Goldhammer, Frank; Kroehne, Ulf; Hahnel, Carolin; De Boeck, Paul

Controlling speed in component skills of reading improves the explanation of reading comprehension

formal und inhaltlich überarbeitete Version der Originalveröffentlichung in:
formally and content revised edition of the original source in:
The Journal of educational psychology 113 (2021) 5, S. 861-878

Controlling speed in component skills of reading improves the explanation of reading comprehension

Frank Goldhammer[1,2], Ulf Kroehne[1], Carolin Hahnel[1,2], Paul De Boeck[3]

[1]DIPF | Leibniz Institute for Research and Information in Education

[2]Centre for International Student Assessment (ZIB)

[3]Ohio State University

Correspondence concerning this article should be addressed to Frank Goldhammer, DIPF | Leibniz Institute for Research and Information in Education, Rostocker Straße 6, 60323 Frankfurt/Main, Germany, Phone: +49 (0) 69.24708 323, Fax: +49 (0) 69.24708 444, E-mail: Goldhammer@dipf.de.

Abstract

Efficiency in reading component skills is crucial for reading comprehension, as efficient sub-processes do not extensively consume limited cognitive resources, making them available for comprehension processes. Cognitive efficiency is typically measured with speeded tests of relatively easy items. Observed responses and response times indicate the latent variables of ability and speed. Interpreting only ability or speed as efficiency may be misleading because there is a within-person dependency between both variables (speed-ability tradeoff, SAT). Therefore, the present study measures efficiency as ability conditional on speed by controlling speed experimentally with item-level time limits. The proposed timed ability measures of reading component skills are expected to have a clearer interpretation in terms of efficiency and to be better predictors for reading comprehension. To support this claim, this study investigates two component skills, visual word recognition and sentence-level semantic integration (sentence reading), to understand how differences in ability in a timed condition are related to differences in ability and speed in a traditional untimed condition. Moreover, untimed and timed reading component skill measures were used to explain reading comprehension. A German subsample from PISA 2012 completed the reading component skills tasks with and without item-level time limits and PISA reading tasks. The results showed that timed ability is only moderately related to untimed ability. Furthermore, timed ability measures proved to be stronger predictors of sentence-level and text-level reading comprehension than the corresponding untimed ability and speed measures; although using untimed ability and speed jointly as predictors increased the amount of explained variance.

Keywords: efficiency in reading component skills; reading comprehension; experimental control of speed; item-level time limits; conditional effects of speed and ability

Educational Impact and Implications Statement

The study suggests that the assessment of reading component skills (e.g., word recognition) can be improved by controlling the time a reader spends solving individual tasks. The new measures provide a clearer picture of how strongly lower level processes support or hamper reading comprehension.

Controlling speed in component skills of reading improves the explanation of reading comprehension

Cognitive efficiency can be conceptualized as the product in a cognitive task (work output, e.g., accuracy) in relation to its costs (work input, e.g., invested time, cognitive effort, conscious control). Given the work output, efficiency is higher if the work input was lower, and, inversely, given the work input, efficiency is higher if the work output is of higher quality. Conceptions of cognitive efficiency are used in many domains, for instance, from neurological, instructional, and learning perspectives (Hoffman, 2012; Perfetti, 2007), and play an important role in understanding cognitive information processing that is organized hierarchically.

Cognitive higher-order constructs (e.g., broad abilities, competencies) rely on sub-processes or elementary component skills (e.g., perceptual speed, Ackerman & Beier, 2007). If these information processing elements can be performed efficiently, that is, fast and correct in an automatic fashion, (limited) cognitive resources, such as working memory capacity become available for higher-order cognitive processing. The benefit of efficient component skills applies to various domains, such as general cognitive functioning (e.g., Salthouse, 1996), reading comprehension (Perfetti & Hart, 2002), and even to problem solving, which by definition is controlled processing (Carlson et al., 1990).

