

## Serious game-based word-to-text integration intervention effects in English as a second language

Evelien Mulder<sup>a,\*</sup>, Marco van de Ven<sup>b</sup>, Eliane Segers<sup>b</sup>, Alexander Krepel<sup>c</sup>, Elise H. de Bree<sup>e</sup>, Han van der Maas<sup>d</sup>, Peter F. de Jong<sup>c</sup>, Ludo Verhoeven<sup>b</sup>

<sup>a</sup> Department of Special Education, University of Groningen, the Netherlands

<sup>b</sup> Behavioural Science Institute, Radboud University, the Netherlands

<sup>c</sup> Research Institute of Child Development and Education, University of Amsterdam, the Netherlands

<sup>d</sup> Psychology Research Institute, University of Amsterdam, the Netherlands

<sup>e</sup> Education and Pedagogy - Development & Education of Youth in Diverse Societies, University of Utrecht, the Netherlands

### ABSTRACT

Word-to-text integration (WTI) is the ability to integrate words into a mental representation of the text and is important for reading comprehension, but challenging in English as a second language (ESL). However, it remains unclear whether WTI can be trained in seventh grade ESL learners, who often struggle with reading comprehension and display large individual differences. To pay attention to individual differences, the present study examined an adaptive computer game-based WTI-intervention. The intervention, replacing 50 min of ESL classroom instruction, comprised a 12-week program in which students had to complete WTI-based assignments within four serious games, targeting morphosyntactic awareness, translation of words within sentences, recognizing idioms from words in contexts, and a filler game targeting dictation. The intervention group ( $n = 164$ ) was compared to a control group ( $n = 166$ ), who only received regular ESL classroom instruction. Both groups completed the following reading measures: decoding, morphological, and syntactic awareness, WTI (argument and anomaly reading speed and processing), and reading comprehension tasks at the beginning (T1) of the school year and at the end (T2) of the school year. Results demonstrated an intervention effect on decoding and anomaly processing as reflected by an interaction between time (T1 vs. T2) and group (intervention vs. control) in a repeated measures MANOVA. Follow-up mediation analyses for the intervention group only - with game performance as mediators between reading measures at T1 and T2 - indicated that students with better T1 scores on reading measures showed more growth in performance within games. More performance growth within the translation game and the idiom recognition game was related to better reading scores at T2. Both high-achieving and low-achieving students displayed performance growth within games, indicating that a WTI intervention yields promising results for a broad variety of ESL readers.

### 1. Serious Game-based Word-to-text integration intervention effects in English as a second language

Reading comprehension is challenging, especially in a second language (L2; Lesaux et al., 2006). Comprehension problems often have diverse causes (Cain et al., 2005; Degand & Sanders, 2002; Perfetti & Hart, 2002; Tong et al., 2011). As a result, large individual differences in English as an L2 or foreign language (henceforth ESL) proficiency continue to exist. For teachers, it can be difficult to both differentiate between and meet the needs of higher and lower achieving students (Thijs et al., 2011). In the influential, comprehensive model of reading comprehension, the Reading Systems Framework (Perfetti & Stafura, 2014), Word-to-text integration (WTI; the ability to integrate words that are read into a mental model of the text) is an essential component of both first language (L1) and L2 reading comprehension (Perfetti et al., 2008). However, no studies have yet examined whether WTI in ESL can

be trained and whether individual differences in reading skills and performance within an intervention are predictive of reading development. Furthermore, WTI is an important, yet understudied predictor of reading comprehension and is therefore useful to target within an intervention. Adaptive digital learning environments may make a useful contribution to differentiation (Sandberg et al., 2014). In other words, problems with reading comprehension have diverse sources, thus calling for an adaptive approach instead of a more traditional classroom-based approach. Therefore, the purpose of the present study was to examine the effects of an adaptive computer game-based intervention (henceforth adaptive game-based intervention), targeting WTI on reading comprehension and its predictors. The intervention embodied a 12-week program in which, for 50 min a week, students were to complete WTI-based assignments (series of test items) within four serious games. The effect of the intervention on several reading measures was examined in ESL learners.

\* Corresponding author at: University of Groningen, the Netherlands.

E-mail address: [e.m.b.mulder@rug.nl](mailto:e.m.b.mulder@rug.nl) (E. Mulder).

<https://doi.org/10.1016/j.cedpsych.2021.101972>

### 1.1. Components of reading comprehension

Comprehending text arises from an interaction between knowledge sources and processes (Helder & Perfetti, 2021). To comprehend text, readers need to activate, among others, knowledge about word meaning, morphology, and syntactic structures. The processes involved with reading comprehension entail, among others, decoding and word-to-text integration (WTI). We explain these knowledge sources and processes required for reading comprehension in more detail below.

In order to understand text, words need to be decoded during reading, thus activating the phonological (sound), orthographic (spelling), and semantic (meaning) representation of a word (Ouellette & Beers, 2010). Robust, high quality representations may result in smoother retrieval of words during reading (Perfetti & Hart, 2002; Liu et al., 2017). Intertwined with the semantics of words are morphological and syntactic awareness. Morphological awareness concerns awareness of the smallest meaningful units in words and syntactic awareness refers to the rules about grammar and structure. Morphological awareness can improve the ease with which words are retrieved and in turn can be integrated into a sentence context (Bowers et al., 2010). Syntactic awareness may further help students to integrate words into a syntactic structure and help in binding words in a sentence (Cain, 2007). Both thus help Word-to-text Integration (WTI).

WTI is a part of the reading comprehension process. It encompasses combining words, sentences, and larger units of text into a mental representation of the text. This is a vital component of reading comprehension, according to the influential Reading Systems Framework (Perfetti & Stafura, 2014). During the process of activating single word meanings, WTI can take place to unify words into sentences and larger text units (Perfetti et al., 2008), and is thus a key process in building up a mental model of the text (Stafura et al., 2015). During WTI, words are linked both prospectively, anticipating upcoming words, and retrospectively, linking words to previously read words (Zwaan, 2016). The complexity of a text influences the ease with which WTI can take place (Mulder et al., 2020).

The process of WTI depends on the demands posed by the text, for example as a result of argument overlap or the presence of an anomaly. In some text passages, students need to be able to make an inference between two sentences as a result of argument overlap. For example, in the passage: *'After being dropped from the plane, the bomb hit the ground and exploded. The explosion was quickly reported to the commander'* (Yang et al., 2007) the ease with which the meaning of the word 'explosion' can be inferred and integrated depends on the binding options within the previous sentence in the passage. In other cases, the texts demand readers to semantically bind words to detect possible anomalies and update a mental model of the text accordingly. For example, it is easier to integrate the word 'butter' into a mental model of the passage: *'He spread the warm bread with butter'* than integrating the word 'socks' into a mental model of the passage: *'He spread the warm bread with socks'* (van Berkum et al., 1999). Students need to be able to detect contextual cues to integrate words into a sentence context, but also to recognize idioms (Cain et al., 2009).

### 1.2. Individual differences in reading comprehension

Although some elements required for reading comprehension develop similarly in L1 and L2 learners (Barber et al., 2020; Verhoeven & van Leeuwe, 2012), sources of comprehension problems are often diverse and may be rooted in problems with lexical quality (Melby-Lervåg & Lervåg, 2014) and integration processes (Cain & Oakhill, 2006; Droop et al., 2016; LARCC, 2019). L2 students often have decoding levels comparable to age-matched L1 learners (Verhoeven & van Leeuwe, 2012). However, students learning ESL in secondary school have less linguistic experience than, for example, balanced bilinguals, or L2 learners in immersion programs (Gebauer et al., 2013; Goriot et al., 2018). Moreover, L2 learners often have less L2 vocabulary knowledge

than L1 learners (Melby-Lervåg & Lervåg, 2014). Furthermore, the efficiency of the WTI process, as reflected by online measures of WTI, appears to be lower in L2 compared to L1 learners (Yang et al., 2005). L2 learners often have poorer L2 morphological and syntactic awareness (Tong et al., 2011), may be less sensitive to contextual cues required to bind words within and across sentences (Degand & Sanders, 2002), and show difficulty understanding idioms (Cain et al., 2009). These problems might be explained by the more limited experienced of L2 learners in their L2, resulting in less knowledge about idioms and syntactic structures in that language (Ellis, 2015). Students that have weak linguistic skills in general may also have fewer lexical entries in their L1, resulting in even less information to rely on while reading in an L2 (Koda, 2007). It is thus important to examine if and how WTI and reading comprehension in ESL can be trained with a WTI intervention and to what extent individual differences are predictive of responsiveness to intervention.

#### Reading Comprehension Interventions

Interventions enhancing L1 and L2 reading comprehension often focus on vocabulary (see Edmonds et al., 2009 for a review; Scamacca et al., 2015 for a meta-analysis; Wright & Brown, 2006; Wright, & Cervetti, 2017) or on metacognitive reading strategies (e.g., Muijselaar et al., 2017). Interventions focusing on fluency, vocabulary, word studying, comprehension or a combination of the aforementioned have been shown to positively affect reading comprehension in 11- to 21-year old struggling L1 readers (Edmonds et al., 2009). Although reading strategies improved in an intervention study focused on teaching reading strategies to 11-to 12-year old students, learning French or Spanish as an L2, it was complicated to disentangle whether this intervention indeed improved reading strategies (Wright & Brown, 2006). However, in fourth grade students L1 reading strategies indeed improved after an intervention, but results did not transfer to reading comprehension (Muijselaar et al., 2017). A systematic review of vocabulary intervention effects on reading comprehension in prekindergarten throughout secondary school indicates that actively processing words and monitoring understanding seemed to support comprehension. Nevertheless, the review also found limited evidence that comprehension could be improved through teaching word meaning (Wright & Cervetti, 2017). Although 11-year old L1 and L2 students mostly improve on words that are addressed during the intervention, transfer to other words generally does not take place (e.g., Lesaux et al., 2010).

