00
56	stout	moedig	/‘stʌut/	6	159	5	sheriff	They said the sheriff was very stout	0.41	child	They said the child was very stout	0.07
57	rover	zwerver	/‘rovər/	4	70	5	wild	The wild man was a rover	0.22	old	The old man was a rover	0.12
58	brink	rand	/‘brɪŋk/	3	107	5	war	They are on the brink of war	0.32	fight	They are on the brink of a fight	0.14
59	lover	geliefde	/‘lɔvər/	7	430	5	marry	He wants to marry his lover	0.47	talk	He wants to talk to his lover	0.12
60	lap	schoot	/‘lɑp/	6	324	3	baby	The baby sat on the doctor’s lap	0.35	child	The child sat on the doctor’s lap	0.13
61	boon	zegen	/‘bon/	4	715	4	fortune	He thinks having a fortune is a real boon	0.21	cat	He thinks having a cat is a real boon	-0.01
62	pace	tempo	/‘pes/	5	581	4	running	When he is running he changes his pace	0.40	hockey	When he is playing hockey he changes his pace	0.18
63	mind	geest	/‘mɪnt/	5	5530	4	thoughts	Those thoughts stayed in her mind	0.67	pictures	Those pictures stayed in her mind	0.14
64	need	noodzaak	/‘net/	5	4337	4	helped	He helped, though there was no need	0.27	laughed	He laughed, though there was no need	0.12
Control words												
65	moon	maan	—	2	951	4	astronauts	Some astronauts walk on the moon	.65	people	Some people walk on the moon	0.09
66	ease	gemak	—	4	456	4	simple	It was simple and done with ease	0.31	nice	It was nice and done with ease	0.14
68	corn	mais	—	4	429	4	farm	She goes to the farm to get corn	0.57	shop	She goes to the shop to get corn	0.02
69	crack	barst	—	5	250	5	earth	The earth opened with a crack	0.27	coconut	The coconut opened with a crack	0.10

APPENDIX A (*cont.*)

Item number	English target	Dutch translation	Transcription	LD	Target frequency	Target length	Prime 1	Sentence 1	LSA 1	Prime 2	Sentence 2	LSA 2
70	dusk	schemer	—	7	107	4	dark	It is dark in the dusk	0.61	scary	It is scary in the dusk	0.04
71	cheek	hout	—	5	1166	5	kissed	She kissed him on the cheek	0.40	toys	She hit him on the cheek	0.02
72	wood	wang	—	3	268	4	trees	Trees are made out of wood	0.68	hit	Toys are made out of wood	-0.04
73	near	dichtbij	—	8	434	4	sea	The sea and the beach are near	0.39	church	The church and the palace are near	0.09
74	cave	grot	—	4	1081	4	tunnels	There were tunnels in the cave	0.48	rats	There were rats in the cave	0.15
75	ledge	rand	—	5	491	5	rock	He walks on the rock and stands on a ledge	0.40	street	He walks on the street and stands on a ledge	0.13
76	fibre	vezel	—	5	148	5	wool	Many times, wool is full of fibre	0.21	food	Many times, food is full of fibre	0.15
77	chess	schaak	—	4	264	5	game	They like a game of chess	0.54	round	They like a round of chess	0.13
78	cabbage	kool	—	7	1023	7	cook	They wanted to cook a lot of cabbage	0.39	buy	They wanted to buy a lot of cabbage	0.12
79	odd	ongewoon	—	7	144	3	stupid	The stupid man is very odd	0.41	fat	The fat man is very odd	0.16
80	ache	pijn	—	4	1069	4	hurt	I hurt my leg and felt an ache	0.47	moves	I moved my leg and felt an ache	0.04
81	pond	vijver	—	6	257	4	water	There is a lot of water in the pond	0.44	dirt	There is a lot of dirt in the pond	0.13
82	chart	kaart	—	2	238	5	numbers	The numbers are on the chart	0.34	letters	The letters are on the chart	0.10
83	dust	stof	—	4	762	4	clean	She doesn't want to clean the dust	0.31	hide	She doesn't want to see the dust	0.15
84	seat	stoel	—	4	1396	4	driver	The driver was in his seat	0.55	man	The man was in his seat	0.13
85	desk	bureau	—	5	1473	4	office	In his office he had a desk	0.47	house	In his house he had a desk	0.09
86	war	oorlog	—	5	6082	3	fighting	People were fighting during the war	0.69	crying	People were crying during the war	0.04
87	sign	bord	—	4	1557	4	walking	The walking route was on the sign	0.44	cycling	The cycling route was on the sign	0.01
88	part	rol	—	4	8361	4	actor	The actor only has a small part	0.27	woman	The woman only has a small part	0.12
89	farmer	boer	—	4	1017	6	chickens	The chickens stay with the farmer	0.64	boys	The boys stay with the farmer	0.07
90	tears	tranen	—	4	995	5	eyes	Her eyes were full of tears	0.69	paper	The paper was full of tears	0.07
91	branch	tak	—	5	961	6	tree	He sat in the tree on a branch	0.52	forest	He sat in the forest on a branch	0.06
92	judge	rechter	—	6	732	5	criminal	The criminal had to see a judge	0.78	woman	The woman had to see a judge	0.13
93	vote	stem	—	3	477	4	president	The president had his vote	0.35	person	That person had his vote	0.12
94	goal	doel	—	2	531	4	player	The player made a goal	0.45	man	The man made a goal	0.08
95	rope	touw	—	3	552	4	cowboy	The cowboy wants to use the rope	0.27	girl	The girl wants to use the rope	0.13
96	acids	zuren	—	5	107	5	lemons	Lemons are full of acids	0.51	bombs	Bombs are full of acids	0.00