Few interventions focus on inference making, indicating inference making can be trained in order to improve reading comprehension in 7–8 and 10–11-year-old L1 learners (Elbro & Buch-Iversen, 2013; Yuill & Oakhill, 1988) or on online processing during reading comprehension (e.g., McMaster et al., 2015). The aforementioned interventions do not provide an integrated approach focused on processing text and often do not take into account the aforementioned challenges L2 learners are faced with. Furthermore, despite the available knowledge on WTI and its development, it remains to be unraveled whether WTI can be trained (Yang et al., 2017) in order to foster reading comprehension. As WTI reflects the ability to integrate words into the rest of the text (Perfetti et al., 2008), transfer from the intervention to measures of WTI and reading comprehension may be expected.

The individual differences in (predictors of) reading comprehension require a tailored approach. Computer assisted language learning (CALL) seems promising for this purpose: it can provide integrated language learning and control over the learning process (Dina & Ciornei, 2013) and independent learning with attention for individual differences (Nouri & Pargman, 2016). Furthermore, digital game-based learning has been found to have positive outcomes on language acquisition: L2 learners that were involved in CALL activities outperformed students using traditional practices on language proficiency (Sandberg et al., 2014). More specifically, in terms of instructional design, adaptivity may foster further differentiation within CALL (Sandberg et al., 2014). Previous foreign language teaching programs that provided the

right amount of challenge had positive effects for foreign language learning (Porter, 2019). Many of the studies using CALL included university students (Hung et al., 2018) and focused on vocabulary and general proficiency. Much less attention has been devoted to WTI (Persson & Nouri, 2018). Providing direct feedback has also been found to positively affect L2 learning outcomes (Li, 2010) and should also be taken into account when designing a WTI intervention.

### 1.3. Present study

To summarize, many ESL learners struggle with reading comprehension and the underlying WTI processes. For teachers, it can be difficult to provide instruction and practice suiting the needs of individual students (Thijs et al., 2011) while also providing direct feedback (Li, 2010). Previous intervention studies using CALL seem promising for improving reading comprehension. Thus far, vocabulary and reading strategy interventions have been conducted, showing limited (transfer) effects. In contrast, an adaptive game-based WTI intervention for ESL might facilitate transfer. Such an intervention has not been examined before.

Therefore, in the present study, Dutch seventh grade students followed a 12-week adaptive game-based intervention focused on enhancement of WTI in ESL and were compared to a control group that continued ESL lessons as usual. Students in the intervention group completed WTI-based assignments within four different serious games: one game focused on morphosyntactic awareness, the second on the use of contextual cues by translating words within sentences, the third on recognizing idioms, and finally there was a filler game about dictation. Students were to complete 60 assignments a week within each game, boiling down to a total of 240 assignments a week.

Both groups received between 135 and 180 min of ESL instruction divided across three lessons. In the intervention group, 50 min of the instruction was replaced by the intervention. In terms of instructional design, teachers were supported in two ways: the provided assignments within the intervention were adapted to the level of the students, thus attending to individual differences, and direct feedback was provided, which has been found useful to foster L2 learning (Li, 2010). The intervention was carried out using the digital learning environment *Words&Birds* (Oefenweb (Now: Prowise), 2015). *Words&Birds* (Oefenweb, 2015) is an adaptive digital learning environment that consists of several serious games, targeting different ESL subskills. Within each game, students completed assignments that were adapted to the performance level of the student following Elo chess ratings (Elo, 1978). This way, students were always provided with the right amount of challenge. Both the intervention and control group completed measures for decoding, morphological and syntactic awareness, inference making as a result of argument overlap, anomaly detection and reading comprehension (reading measures) at the beginning (T1) and end (T2) of the school year. For the intervention group to gain insight into gaming behavior, the number of completed assignments and the students' performance level for each game were recorded as well. The gaming measures were included as mediators between the reading measures at T1 and reading measures between T2 for the intervention group.

In the present study, we addressed the following questions:

- 1) What are the effects of a game-based adaptive WTI intervention on ESL reading measures, namely: decoding, morphological and syntactic awareness, WTI ability, and reading comprehension development?
- 2) To what extent are individual differences in T1 levels of ESL reading measures, namely: decoding, morphological and syntactic awareness, WTI ability, and reading comprehension predictive of performance within the intervention and in turn for reading measures at T2?

Our first hypothesis was that participants in the intervention group

would improve more on the reading measures than those in the control group, because WTI is not as such addressed in the regular ESL curriculum. Therefore, we especially expected larger growth in WTI ability for the intervention group than the control group. We also expected a larger growth in reading measures in the intervention group than in the control group, as the intervention provided students with specific and adaptive WTI-based assignments tailored to students' performance level as well as direct and tailored feedback to foster ESL reading.

The other hypotheses concern the intervention group only. Our second hypothesis was that students who completed more assignments within the intervention showed more improvement between T1 and T2 on reading measures performance. The third hypothesis was that students who had higher levels of reading measures at T1, showed more growth in game performance, and in turn had higher levels of the reading measures at T2. More specifically, we expected that students who were better at decoding, morphological and syntactic awareness, WTI, and reading comprehension at T1 would show most performance growth the morphosyntactic awareness game and translation game, and in turn better scores on the aforementioned reading measures at T2. Furthermore, we expected that students with high decoding, WTI, and reading comprehension scores at T1 would show more performance growth within the idiom game, and in turn higher scores at T2. Finally, we expected that students with high decoding at T1 would benefit most from the dictation game and in turn have higher decoding levels at T2.

### Method

#### 1.4. Design

The present study was part of a larger study<sup>1</sup>. In 2016–2017, we tracked the English as a second language (ESL) development of a control group of 532 seventh grade students. In 2017–2018, we selected a new cohort of seventh grade students. We matched the intervention group with the control group based on school track and English vocabulary knowledge (see *Participants*, below). The intervention took place in 2017–2018. An overview of the design is displayed in [Table 1](#).

While the control group followed (ESL) lessons as usual, the intervention groups followed a 12-week intervention program with *Words&Birds* (Words&Birds, Oefenweb, 2015), which replaced one English lesson during the week. Both the intervention and control group received 135–180 min of ESL classroom instruction divided across three lessons a week.

In the intervention group, 50 min of the classroom instruction were replaced with the intervention: 40 min in the first lesson a week, to complete the assignments within the games, and 10 min during a second lesson in the week in which plenary feedback was provided by the

**Table 1**  
Overview of the Study Design.

Time	Activity
November 2016 (Pretest – T1)	Pretest control group
April 2017 (Posttest – T2)	Posttest control group
November 2017 (Pretest – T2)	Pretest intervention group
January – March 2018	12-week <i>Words&amp;Birds</i> intervention
April 2018 (Posttest – T2)	Posttest intervention group

<sup>1</sup> The present study was part of a project, in which several studies were performed simultaneously. The goal of the present study was to investigate the effect of an intervention focused on improving WTI. Simultaneously, data were collected within a different group of children that followed an orthographic learning intervention, which was irrelevant for the present study. The children included in the present study only followed the WTI intervention. While data for both intervention studies were collected parallel, the studies included different children.

teacher. The remainder of the second lesson as well as the complete third lesson were devoted to ESL classroom instruction in the intervention group. Thus, the control group had 135–180 min of ESL classroom instruction, and the intervention group had 50 min of the Words&Birds intervention, and 85–130 min of ESL classroom instruction. The total intervention program was 600 min. An overview of all variables and the way scores were obtained is presented in Table 2.

### 1.5. Participants

Participants of this study were 330 seventh grade students in the Netherlands, who completed all measures in our study, divided into a control group ( $n = 166$ , 52% boys) and an intervention group ( $n = 164$ , 56% boys). The participants in both groups were between 11 and 13 years old (control group:  $M = 12.51$ ,  $SD = 0.41$ ; intervention group:  $M = 12.61$ ,  $SD = 0.47$ ). The participants from the intervention group were matched to participants in the control group based on educational track and vocabulary level. The intervention group consisted of two groups of students in prevocational education, two in higher vocational and pre-university education, and two in preuniversity education. Groups in the control condition were matched with the groups in the intervention condition based on educational track and average score on the Peabody Picture Vocabulary Test.

These students had received approximately 24 months of ESL education in primary school. Although ESL is obligatory in the Netherlands from the fifth grade on, there are substantial differences in the amount of time that is spent on ESL education, with individual differences between students as a result (Thijs et al., 2011). Their formal English education had commenced at the start of seventh grade, which was three months before the beginning of this study. By the end of the school year, all students had received eight months of English education in the secondary school setting.

### 1.6. Word-to-text integration (WTI) intervention

The intervention group completed a WTI intervention, which replaced one of the three English lessons during the week. The intervention entailed completing WTI-based assignments within four serious games using *Words&Birds* (Oefenweb, 2015) during the first lesson in the week and plenary feedback during the second lesson in the week. During the first lesson, students were to complete 60 assignments within each of four games within *Words&Birds* (Oefenweb, 2015) on a weekly basis, rendering a total of 240 assignments a week. Assignments that were not completed within the lesson became homework. During a second lesson later on in the week, the teacher discussed common errors that were made during the assignments (as analyzed by the researchers) in a 10-minute plenary session. Prior to the intervention, teachers were trained and received a manual with instructions. Furthermore, a researcher was present during the first intervention lesson in each class. The researchers contacted the teachers once a week to ensure treatment fidelity, and discuss progress and commonly made errors.

*Words&Birds* is an adaptive practice and monitoring system in which students are presented with items that match their skill levels, similar to computerized adaptive testing in Item Response Theory (van der Linden & Glas, 2000). *Words&Birds* applies an adaptation of the Elo (1978) rating system designed for chess competitions to adjust the user and item ratings after each trial (for more details, see (Klinkenberg et al., 2011)). This way, the algorithm can estimate student skill and item difficulty simultaneously, resulting in a game score for each of the four games, which researchers, students, and teachers could view. Items are selected based on the probability of a correct answer on the item by the student. By default, the system selects items with a probability of 0.75 of answering an item correctly. Students can manually adjust the difficulty to easy (0.90 probability of a correct answer) or hard (0.60 probability of a correct answer). As items are selected based on item and student characteristics, students did not necessarily complete the same items.

**Table 2**  
Overview of Variables, Definitions, and Measures

Variable	Definition	Measure
Reading comprehension	The ability to understand the text, which is often theorized as a product of identifying words during reading (decoding) and understanding language.	Accuracy on comprehension questions presented with texts (College voor Toetsen en Examens – Board for Assessment and Exams, 2016)
Decoding	Combining letters to the corresponding sounds and blending the letter-sound combinations to read the entire word.	Accuracy on TOWRE (Torgesen et al., 2012)
Morphological awareness	Awareness of morphemes: the smallest meaningful units in words. For example: in the word <i>violinist</i> , the morpheme <i>-ist</i> indicates that a violinist is a person	Accuracy on Singson Task (Singson et al., 2000)
Syntactic awareness	Awareness of the grammar rules and rules about the structure of sentences; for example, that if a subject is plural, the corresponding verb also needs to be plural.	Accuracy on Gap Text (Siegel, 2008)
Anomaly speed (WTI)	Detecting anomalies is a part of the ability to integrate words into a mental representation of text, i.e. word-to-text integration (WTI). Here, we mean the speed with which anomalous, compared to continuous passages are read.	Logged reading times of anomaly detection passages during a self-paced reading task (Mulder et al., 2020)
Anomaly processing (WTI)	Sensitivity of readers to encountering anomalies: when an anomaly is encountered, the reader can continue reading at the same pace, or slow down in order to try and fit the anomaly into a mental representation of the text.	Random effect of sensitivity to anomaly detection passages during self-paced reading task, extracted from a multilevel model (Mulder et al., 2020)
Argument overlap speed (WTI)	Inferencing as a result of argument overlap is part of the ability to integrate words into a mental representation of the text. Here, we mean the speed with which passages with implicit inferences, compared to passages with explicit repetitions are read.	Logged reading times of argument overlap passages during a self-paced reading task (Mulder et al., 2020)
Argument processing (WTI)	Sensitivity of readers to encountering implicit inferences: when an implicit inference is required to be made to understand the meaning of a word, the reader can continue reading at the same pace, or slow down in order to try and uncover the meaning of a word and integrate it into a mental representation of the text.	Random effect of sensitivity to complexity to argument overlap passages during self-paced reading task, extracted from a multilevel model (Mulder et al., 2020)
Completed assignments within intervention Performance		Number of completed assignments within intervention within each of the four games Accuracy and speed on assignment within each of the four games

In *Words&Birds* trials, students are first presented with the main screen with four flying birds, each representing one of the games. When students click on a bird, the game starts and they are presented with ten subsequent assignments. For each item, a question is presented with an answer box or multiple-choice options to provide the answer. When students do not know the answer, they either guess an answer or click on a question mark. Students need to answer each question within 20 s; a count-down bar is presented at the bottom of the screen (see Figs. 1). When they respond, the correct answer is immediately displayed. After that, they are presented with the next assignment. Upon finishing ten assignments, the game award is presented and students return to the main screen, where they can click the same or another bird.

Gaming elements within *Words&Birds* are that students can collect coins to buy virtual awards. Furthermore, they are rewarded for both accurate and fast responses: Faster responses are rewarded with more coins. These gaming elements stimulate students to successfully complete the assignments. While a time pressure or reward system may cause anxiety, it must be noted that the algorithm always provides students with assignments that they have a 75% chance of completing correctly. If students cannot complete an assignment within the time limit, they will be provided with an easier assignment next, thus ensuring they are always working in their zone of proximal development and will not be presented with items they cannot finish within the time limit.

During the intervention, students played the four different games ‘Shaper’ (Morphosyntactic operations), ‘Wordo’ (Translation in Sentence), ‘Twinny’ (Idiom Recognition), and a filler game ‘Ducktator’ (Dictation). The first three games were included to enhance aspects of WTI, the filler game was included to ensure students would continue to be motivated to complete 60 assignments for each game during a time span of 12 weeks. Screenshots of all games, in which an example of one assignment is shown, are provided in Figs. 1.

**Morphosyntactic awareness.** In ‘Shaper’, students performed assignments focused on training morphosyntactic awareness: 1) superlative completion, 2) plural formation, and 3) morpheme addition. Students received feedback containing the correct answer. An example of each assignment is displayed below:

1. Question: ‘..., holier, holiest’

Options: open ended

Feedback: ‘*holy, holier, holiest*’

2. Question: ‘One pigtail. Two ...’

Options: open ended

Feedback: ‘*One pigtail. Two pigtails*’

3. Question: ‘Marco Polo made an unexpected discover...’

Options: ‘*\*discoveree, discoverer, discovery, \*discoverness*’

Feedback: ‘*Marco Polo made an unexpected discovery*’

The game contained a total of 2307 items.

**Translation in sentence.** In ‘Wordo’, students were presented with sentences in which one word was highlighted. Students were to translate the highlighted word from English into Dutch, choosing from four multiple-choice options. Students received feedback containing the Dutch translations of the full sentence that was presented as the question. For example:

Question: ‘*I see one goose and six ducks in the pond.*’

Options: ‘*onderdorpeling (dip), dak (roof), eend (duck), rok (skirt), lende (loin), haag (hedge)*’

Feedback: ‘*Ik zie een gans en zes eenden in de vijver.*’

The game contained a total of 1734 items.

**Idiom recognition.** In ‘Twinny’, students were presented with different items about idioms: They either had to complete an idiom of which one word was omitted, or choose, out of four options, which sequence of words was in the right order. An example is displayed below:

Question: ‘*a ... of stars*’

Options: ‘*forest, fleet, galaxy, bouquet*’

Feedback: ‘*a galaxy of stars*’

The game contained a total of 651 items.

**Dictation.** The game ‘Ducktator’ was added as a filler task to provide students with sufficient variation and thus motivation. It is a dictation game, in which they heard an English word and had to type the correct spelling, for example: ‘*thirsty*’. The game contained a total of 1499 items.

## 2. Materials

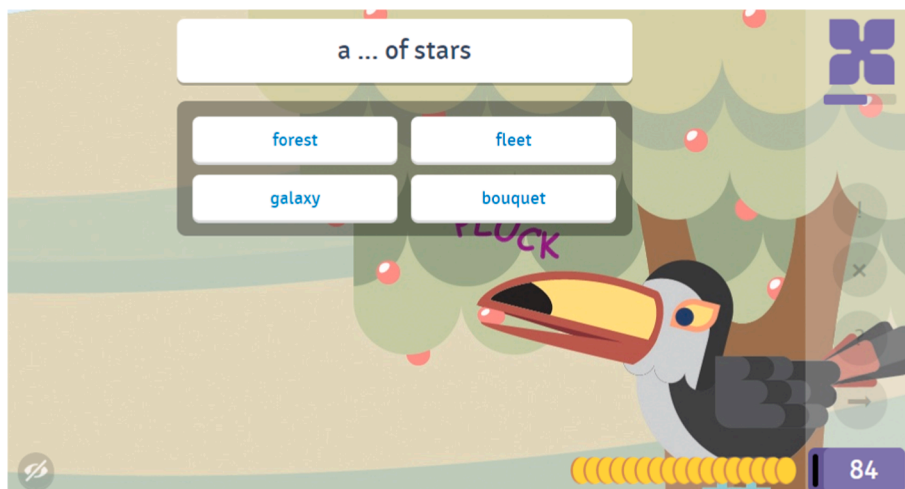
**Reading comprehension.** To measure reading comprehension, items from a nationally standardized reading comprehension test were selected. The test was normed on final-year students in prevocational education ([College voor Toetsen en Examens – Board for Assessment and Exams, 2016](#)). Students were presented with three different texts and 12 questions, of which 10 were multiple-choice questions, and two were open-ended questions. Each correct answer was granted 1 point, and the total amount of correct points formed the reading comprehension score with a maximum of 12. Reliability of the reading comprehension scores was acceptable (based on [Kline, 1993](#)) at both Time points ( $\alpha = 0.60$  at



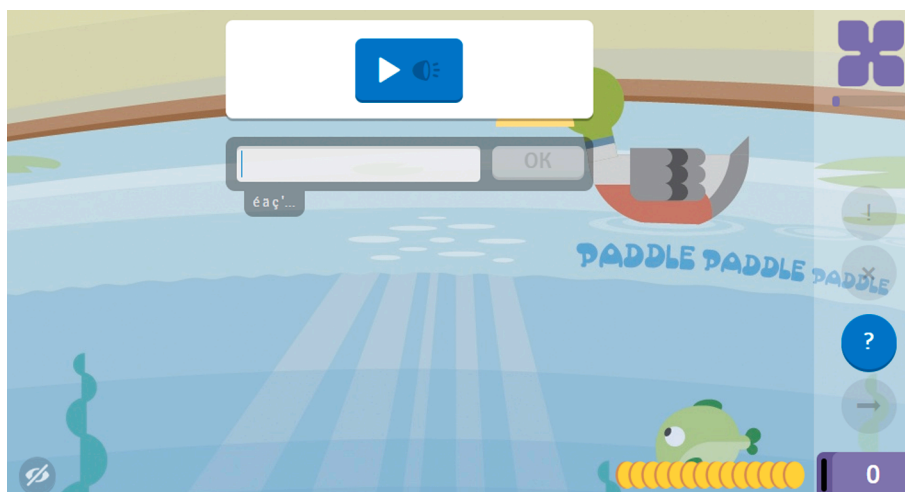
Fig. 1a. Screenshot of an Assignment within the Morphosyntactic Awareness Game, ‘Shaper’



**Fig. 1b.** Screenshot of an Assignment within the Translation of Words within Sentences Game, 'Wordo'. Note. The Dutch words translate into English as follows: onderdompeling (dip), dak (roof), eend (duck), rok (skirt), lende (loin), haag (hedge).