ACKNOWLEDGMENTS

This research was supported by Grant nr 405-14-304 from the NRO Programme Council for Educational Research (PROO). We thank all university students, participants, schools, and staff that helped to make this project possible. We would also like to thank Hubert Voogd from the Technical Support Group, Faculty of Social Sciences, Radboud University, for his valuable technical support in the creation of the experiment. Finally, we thank Bernard Westwell for his useful advice.

REFERENCES

Adelman, J. S., Brown, G. D. A., & Quesada, J. F. (2006). Contextual diversity, not word frequency, determines word naming and lexical decision times. *Psychological Science*, *17*, 814–823.

Adlof, S., Frishkoff, G., Dandy, J., & Perfetti, C. (2016). Effects of induced orthographic and semantic knowledge on subsequent learning: A test of the partial knowledge hypothesis. *Reading and Writing*, *29*, 475–500.

Alderson, J. C., Nieminen, L., & Huhta, A. (2016). Characteristics of weak and strong readers in a foreign language. *Modern Language Journal*, *100*, 853–879.

Audacity Team. (2015). Audacity®: Free Audio Editor and Recorder. Version 2.1.2. Retrieved October, 29, 2015, from http://audacity.sourceforge.net/

Baayen, R. H. (2008). *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge: Cambridge University Press.

Baayen, R. H., Piepenbrock, R., & van Rijn, H. (1993). *The CELEX lexical database [CD-ROM]*. Philadelphia, PA: University of Pennsylvania Linguistic Data Consortium.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*, 1–48.

Beck, I. L., McKeown, M. G., & McCaslin, E. S. (1983). Vocabulary development: All contexts are not created equal. *Elementary School Journal*, *83*, 177–181.

Belsley, D. A., Kuh, E., & Welsh, R. E. (1980). *Regression diagnostics: Identifying influential data and sources of collinearity*. New York: Wiley.

Bialystok, E. (2015). Bilingualism and the development of executive function: The role of attention. *Child Development Perspectives*, *9*, 117–121.

Boersma, P. (2001). Praat: A system for doing phonetics by computer. *Glot International*, *5*, 341–345.

Bolger, D. J., Balass, M., Landen, E., & Perfetti, C. A. (2008). Context variation and definitions in learning the meanings of words: An instance-based learning approach. *Discourse Processes*, *45*, 122–159.

Brenders, P., van Hell, J. G., & Dijkstra, T. (2011). Word recognition in child second language learners: Evidence from cognates and false friends. *Journal of Experimental Child Psychology*, *109*, 383–396.

Breslow, N. E., & Clayton, D. G. (1993). Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association*, *88*, 9–25.

Bultena, S., Dijkstra, T., & van Hell, J. G. (2014). Cognate effects in sentence context depend on word class, L2 proficiency, and task. *Quarterly Journal of Experimental Psychology*, *67*, 1214–1241.

Cain, K., Oakhill, J., & Lemmon, K. (2004). Individual differences in the inference of word meanings from context: The influence of reading comprehension, vocabulary knowledge, and memory capacity. *Journal of Educational Psychology*, *96*, 671.

Chrabaszcz, A., & Gor, K. (2017). Quantifying contextual effects in second language processing of phonolexically ambiguous and unambiguous words. *Applied Psycholinguistics*, *38*, 909–942.