**Fig. 1c.** Screenshot of an Assignment within the Idiom Recognition Game, 'Twinky'



**Fig. 1d.** Screenshot of an Assignment within the Dictation Game, 'Ducktator'. Note. Upon clicking the play button, students hear a word and have to type the word in the field below.

T1,  $\alpha = 0.68$  at T2) and could not be further improved.

**Decoding.** The Sight Word Efficiency subtest of the Test of Word Reading Efficiency second edition (TOWRE; Torgesen et al., 2012) was used to measure word decoding. A list of 109 regular and irregular English words (such as ‘go’ and ‘embassy’) with increasing difficulty was presented to the students. Students were told to read the words on the list as quickly and accurately as possible within 45 s. The decoding score consisted of the number of words read aloud correctly. Reliability of the decoding scores was excellent (based on Kline, 1993) at both Time points ( $\alpha = 0.94$  at T1 and  $\alpha = 0.96$  at T2).

**Morphological awareness.** Following Siegel (2008), we used the first two subtests of the Singson Tasks (Singson et al., 2000) to measure morphological awareness. In the first subtest, students had to complete words of which one morpheme was omitted. They were presented with ten sentences that each contained an incomplete word and had to choose the right morpheme to complete the word from four answer options, for example: ‘She hoped to make a good \_\_\_\_ A. impressive, B. impressionable, C. impression, D. impressively.’ The second subtest was similar, but the word that needed to be completed was a pseudoword, for example: ‘I admire her \_\_\_\_ A. sufilive, B. sufilify, C. sufflation, D. sufilize’. The maximum score was 20. Reliability was good at T1 and excellent at T2 ( $\alpha = 0.77$  at T1,  $\alpha = 0.83$  at T2; Kline, 1993).

**Syntactic awareness.** A gap text was used to measure syntactic awareness (Siegel, 2008). Students were presented with 20 sentences in which one word was omitted and were instructed to fill out a semantically, grammatically, and orthographically correct word. For example, in the sentence: ‘The \_\_\_\_\_ put his dairy cows in the barn.’, ‘farmer’ is semantically, syntactically, and orthographically correct. Students could obtain a maximum score of 20. Reliability was excellent at both Time points ( $\alpha = 0.81$  at T1,  $\alpha = 0.84$  at T2; Kline, 1993).

**WTI.** A computerized word-by-word self-paced reading task, with a moving window paradigm, was created in Inquisit 4 (2015) to measure WTI (Mulder et al., 2020). In the self-paced reading task, students read sentence passages with three different manipulations: anaphora resolution, argument overlap, and anomaly detection. WTI was proposed to be reflected by the effect of complexity on reading times on the target word, target plus 1, and target plus 2 (similar to Bultena et al., 2015), for each text manipulation (Mulder et al., 2020). Therefore, we created a complex and simple version of each sentence passage, resulting in twelve pairs of complex and simple sentence passages per text manipulation. Simple and complex sentence passages were identical, with the

exception that each complex sentence passage contained a complex target word, whereas each simple sentence passage contained a simple target word. A graphic overview of the design of the task, including example passages, is provided in Fig. 2. As the effect of the complexity manipulation was absent in the anaphora resolution manipulation, these passages were excluded from the present study.

Argument overlap passages were adapted from a study by Yang and colleagues (2007) and always consisted of two sentences. The complex passages contained a semantic construction that required readers to make an implicit inference, whereas simple passages contained an explicit repetition of a word earlier in the passage. On average, the passages contained fifteen words ( $M = 15.83$ ,  $SD = 2.37$ ; range = 12–19). In each passage, the target was placed at the beginning of the second sentence and was between the eighth and seventeenth position ( $M = 11.38$ ,  $SD = 2.66$ ).

The anomaly detection passages were constructed for the purpose of the present study. Each complex anomaly detection passage contained an anomaly, whereas simple anomaly detection sentences did not (Mulder et al., 2020). Average passage length was  $M = 10.08$ ,  $SD = 1.89$  (range = 7–14). The target word was placed between the fourth and the tenth position of the sentence ( $M = 7.03$ ,  $SD = 2.37$ ).

At each time point (T1 or T2), students read six complex and six simple versions of 12 passages per text manipulation, counterbalanced across time. For example, if they read the complex versions of items 1–6 at T1, they read the simple versions of these items at T2. Simple and complex passages were pseudorandomized and divided into 30 different matched lists, to control for order effects. We distinguished between WTI speed and WTI processing for each text manipulation. WTI speed was the average logged reading time, within a text manipulation on the target, target plus 1, and target plus 2, regardless of complexity. WTI processing was the beta coefficient for complexity (simple versus complex) extracted from a mixed-effects model predicting logged reading times on the target, target plus 1 and target plus 2 (see the ‘Analyses’ section for more details). Reliability can be considered acceptable (Kline, 1993):  $\alpha = 0.79$  and  $\alpha = 0.72$  for argument overlap and anomaly detection passages respectively.

We instructed students to read the passages silently and carefully at a normal pace. Students were told they merely had to comprehend the passages, and not try to memorize them. Students were presented with one practice trial of each text manipulation. Upon completion of the practice trials, students were allowed to ask questions. Half-way through

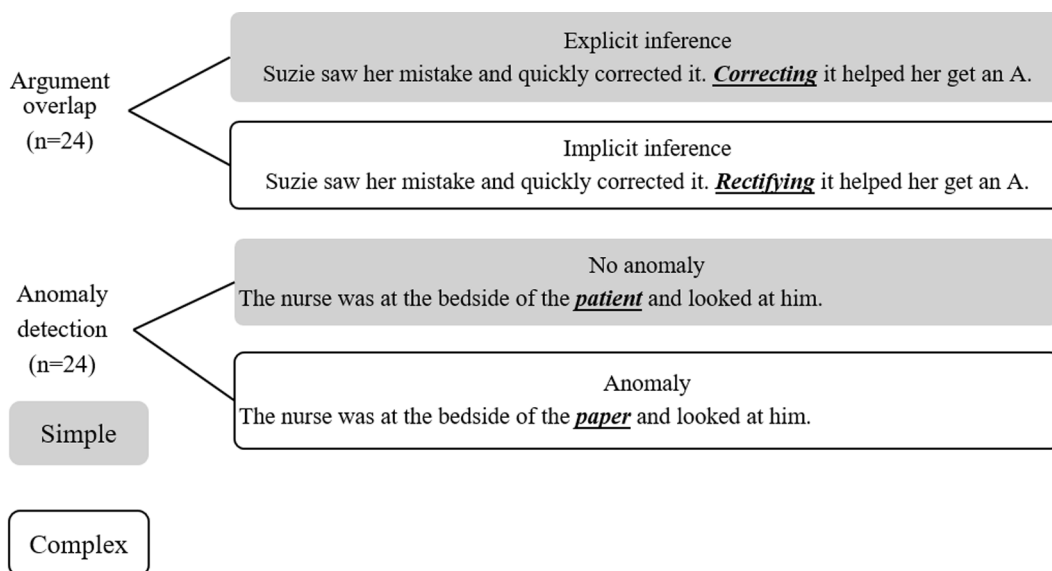


Fig. 2. Graphic Overview of Different WTI Text Manipulations and their Corresponding Complex and Simple Sentence Passage. Targets are Underlined, Printed in Bold and in Italics.

the task, they received a one-minute break. In each trial, students were presented with a screen that had a dash representing each word of the passage. The passage appeared one word at a time, and students had to press the space bar when they had read the word. Subsequently, the current word would disappear and the next word would appear.

**Completed assignments.** The number of completed assignments within each game was recorded, to examine whether students who completed more assignments, showed more improvement on reading measures between T1 and T2.

**Game performance growth.** To assess performance of intervention group students within each game, for each game when students started playing, their initial skill level was determined based on their performance on the first 40 assignments they completed. Their performance at the end of the intervention was also measured and the difference between initial and final performance was included as the variable performance growth.

### 2.1. Procedure

Pretest (T1) and posttest (T2) measures were conducted in November and April respectively. All participants were submitted to two 50 min plenary classroom sessions and one 45 min individual session, in which reading measures were administered. Morphological and syntactic awareness, and reading comprehension were assessed in a plenary classroom setting. Decoding and WTI were measured during the individual session. Students were seated approximately 30 cm away from the computer screen during the online WTI task, in which words were presented in Consolas font. The other tasks were paper and pencil tasks.

### 3. Analyses

We performed several preparatory analyses. First, we extracted the online processing measures for argument overlap and anomaly detection in order to estimate the effect of complexity on logged reading times. We did so by creating separate mixed-effects models for each text type (argument overlap, and anomaly detection) and for each time (T1 and T2) in R version 3.6.3. (R Core team, 2020) with lme4 (Bates et al., 2015), resulting in four models: argument overlap T1, argument overlap T2, anomaly detection T1, and anomaly detection T2. The mixed-effects models had logged reading time as the dependent variable, and word position (target, target plus 1, target plus 2), frequency (logged frequency of the word), complexity (simple or complex), reading speed (average reading speed on words preceding the target words) as fixed effects, participant and item as random intercepts, and complexity as random slopes for both participant and item. Reading times for which the standardized residuals were larger than 2.5 or smaller than -2.5 were removed from the mixed-effects model (Baayen, 2008). After fitting the models for each text type and time point, we extracted the regression coefficient of complexity, using the 'coef' function. Thus, this coefficient reflects the mean effect of complexity (across participants) on logged reading times and the participant-specific deviation from this mean. This resulted in separate processing values for argument overlap and anomaly detection at T1 and T2. WTI speed, logged reading times on the target words regardless of complexity, was also included in subsequent analyses.

To examine whether the intervention and control group differed at T1, we performed independent samples t-tests and used the Holm-Bonferroni procedure in order to adjust *p*-values and control for Type I error (Abdi, 2010). Subsequently, we performed a Multivariate Analysis of Variance (MANOVA) to examine the overall intervention effect. Correlations above 0.90 between the dependent variables were considered problematic (Tabachnick & Fidell, 2012), but this was never the case in the data. In order to analyze the effect of performance growth on games within the intervention, on reading measures development, parallel mediation models were fitted with multiple mediation analysis with the PROCESS plug-in (Hayes, 2018) in SPSS 26 (IBM Corp, 2019).

Bootstrapping was set to 5000 cycles (Preacher & Hayes, 2004). Total effects are combinations of both the direct and indirect effects of independent variables on the dependent variable. Multicollinearity was checked for all mediation models. VIF values larger than greater than 10 are considered problematic (Myers, 1990), but none of the models displayed problematic VIF values.

### 4. Results

Descriptive statistics of all reading measures can be found in Table 3. Descriptive statistics of game performance and number of completed assignments can be found in Table 4. Correlations between all reading measures, completed assignments, and game performance growth can be found in Appendix A and B. Although the intervention group had lower scores on some pretest (T1) measures, results from the independent samples t-tests, corrected for Type I errors, demonstrated no significant differences between the intervention and control group on the reading measures at T1.

#### 4.1. Development of reading measures and development within intervention

To examine our first hypothesis about the effects of intervention, we first performed a repeated measures MANOVA with time (T1, T2) \* group (intervention, control) \* reading measures (decoding, morphological, and syntactic awareness, argument processing, anomaly processing, argument reading speed, anomaly reading speed, reading comprehension). Intervention effects can be evidenced by an interaction between time and group. The three-way interaction between time, group, and reading measures was not significant,  $F(7, 213) = 1.92, p = .068$ , partial  $\eta^2 = 0.06$ . However, an effect of the intervention was evidenced by a significant two-way interaction between time\*group,  $F(1, 219) = 6.85, p = .009$ , partial  $\eta^2 = 0.03$ . Furthermore, there were significant two-way interactions between reading measures\*group,  $F(7, 213) = 3.42, p = .002$ , partial  $\eta^2 = 0.10$ , and time\*reading measures,  $F(7, 213) = 42.76, p < .001$ , partial  $\eta^2 = 0.58$ . To disentangle these interactions from the intervention effect, we performed follow-up ANOVAs for each reading measure, in which we examined the effect of time, with group as the between-subjects factor. There was a significant interaction between time and group for decoding,  $F(1, 302) = 9.56, p = .002$ , partial  $\eta^2 = 0.03$  (see Fig. 3a) and between time and group for anomaly processing,  $F(1, 252) = 8.07, p = .005$ , partial  $\eta^2 = 0.03$  (see Fig. 3b). There was a significant difference on decoding between pretest and posttest for both the intervention  $t(159) = 12.80, p < .001, d = 0.552$ ; and the control group  $t(143) = 6.10, p < .001, d = 0.345$ . There was a significant difference on anomaly processing between pretest and posttest for the intervention  $t(109) = 2.85, p = .005, d = 0.380$ . The control group did not improve between pretest and posttest  $t(143) = 0.504, p = .615, d = -0.061$ . The absence of an interaction between time and group for the other reading measures indicates that development of the intervention and control group was similar on these other measures.

#### 4.2. Effect of completed assignments on reading measures development

Our second hypothesis concerned the effect of completed assignments on the effectiveness of the intervention. To test this hypothesis, we performed repeated measures ANOVAs for each reading measure, with number of completed assignments during the intervention as a between-subjects factor. There were significant interactions between time and completed idiom recognition assignments for argument reading speed,  $F(94, 15) = 2.60, p = .020$ , partial  $\eta^2 = 0.94$ , and for reading comprehension,  $F(99, 17) = 2.21, p = .033$ , partial  $\eta^2 = 0.93$ . These outcomes indicate that students who completed more idiom recognition assignments also showed more growth in argument reading speed and reading comprehension than those who completed fewer



**Table 3**  
Descriptive Statistics of the Reading Measures at Pretest (T1) and Posttest (T2) per Group and Cohen's d

	Control				Intervention				Cohen's d		
	T1		T2		T1		T2		Control - Intervention		T1 - T2
	M	SD	M	SD	M	SD	M	SD	T1	T2	Intervention
1. Decoding	60.194	10.435	63.556	9.030	57.445	10.743	63.025	9.430	-0.260	-0.058	0.552
2. Morphological awareness	10.222	4.111	11.528	4.028	9.713	3.759	11.571	4.349	-0.129	0.010	0.457
3. Syntactic awareness	7.076	3.954	8.611	4.258	6.433	4.002	8.239	4.449	-0.162	-0.085	0.427
4. Argument processing (WTI)	0.112	0.089	0.127	0.101	0.107	0.095	0.113	0.092	-0.054	-0.145	0.064
5. Argument reading speed (WTI)	6.373	0.257	6.262	0.249	6.388	0.241	6.310	0.241	0.060	0.196	-0.324
6. Anomaly processing (WTI)	0.054	0.066	0.050	0.066	0.026	0.105	0.064	0.095	-0.320	0.171	0.380
7. Anomaly reading speed (WTI)	6.346	0.257	6.234	0.252	6.391	0.260	6.321	0.254	0.174	0.344	-0.272
8. Reading comprehension	5.069	2.402	6.118	2.587	4.472	2.393	5.414	2.669	-0.249	-0.268	0.372

Note. Cohen's d displays the difference between 1) the control group and intervention group on T1, 2) control group and intervention group on T2, and 3) between T1 and T2 for the intervention group.

**Table 4**  
Means (Standard Deviations) of Game Performance, Game Performance Growth, Cohen's d, and Number of Completed Assignments

	Morphosyntactic awareness	Translation in sentence	Idiom recognition	Dictation
Game performance at beginning	5.63 (1.69)	5.69 (1.99)	1.56 (2.62)	5.49 (1.95)
Game performance at end	7.37 (1.68)	8.42 (1.35)	5.32 (1.62)	6.66 (1.88)
Game performance growth	1.76 (1.63)	2.73 (2.63)	3.76 (2.03)	1.24 (1.84)
Cohen's d	1.03	1.61	1.73	0.61
Completed assignments	492.34 (175.30)	454.66 (159.30)	469.70 (155.12)	427.78 (173.22)

Note. Game performance growth refers to the difference between game performance at the beginning and at the end of the intervention; Cohen's d displays the difference between the beginning and end of the intervention

idiom recognition assignments. There were no significant interactions between time and completed assignments for the other reading measures and games.

**4.3. Mediation effects of performance growth on reading measures development**

To address our third hypothesis, we examined the effect of game performance growth on intervention effects more closely, and created mediation models for each reading measure. The independent variable

was always the initial level, i.e., the score on a reading measure at T1; the dependent variable was always the same measure at T2, and for each of the four games, performance growth was included as a mediator. An overview of all mediation effects can be found in Table 5. For the sake of parsimony, only significant effects are reported in the text. Except for argument processing and anomaly processing, there were significant autoregressive effects of each reading measure, with higher T1 scores resulting in higher scores at T2. We first report the effects of T1 reading measures on performance growth within games and then discuss the effect of performance growth within games on T2 reading measures.

**Effects of T1 reading measures on performance growth within games.** All initial levels (T1) of the reading measures were related to performance growth of at least one game, except for argument processing. In more detail, students with better decoding skills at T1 also showed more performance growth on the morphosyntactic awareness and translation games. Furthermore, on the one hand, students with better T1 levels of morphological awareness displayed more performance growth on the morphosyntactic awareness and translation games. On the other hand, students with poorer T1 levels of morphological awareness displayed more performance growth on the idiom recognition game than did students with better T1 levels of morphological awareness. T1 levels of syntactic awareness were positively related to performance growth on the morphosyntactic awareness and translation games, but negatively related to performance growth on the idiom recognition game. Students who initially read argument overlap passages more quickly showed more performance growth within dictation than those who read these more slowly. Students who initially read anomaly detection passages more quickly displayed more performance growth within dictation than those who read these more slowly. A larger effect of sensitivity to anomalies (i.e., higher anomaly processing) at T1

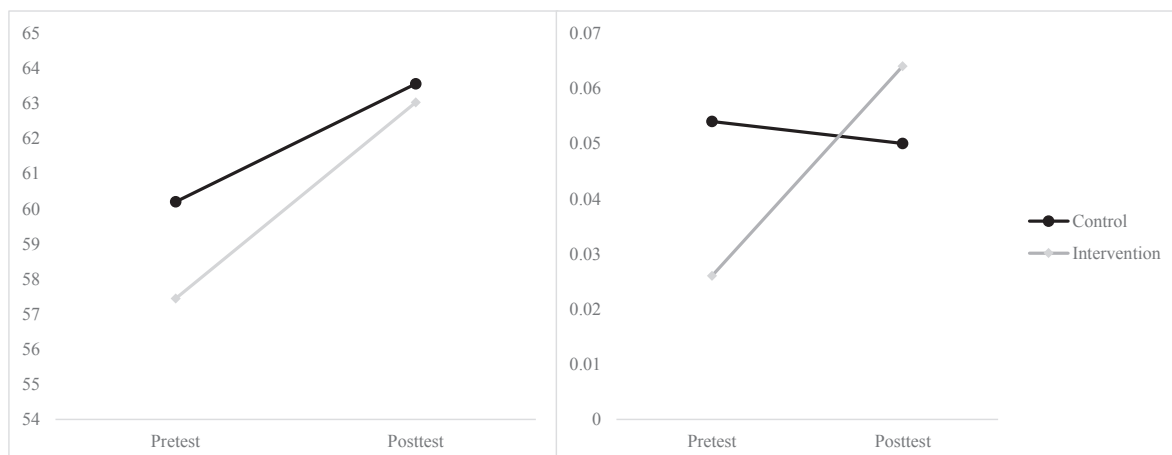


Fig. 3. Outcomes at Pretest and Posttest (x-axis) per group on Decoding (Left-hand Panel) and Anomaly Processing (Right-hand Panel).