Crossley, S. A., Salsbury, T., & McNamara, D. S. (2012). Predicting the proficiency level of language learners using lexical indices. *Language Testing*, *29*, 243–263.

Daneman, M., & Green, I. (1986). Individual differences in comprehending and producing words in context. *Journal of Memory and Language*, *25*, 1–18.

Dautriche, I., Mahowald, K., Gibson, E., Christophe, A., & Piantadosi, S. T. (2017). Words cluster phonetically beyond phonotactic regularities. *Cognition*, *163*, 128–145.

de Groot, A. M. (1992). Determinants of word translation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*, 1001–1018.

de Groot, A. M., Dannenburg, L., & van Hell, J. G. (1994). Forward and backward word translation by bilinguals. *Journal of Memory and Language*, *33*, 600–629.

de Groot, A. M., & Keijzer, R. (2000). What is hard to learn is easy to forget: The roles of word concreteness, cognate status, and word frequency in foreign-language vocabulary learning and forgetting. *Language Learning*, *50*, 1–56.

de Groot, A. M., & Nas, G. L. (1991). Lexical representation of cognates and noncognates in compound bilinguals. *Journal of Memory and Language*, *30*, 90–123.

Diependaele, K., Lemhöfer, K., & Brysbaert, M. (2013). The word frequency effect in first- and second-language word recognition: A lexical entrenchment account. *Quarterly Journal of Experimental Psychology*, *66*, 843–863.

Dijkstra, T., Grainger, J., & van Heuven, W. J. (1999). Recognition of cognates and interlingual homographs: The neglected role of phonology. *Journal of Memory and language*, *41*, 496–518.

Elgort, I., Perfetti, C. A., Rickles, B., & Stafura, J. Z. (2015). Contextual learning of L2 word meanings: Second language proficiency modulates behavioural and event-related brain potential (ERP) indicators of learning. *Language, Cognition and Neuroscience*, *30*, 506–528.

Elgort, I., & Warren, P. (2014). L2 vocabulary learning from reading: Explicit and tacit lexical knowledge and the role of learner and item variables. *Language Learning*, *64*, 365–414.

Ellis, N. C. (2013). Second language acquisition. In G. Trousdale & T. Hoffmann (Eds.), *Oxford handbook of construction grammar* (pp. 365–378). Oxford: Oxford University Press.

Ellis, S. M., & Steyn, H. S. (2003). Practical significance (effect sizes) versus or in combination with statistical significance (*p*-values): Research note. *Management Dynamics*, *12*, 51–53.

Forster, K. I. (1976). Accessing the mental lexicon. In E. W. Walker (Ed.), *Explorations in the biology of language* (pp. 139–174). Montgomery, VT: Bradfort Books.

Gellert, A. S., & Elbro, C. (2013). Cloze tests may be quick, but are they dirty? Development and preliminary validation of a cloze test of reading comprehension. *Journal of Psychoeducational Assessment*, *31*, 16–28.

Hancin-Bhatt, B., & Nagy, W. (1994). Lexical transfer and second language morphological development. *Applied Psycholinguistics*, *15*, 289–310.

Hauk, O., & Pulvermüller, F. (2004). Effects of word length and frequency on the human event-related potential. *Clinical Neurophysiology*, *115*, 1090–1103.

Howard, M. W., & Kahana, M. J. (2002). When does semantic similarity help episodic retrieval? *Journal of Memory and Language*, *46*, 85–98.

Huensch, A., & Ventura, N. (2017). Understanding second language fluency behavior: The effects of individual differences in first language fluency, cross-linguistic differences, and proficiency over time. *Applied Psycholinguistics*, *38*, 755–785.

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language*, *59*, 434–446.

Keenan, J. M., & Meenan, C. E. (2014). Test differences in diagnosing reading comprehension deficits. *Journal of Learning Disabilities*, *47*, 125–135.

Kuppens, A. H. (2010). Incidental foreign language acquisition from media exposure. *Learning, Media and Technology*, *35*, 65–85.

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review*, *104*, 211–240.

Landauer, T. K., Foltz, P. W., & Laham, D. (1998). An introduction to latent semantic analysis. *Discourse Processes*, *25*, 259–284.

Levenshtein, V. I. (1966). Binary codes capable of correcting deletions, insertions and reversals. *Soviet Physics Doklady*, *10*, 707–710.