**Table 5**  
 Results of Mediation Analyses with Reading Measures at T1 (1) as the Independent Variable, Performance Growth on: Morphosyntactic Awareness, Translation, Idiom Recognition, and Dictation as Mediators (M), and Reading Measures at T2 (2) as the Dependent Variables

Path	Mediator (M)																			
	Morphosyntactic Awareness					Translation					Idiom Recognition					Dictation				
	B	SE	T	p	R <sup>2</sup>	B	SE	t	P	R <sup>2</sup>	B	SE	T	p	R <sup>2</sup>	B	SE	t	p	R <sup>2</sup>
De1 → M	0.021 <sup>†</sup>	0.012	1.789	0.076	0.020	0.044*	0.019	2.286	0.024	0.032	-0.012	0.015	-0.821	0.413	0.004	0.016	0.013	1.180	0.240	0.094
M → De2	0.333	0.255	1.307	0.193		0.090	0.159	0.568	0.571		-0.010	0.194	-0.052	0.959		0.035	0.211	0.165	0.869	
De1 → De2																0.746*	0.036	20.505	<0.001	
Model De2																				0.750*
Mo1 → M	0.069 <sup>†</sup>	0.035	1.991	0.048	0.026	0.138*	0.057	2.427	0.016	0.039	-0.107*	0.045	-2.397	0.018	0.038	0.019	0.042	0.457	0.648	0.001
M → Mo2	0.215	0.177	1.212	0.228		0.121	0.107	1.130	0.260		-0.002	0.128	-0.0117	0.991		-0.268 <sup>†</sup>	0.1373	-1.952	0.053	
Mo1 → Mo2																0.800*	0.070	11.428	<0.001	
Model Mo2																				0.726*
Sy1 → M	0.112*	.31	3.584	<0.001	0.078	0.167*	0.051	3.297	0.001	0.064	-0.114*	0.039	-2.915	0.004		-0.020	0.037	-0.546	0.586	
M → Sy2	-0.195	0.168	-1.157	0.249		0.108	0.102	1.057	0.292		-0.093	0.128	-0.724	0.470		0.089	0.136	0.655	0.514	
Sy1 → Sy2																0.802*	0.066	12.074	<0.001	
Model Sy2																				0.534*
ArS1 → M	-0.410	0.599	-0.684	0.495	0.004	0.652	1.088	0.599	0.550	0.003	0.538	0.828	0.650	0.517	0.063	1.547*	0.715	2.164	0.033	0.042
M → ArS2	0.004	0.013	0.294	0.768		0.004	0.008	0.464	0.644		-0.013	0.010	-1.379	0.171		0.018	0.011	1.635	0.105	
ArS1 → ArS2																0.572*	0.080	7.148	<0.001	
Model ArS2																				0.379*
ArP1 → M	-0.992	1.517	-0.654	0.515	0.004	1.810	2.751	0.657	0.513	0.004	3.004	2.078	1.445	0.151	0.019	0.631	1.846	0.342	0.733	0.001
M → ArP2	-0.003	0.006	-0.487	0.627		0.007*	0.004	2.044	0.044		-0.001	0.005	-0.082	0.935		-0.001	0.005	-0.191	0.849	
ArP1 → ArP2																0.098	0.094	1.038	0.302	
Model ArP2																				0.058
AnS1 → M	-0.197	0.556	-0.354	0.724	0.001	0.877	1.006	0.872	0.385	0.007	0.551	0.766	0.719	0.474	0.005	1.117 <sup>†</sup>	0.668	1.673	0.097	0.025
M → AnS2	0.010	0.013	0.730	0.467		-0.006	0.007	-0.805	0.423		-0.011	0.010	-1.1887	0.237		0.016	0.011	1.510	0.134	
AnS1 → AnS2																0.643*	0.072	8.885	<0.001	
Model AnS2																				0.459*
AnP1 → M	4.622*	1.300	3.557	<0.001	0.105	2.890	2.476	1.167	0.246		0.737	1.895	0.389	0.698	0.001	-2.020	1.658	-1.218	0.226	0.014
M → AnP2	-0.005	0.007	-0.726	0.469		0.009*	0.004	2.610	0.010		0.002	0.005	0.455	0.650		0.004	0.005	0.792	0.430	
AnP1 → AnP2																0.025	0.092	0.268	0.789	
Model AnP2																				0.066
RC1 → M	0.078	0.065	1.204	0.231	0.013	0.103	0.103	0.998	0.321	0.009	-0.233*	0.074	-3.132	0.002	0.079	-0.027	0.070	-0.382	0.704	0.001
M → RC2	0.150	0.139	1.075	0.285		0.090	0.088	1.025	0.308		0.132	0.112	1.178	0.242		0.087	0.116	0.750	0.455	
RC1 → RC2																0.562*	0.089	6.323	<0.001	
Model RC2																				0.304*

Note. † significant at  $p < .10$ , \* significant at  $p < .05$ ; De = decoding, Mo = morphological awareness, Sy = syntactic awareness, ArS = argument reading speed, ArP = argument processing, AnS = anomaly speed, AnP = anomaly processing, RC = Reading Comprehension

was positively related to performance growth within the morphosyntactic awareness game. Finally, students with poorer T1 levels of reading comprehension showed more performance growth in idiom recognition than those with better reading comprehension.

#### **Effects of performance growth within games on reading levels at T2.**

Performance growth in the translation and idiom recognition games were related to reading measures at T2. More specifically, students who displayed more performance growth on the idiom recognition game also showed higher argument reading speed at T2. Performance growth within the translation game was positively related to a larger effect of complexity on students' reading times for anomaly detection passages at T2. There were no significant effects of performance growth in the morphosyntactic awareness and the dictation games on reading measures at T2.

## **5. Discussion**

The aim of the present study was to examine the effects of an adaptive game-based word-to-text integration (WTI) intervention on reading comprehension and its predictors in English as a second language (ESL). An intervention group was matched with a control group based on educational track and vocabulary levels.

The main findings were that the intervention and control group had similar scores at T2 on all reading measures, but that the intervention group improved more on decoding and anomaly processing than did the control group. The T1 levels of all reading measures were related to performance growth within the intervention games; however, only performance game growth on the translation and idiom game were related to reading measures at T2. The results are discussed in more detail below.

With regard to our first hypothesis, the results indicated that the intervention group improved more on decoding and anomaly processing than the control group. The improvement on decoding is in line with previous studies examining effects of a reading intervention on decoding ability (e.g., [Nayak & Sylva, 2013](#); [Saine et al., 2011](#)). With respect to WTI, our findings are the first to show an effect of a WTI intervention on WTI development, namely on anomaly processing. We did not find intervention effects on any of the other reading measures, as growth was attested in both the teaching as usual and the adaptive game intervention groups. Inference making (as a result of argument overlap), i.e., argument processing, was not specifically trained in the intervention. Rather, during the intervention students had to perform morphosyntactic operations on words within a sentence and translate words within a sentence. The latter means that students had to activate the semantic representations of primes in the sentence preceding an unfamiliar target word, which had to be translated. This may explain why an effect was found on anomaly processing, which also requires students to activate semantic properties of sentences ([van Berkum et al., 1999](#)), but no effect was found for argument processing, which was not specifically trained as such during the intervention.

Regarding our second hypothesis, indeed, we found more growth between T1 and T2 on argument reading speed (WTI) and reading comprehension for students who completed more assignments within the idiom recognition game. No effects of any other game on any other reading measure were found.

With regard to our third hypothesis, results indicated that students who completed more idiom recognition assignments showed more growth in argument reading speed and reading comprehension compared to students who completed fewer assignments, although we did not find an intervention effect. This may be explained by the fact that there were individual differences within the intervention group as a result of which some students did benefit from completing more assignments, whereas others did not. Furthermore, the results indicated that higher T1 levels of ESL decoding, morphological awareness, argument reading speed, anomaly speed and anomaly processing were predictive of game performance growth within the intervention.

Performance growth on all games within the intervention was predictive of some reading levels at T2, consistent with previous studies that found reading games can foster the development of reading skills in poor readers ([Van Gorp et al., 2017](#)).

In more detail, when looking at the relationship between decoding and game performance growth, students with better T1 decoding skills also showed more performance growth in morphosyntactic awareness and translation skills from first to second language (L2) during the intervention. It seems that L2 decoding skills foster the development of morphosyntactic skills ([Levesque et al., 2017](#)) and students' ability to improve the quality of their lexical representations using the rich context in which the target words were embedded in the translation game. This might be explained by the fact that better decoders had more cognitive resources available ([Torgesen, 1986](#)) to utilize morphological cues.

Concerning the relationship between anomaly processing and game performance growth, better T1 anomaly processing, i.e., more sensitivity to the presence of an anomaly, was related to more performance growth in morphosyntactic awareness in the intervention and more performance growth on translation of words within sentences was related to better anomaly processing at T2. Possibly, students who show larger sensitivity to semantically incongruent (relative to syntactically congruent) words at T1 have sufficient morphological and syntactic skills to benefit from the morphosyntactic awareness game ([Goodwin, 2016](#)). Further, it seems that performance growth in the translation game, which probably indirectly reflects improvement in lexical quality ([Torgesen, 1986](#)), can predict the anomaly processing at T2. This is in line with the assumption that learners with specific lexical representations were more likely to detect anomalies and show semantic incongruity effects.

The present study has limitations that need mentioning. First, while reliability of the reading comprehension measure was acceptable, it could be improved (although current reliability could not be improved by deleting items) and the results should be interpreted with caution. Furthermore, item-characteristics, such as word frequency, were not used as a predictor in the analyses, though previous studies have shown that item cues may influence performance ([De Bree et al., 2017](#)), albeit on Dutch and English spelling. Furthermore, future studies could examine an intervention with a longer duration, as intervention effects often emerge after a longer period of time than the 12-week intervention used in the present study out of feasibility considerations (e.g., [Droop et al., 2016](#)). Finally, WTI is a key process in the ability to build up a mental representation of the text, but in the present study, situation model building was not explicitly trained or assessed. Situation model building is predictive of reading comprehension in L2 learners ([Raudszus et al., 2019](#)) and could be explicitly trained or assessed in future studies.

The present study yields theoretical implications. First, the present study adds to the existing body of literature that components of WTI can be trained, although trainability of such skills has been found to be hard ([Yang et al., 2017](#)). Second, no transfer to reading comprehension arose. Third, the present study confirms the interactive nature of reading, proposed by the interactive view of reading ([Verhoeven & Perfetti, 2008](#)): training WTI does not only yield positive effects for WTI, but also for decoding. The interactive view of reading has mainly been examined in adult L1 learners, but we have now demonstrated this also applies to novice L2 learners.

The current study also yields practical implications, namely an adaptive game-based intervention may be promising in enhancing predictors of reading comprehension, thus supporting teachers in taking individual differences into account. Furthermore, teachers need to be sensitive to individual differences, as the present study confirmed that responsiveness to an ESL intervention may be (partially) dependent on students' initial English language proficiency. Finally, it seems that all components targeted in the intervention, namely dictation, morphosyntactic awareness, translation of words within sentences, and idiom

recognition, contributed to some extent to ESL reading proficiency. Therefore, all of these elements should be addressed when designing an intervention to foster WTI in ESL.

To summarize, the present study found that a WTI intervention can benefit decoding ability and anomaly processing and these effects seem to be influenced by students' ESL proficiency. Overall, our study suggests that some linguistic components of WTI can be trained by means of an adaptive, game-based learning environment. Furthermore, individual differences in initial levels of the reading measures predicted

intervention effects.

### Acknowledgements

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 facilitated realizing this project. We also thank Abe Hofman for his help with data extraction.

### Appendix 6A

Correlations between Reading Measures at T1 and Game Performance Growth Scores (for the Intervention Group)

	1	2	3	4	5	6	7	8	9	10	11
T11. Decoding	–										
2. Morphological awareness	0.55***	–									
3. Syntactic awareness	0.62***	0.66***	–								
4. Argument processing (WTI)	-0.04	-0.07	-0.05	–							
5. Argument reading speed (WTI)	-0.28**	-0.26**	-0.24*	0.30**	–						
6. Anomaly processing (WTI)	0.15	0.10	0.23*	0.03	0.05	–					
7. Anomaly speed (WTI)	-0.28**	-0.23*	-0.16	0.21*	0.87***	0.04	–				
8. Reading comprehension	0.29***	0.38***	0.40***	-0.20	-0.16	0.07	-0.06	–			
Game performance growth9. Dictation	0.09	0.02	-0.04	0.03	0.20*	-0.12	0.16	-0.06	–		
10. Morphosyntactic awareness	0.14	0.21**	0.27***	-0.06	-0.07	0.32***	-0.03	0.12	0.15	–	
11. Translation	0.17*	0.18*	0.25**	0.06	0.06	0.11	0.08	0.13	0.02	0.35***	–
12. Idiom recognition	-0.06	-0.18*	-0.22**	0.14	0.06	0.04	0.07	-0.24**	0.15	0.08	-0.15

Note.  $p < .001$ , \*\*\*,  $p < .01$ , \*\*,  $p < .05$ .

### Appendix 6B

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
T21. Decoding	–														
2. Morphological awareness	0.53***	–													
3. Syntactic awareness	0.57***	0.69***	–												
4. Argument processing (WTI)	0.17	0.34***	0.30**	–											
5. Argument reading speed (WTI)	-0.30**	-0.03	-0.07	0.13	–										
6. Anomaly processing (WTI)	-0.02	0.21*	0.07	0.30**	0.22*	–									
7. Anomaly speed (WTI)	-0.31**	-0.06	-0.12	0.12	0.92***	0.13	–								
8. Reading comprehension	0.52***	0.56***	0.57***	0.32**	-0.02	0.24*	-0.09	–							
Game performance growth9. Dictation	0.09	-0.07	-0.01	-0.04	0.23*	0.05	0.21*	0.14	–						
1. Morphosyntactic awareness	0.19*	0.17*	0.15	0.00	0.00	0.02	0.02	0.20*	0.15	–					
11. Translation	0.20*	0.25**	0.23**	0.21*	0.10	0.23*	0.02	0.17*	0.02	0.35***	–				
12. Idiom recognition	-0.06	-0.14	-0.21**	-0.06	-0.06	-0.01	0.00	-0.03	0.15	0.08	-0.15	–			
Completed assignments13. Dictation	0.09	0.11	0.12	-0.02	0.01	-0.03	0.01	0.05	0.10	0.20*	0.17*	0.14	–		
14. Morphosyntactic awareness	0.20*	0.03	0.14	-0.05	-0.18	-0.05	-0.15	0.04	0.11	0.23**	0.16*	0.19*	0.59***	–	
15. Translation	0.08	0.02	0.06	-0.08	-0.03	0.00	-0.08	-0.03	0.15	0.15	0.23**	0.13	0.64***	0.74***	–
16. Idiom recognition	0.04	-0.01	-0.03	-0.04	-0.15	-0.09	-0.15	-0.01	0.14	0.02	-0.02	0.25**	0.54***	0.63***	0.71***

Correlations between Reading Measures at T2, Game Performance Growth Scores, and Completed Assignments (for the Intervention Group)

### References

Abdi, H. (2010). Holm's sequential Bonferroni procedure. *Encyclopedia of Research Design*, 1(8), 1–8.

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

Barber, A. T., Cartwright, K. B., Stapleton, L., Klauda, S. L., Archer, C., & Smith, P. (2020). Direct and Indirect Effects of Executive Functions, Reading Engagement, and Higher Order Strategic Processes in the Reading Comprehension of Dual Language Learners and English Monolinguals. *Contemporary Educational Psychology*, 61, 1–17.

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

Berkum, Jos J. A. van, Hagoort, Peter, & Brown, Colin M. (1999). Semantic integration in sentences and discourse: Evidence from the N400. *Journal of Cognitive Neuroscience*, 11(6), 657–671.

Bowers, P. N., Kirby, J. R., & Deacon, S. H. (2010). The effects of morphological instruction on literacy skills: A systematic review of the literature. *Review of Educational Research*, 80(2), 144–179.

BULTENA, SYBRINE, DIJKSTRA, TON, & VAN HELL, JANET. G. (2015). Language switch costs in sentence comprehension depend on language dominance: Evidence from self-paced reading. *Bilingualism: Language and Cognition*, 18(3), 453–469.

CAIN, KATE (2007). Syntactic awareness and reading ability: Is there any evidence for a special relationship? *Applied Psycholinguistics*, 28(4), 679–694.

Cain, K., & Oakhill, J. (2006). Profiles of children with specific reading comprehension difficulties. *British Journal of Educational Psychology*, 76(4), 683–696.

Cain, K., Oakhill, J., & Lemmon, K. (2005). The relation between children's reading comprehension level and their comprehension of idioms. *Journal of Experimental Child Psychology*, 90(1), 65–87.

Cain, K., Towse, A. S., & Knight, R. S. (2009). The development of idiom comprehension: An investigation of semantic and contextual processing skills. *Journal of Experimental Child Psychology*, 102(3), 280–298.