Lindgren, E., & Muñoz, C. (2013). The influence of exposure, parents, and linguistic distance on young European learners' foreign language comprehension. *International Journal of Multilingualism*, *10*, 105–129.

Lotto, L., & de Groot, A. M. (1998). Effects of learning method and word type on acquiring vocabulary in an unfamiliar language. *Language Learning*, *48*, 31–69.

Ma, T., Chen, B., Lu, C., & Dunlap, S. (2015). Proficiency and sentence constraint effects on second language word learning. *ActaPsychologica*, *159*, 116–122.

Martens, R. L., Gulikers, J., & Bastiaens, T. (2004). The impact of intrinsic motivation one-learning in authentic computer tasks. *Journal of Computer Assisted Learning*, *20*, 368–376.

Monaghan, P., Chang, Y. N., Welbourne, S., & Brysbaert, M. (2017). Exploring the relations between word frequency, language exposure, and bilingualism in a computational model of reading. *Journal of Memory and Language*, *93*, 1–21.

Nassaji, H. (2003). Higher-level and lower-level text processing skills in advanced ESL reading comprehension. *Modern Language Journal*, *87*, 261–276.

Nation, K., & Snowling, M. J. (1998). Semantic processing and the development of word-recognition skills: Evidence from children with reading comprehension difficulties. *Journal of Memory and Language*, *39*, 85–101.

Nation, K., & Snowling, M. J. (2004). Beyond phonological skills: Broader language skills contribute to the development of reading. *Journal of Research in Reading*, *27*, 342–356.

Perfetti, C. A., & Hart, L. (2002). The lexical quality hypothesis. *Precursors of Functional Literacy*, *11*, 67–86.

Pytlyk, C. (2017). Are orthographic effects language specific? The influence of second language orthography on second language phoneme awareness. *Applied Psycholinguistics*, *38*, 233–262.

Rayner, K., Ashby, J., Pollatsek, A., & Reichle, E. D. (2004). The effects of frequency and predictability on eye fixations in reading: Implications for the EZ Reader model. *Journal of Experimental Psychology: Human Perception and Performance*, *30*, 720.

Reichle, E. D., & Perfetti, C. A. (2003). Morphology in word identification: A word-experience model that accounts for morpheme frequency effects. *Scientific Studies of Reading*, *7*, 219–237. doi:10.1207/s1532799xssr0703_2

Seidenberg, M. S., & McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. *Psychological Review*, *96*, 523–568.

Sense, F., Meijer, R. R., & van Rijn, H. (2016). On the Link between Fact Learning and General Cognitive Ability. In A. Papafragou, D. Grodner, D. Mirman, & J. C. Trueswell (Eds.), *Proceedings of the 38th Annual Conference of the Cognitive Science Society*. Austin, TX: Cognitive Science Society.

Starreveld, P. A., de Groot, A. M., Rossmark, B. M., & van Hell, J. G. (2014). Parallel language activation during word processing in bilinguals: Evidence from word production in sentence context. *Bilingualism: Language and Cognition*, *17*, 258–276.

University of Colorado at Boulder. (2003). *Latent Semantic Analysis online application.* Available at Latent Semantic Analysis @ CU Boulderwebsite, www.lsa.colorado.edu

van de Ven, M., Tucker, B. V., & Ernestus, M. (2011). Semantic context effects in the comprehension of reduced pronunciation variants. *Memory & Cognition*, *39*, 1301–1316.

van Gelderen, A., Schoonen, R., de Glopper, K., Hulstijn, J., Simis, A., Snellings, P., & Stevenson, M. (2004). Linguistic knowledge, processing speed, and metacognitive knowledge in first- and second-language reading comprehension: A componential analysis. *Journal of Educational Psychology*, *96*, 19.

van Hell, J. G., & Dijkstra, T. (2002). Foreign language knowledge can influence native language performance in exclusively native contexts. *Psychonomic Bulletin & Review*, *9*, 780–789.

Whaley, C. P. (1978). Word–nonword classification time. *Journal of Verbal Learning and Verbal Behavior*, *17*, 143–154.

Wurm, L. H., & Fisicaro, S. A. (2014). What residualizing predictors in regression analyses does (and what it does not do). *Journal of Memory and Language*, *72*, 37–48.

Zhang, D., Chin, C. F., & Li, L. (2017). Metalinguistic awareness in bilingual children's word reading: A cross-lagged panel study on cross-linguistic transfer facilitation. *Applied Psycholinguistics*, *38*, 395–426.