- College voor Toetsen en Examens - Board for Assessment and Exams. (2016). Examen Engels KB VMBO, 1e tijdvak 2016 en VMBO-GL/TL, 2e tijdvak 2012 (English Exam 1st Time Block 2016, and 2nd Time Block 2012). Retrieved from: <https://www.examenblad.nl>.
- Degand, L., & Sanders, T. (2002). The impact of relational markers on expository text comprehension in L1 and L2. *Reading and Writing: An Interdisciplinary Journal*, 15, 739–757.
- Dina, A., & Ciornei, S.-I. (2013). The advantages and disadvantages of computer assisted language learning and teaching for foreign languages. *Procedia-Social and Behavioral Sciences*, 76, 248–252.
- De Bree, et al. (2017). *Language Learning and Development*.
- Droop, et al. (2016). *Journal of Research on Educational Effectiveness*.
- Edmonds, M. S., Vaughn, S., Wexler, J., Reutebuch, C., Cable, A., Tackett, K. K., & Schnakenberg, J. W. (2009). A synthesis of reading interventions and effects on reading comprehension outcomes for older struggling readers. *Review of Educational Research*, 79(1), 262–300.
- Elbro, C., & Buch-Iversen, I. (2013). Activation of background knowledge for inference making: Effects on reading comprehension. *Scientific Studies of Reading*, 17(6), 435–452.
- Ellis, R. (2015). *Understanding second language acquisition 2nd edition - Oxford applied linguistics*. Oxford: Oxford University Press.
- Elo, A. (1978). *The ratings of chessplayers*. Arco: Past and Present.
- Gebauer, S. K., Zaunbauer, A. C. M., & Möller, J. (2013). Cross-language transfer in English immersion programs in Germany: Reading comprehension and reading fluency. *Contemporary Educational Psychology*, 38(1), 64–74.
- Goodwin, A. P. (2016). Effectiveness of word solving: Integrating morphological problem-solving within comprehension instruction for middle school students. *Reading and Writing*, 29(1), 91–116.
- Goriot, C., Broersma, M., McQueen, J. M., Unsworth, S., & van Hout, R. (2018). Language balance and switching ability in children acquiring English as a second language. *Journal of Experimental Child Psychology*, 173, 168–186.
- Hayes, A. F. (2018). *Introduction to mediation, moderation, and conditional process analysis* (2nd Ed.). New York: The Guilford Press.
- Helder, & Perfetti. (2021). *Discourse Processes*.
- Hung, H.-T., Yang, J. C., Hwang, G.-J., Chu, H.-C., & Wang, C.-C. (2018). A scoping review of research on digital game-based language learning. *Computers & Education*, 126, 89–104.
- IBM Corp. Released 2019. IBM SPSS Statistics for Windows, Version 26.0. Armonk, NY: IBM Corp.
- Kline, P. (1993). *Handbook of psychological testing*. London: Routledge.
- Koda, K. (2007). Reading and language learning: Crosslinguistic constraints on second language reading development. *Language Learning*, 57, 1–44.
- Klinkenberg, et al. (2011). *Computers & Education*.
- Language and Reading Research Consortium, Currie, N. K., & Muijselaar, M. M. L. (2019). Inference making in young children: The concurrent and longitudinal contributions of verbal working memory and vocabulary. *Journal of Educational Psychology*, 111(8), 1416–1431.
- Lesaux, Nonie K., Lipka, Orly, & Siegel, Linda S. (2006). Investigating cognitive and linguistic abilities that influence the reading comprehension skills of children from diverse linguistic backgrounds. *Reading and Writing: An Interdisciplinary Journal*, 19(1), 99–131.
- Lesaux, N. K., Kieffer, M. J., Faller, S. E., & Kelley, J. G. (2010). The effectiveness and ease of implementation of an academic vocabulary intervention for linguistically diverse students in urban middle schools. *Reading Research Quarterly*, 45(2), 196–228.
- Levesque, Kyle C., Kieffer, Michael J., & Deacon, S. Hélène (2017). Morphological awareness and reading comprehension: Examining mediating factors. *Journal of Experimental Child Psychology*, 160, 1–20.
- Li, S. (2010). The effectiveness of corrective feedback in SLA: A meta-analysis. *Language Learning*, 60(2), 309–365.
- Liu, Yingyi, Yeung, Susanna Siu-sze, Lin, Dan, & Wong, Richard Kwok Shing (2017). English expressive vocabulary growth and its unique role in predicting English word reading: A longitudinal study involving Hong Kong Chinese ESL children. *Contemporary Educational Psychology*, 49, 195–202.
- McMaster, Kristen L., van den Broek, Paul, Espin, Christine A., Pinto, Viveca, Janda, Bonita, Lam, Elizabeth, ... van Boekel, Martin (2015). Developing a reading comprehension intervention: Translating cognitive theory to educational practice. *Contemporary Educational Psychology*, 40, 28–40.
- Melby-Lervåg, M., & Lervåg, A. (2014). Reading comprehension and its underlying components in second-language learners: A meta-analysis of studies comparing first- and second-language learners. *Psychological Bulletin*, 140(2), 409–433.
- Mulder, et al. (2020). *Reading and Writing*.
- Muijselaar, et al. (2017). *Scientific Studies of Reading*.
- Myers, R. (1990). *Classical and modern regression with applications* (2nd ed.). Boston: Duxbury.
- Nayak, Gopa, & Sylva, Kathy (2013). The effects of a guided reading intervention on reading comprehension: A study on young Chinese learners of English in Hong Kong. *The Language Learning Journal*, 41(1), 85–103.
- Nouri, J., & Pargman, T. C. (2016). In *September*. *When teaching practices meet tablets' affordances. Insights on the materiality of learning* (pp. 179–192). Cham: Springer.
- Ouellette, Gene, & Beers, Ashley (2010). A not-so-simple view of reading: How oral vocabulary and visual- word recognition complicate the story. *Reading and Writing*, 23(2), 189–208.
- Perfetti, C. A., & Hart, L. (2002). The lexical quality hypothesis. In L. Verhoeven, C. Elbro, & P. Reitsma (Eds.), *Precursors of functional literacy* (pp. 67–86). Amsterdam, The Netherlands: John Benjamins Publishing Co.
- Perfetti, Charles, & Stafura, Joseph (2014). Word knowledge in a theory of reading comprehension. *Scientific Studies of Reading*, 18(1), 22–37.
- Perfetti, Charles, Yang, Chin-Lung, & Schmalhofer, Franz (2008). Comprehension skill and word-to-text integration processes. *Applied Cognitive Psychology*, 22(3), 303–318.
- Persson, V., & Nouri, J. (2018). A systematic review of second language learning with mobile technologies. *International Journal of Emerging Technologies in Learning (IJET)*, 13(02), 188–210.
- Porter, A. (2019). An early start to foreign language literacy in English primary school classrooms. *The Language Learning Journal*, 1–16.
- Preacher, Kristopher J., & Hayes, Andrew F. (2004). SPSS and SAS procedures for estimating indirect effects in simple mediation models. *Behavior Research Methods, Instruments, & Computers*, 36(4), 717–731.
- R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>.
- Saine, N. L., Lerkkanen, M. K., Ahonen, T., Tolvanen, A., & Lyytinen, H. (2011). Computer-assisted remedial reading intervention for school beginners at risk for reading disability. *Child Development*, 82(3), 1013–1028.
- Sandberg, Jacobijn, Maris, Marinus, & Hoogendoorn, Pepijn (2014). The added value of a gaming context and intelligent adaptation for a mobile learning application for vocabulary learning. *Computers and Education*, 76, 119–130.
- Scammacca, Nancy K., Roberts, Greg, Vaughn, Sharon, & Stuebing, Karla K. (2015). A meta-analysis of interventions for struggling readers in grades 4–12: 1980–2011. *Journal of Learning Disabilities*, 48(4), 369–390.
- Siegel, L. S. (2008). Morphological awareness skills of English language learners and children with dyslexia. *Topics in Language Disorders*, 28, 15–27.
- Singmon, M., Mahony, D., & Mann, V. (2000). The relation between reading ability and morphological skills: Evidence from derivational suffixes. *Reading and Writing*, 12(3), 219–252.
- Stafura, Joseph Z., Rickles, Benjamin, & Perfetti, Charles A. (2015). ERP evidence for memory and predictive mechanisms in word-to-text integration. *Language, Cognition, and Neuroscience*, 30(10), 1273–1290.
- Tabachnick, B. G., & Fidell, L. S. (2012). *Using multivariate statistics* (6th ed.). Boston, MA: Pearson.
- Thijs, A. M., Trimbo, B., Tuin, D., Bodde, M., & De Graaff, R. (2011). *Engels in het basisonderwijs: Vakdossier [English in primary education: Subject file]*. The Netherlands, Enschede: SLO.
- Tong, X., Deacon, S. H., Kirby, J. R., Cain, K., & Parrilla, R. (2011). Morphological awareness: A key to understanding poor reading comprehension in English. *Journal of Educational Psychology*, 103(3), 523–534.
- Torgesen, Joseph K. (1986). Computers and cognition in reading: A focus on decoding fluency. *Exceptional Children*, 53(2), 157–162.
- Torgesen, J. K., Wagner, R. K., & Rashotte, C. A. (2012). *TOWRE 2: Test of word reading efficiency*. Austin, TX: Pro-ed.
- van der Linden, Wim J., & Glas, Gees A. W. (Eds.). (2000). *Computerized Adaptive Testing: Theory and Practice*. Dordrecht: Springer Netherlands.
- Van Gorp, et al. (2017). *Reading Research Quarterly*.
- Verhoeven, & van Leeuwe. (2012). *Reading and writing*.
- Verhoeven, & Perfetti. (2008). *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition*.
- Wright, Margaret, & Brown, Pamela (2006). Reading in a modern foreign language: Exploring the potential benefits of reading strategy instruction. *The Language Learning Journal*, 33(1), 22–33.
- Wright, Tanya S., & Cervetti, Gina N. (2017). A systematic review of the research on vocabulary instruction that impacts text comprehension. *Reading Research Quarterly*, 52(2), 203–226.
- Words&Birds. (2015). Amsterdam, the Netherlands: Oefenweb. Retrieved from: <https://www.wordsandbirds.nl/>.
- Yang, C., Crain, S., Berwick, R. C., Chomsky, N., & Bolhuis, J. J. (2017). The growth of language: Universal Grammar, experience, and principles of computation. *Neuroscience & Biobehavioral Reviews*, 81(B), 103–119.
- Yang, C. L., Perfetti, C. A., & Schmalhofer, F. (2005). Less skilled comprehenders' ERPs show sluggish word-to-text integration processes. *Written Language & Literacy*, 8(2), 157–181.
- Yang, C. L., Perfetti, C. A., & Schmalhofer, F. (2007). Event-related potential indicators of text integration across sentence boundaries. *Journal of Experimental Psychology*, 33, 55–89.
- Yuill, Nicola, & Oakhill, Jane (1988). Effects of inference awareness training on poor reading comprehension. *Applied Cognitive Psychology*, 2(1), 33–45.
- Zwaan, Rolf A. (2016). Situation models, mental simulations, and abstract concepts in discourse comprehension. *Psychonomic Bulletin & Review*, 23(4), 1028–1034.