In the present study we focus on the domain of reading with visual word recognition as component skill of semantic integration at the sentence level (sentence-level semantic integration) and both visual word recognition and sentence-level semantic integration as sub-processes of reading comprehension at the text level (Hunt, 1978; Perfetti & Stafura, 2013). Note that in the present study sentence-level integration refers to reading and comprehending a single sentence; this semantic integration of word meanings within a sentence is more basic as

compared to higher-order integration across multiple sentence and causal inferencing (e.g., Barnes et al., 2015).

Previous research on the relation of reading component skills, such as word decoding to reading comprehension, has shown that the way reading component skills are measured (e.g., accuracy vs. speed) may affect the strength of the relationship (see e.g., the meta-analyses by Florit & Cain, 2011; García & Cain, 2014). In this work, we argue that for the adequate measurement of efficiency in reading component skills both accuracy and speed need to be considered. Particularly, we aim at capturing the efficiency of visual word recognition and sentence-level semantic integration by a new measurement approach controlling response speed in such a way that assessed individual differences in accuracy clearly represent the relation of work output to work input. By using experimental item-level time limits (Goldhammer & Kroehne, 2014), we expect to obtain a better measurement of the efficiency in reading components skills which is unaffected by individual differences in the speed-ability tradeoff (SAT, van der Linden, 2009). This assumption is empirically tested by investigating whether reading comprehension at sentence level and text level is predicted more strongly by the proposed new (timed) reading component skill measures than by traditional (untimed) component skill measures.

Theoretical models explaining reading comprehension

Reading comprehension is a complex and hierarchically organized cognitive process (Kintsch, 1998; Kintsch & van Dijk, 1978; Perfetti, 1985; Perfetti & Stafura, 2013). It includes bottom up and top down mechanisms working together at the word, sentence, and text levels. At the lowest level, readers need to identify single letters and words. Specifically, visual word recognition relates a written word to a representation in the mental lexicon, involving different types of sub-processes: Phonological recoding assigns letters to phonemes to obtain a phonetic

representation of the written word; orthographic comparison compares the spelling of the word with the mental orthographic representation (Cunningham & Stanovich, 1990); and finally, with the support from high-quality orthographical and phonological representations, word meaning is activated, enabling comprehension at the sentence and text levels (Richter et al., 2013). At the level of sentences, the syntactical structure of a sentence is parsed, and the words are semantically integrated into a coherent and meaningful representation. Furthermore, at the text level, higher-order global coherence must be established among sentences and paragraphs to obtain an integrated propositional representation of the text content. Finally, readers will activate prior knowledge, enabling them to draw additional inferences about the text content, ensuring coherence (situation model; see construction-integration model by Kintsch, 1998).

From a different perspective than the construction-integration model (Kintsch, 1998), which focuses on how the comprehension of a text dynamically evolves over time, the "simple view of reading" model (SVR) explains reading comprehension as a function of underlying capacities at any given point in time (Hoover & Gough, 1990; Hoover & Tunmer, 2018). It claims that reading comprehension is determined by two abilities. Decoding (also referred to as word reading or word recognition) is the ability to quickly and accurately retrieve the meaning of written words from the mental lexicon, whereas language comprehension (also referred to as linguistic or listening comprehension) is the ability to construct and derive linguistic discourse meaning from semantic information. Thus, the SVR posits that reading comprehension involves the same cognitive processes as language comprehension but requires decoding ability as a result of processing written text information rather than oral verbalization. Similarly, semantic integration at sentence level can be expected to rely on language comprehension and, in the case of written sentences, also on decoding. Hoover and Tunmer (2018) emphasize that decoding needs to be accurate and quick as well. If decoding is slow, its outcome might not be cognitively

represented anymore when it needs to be integrated for sentence and discourse interpretation. There is extensive empirical evidence to support the SVR (e.g., Hoover & Gough, 1990; Kim, 2019; Kim, 2015; Language and Reading Research Consortium & Logan, 2017; Vellutino et al., 2007). A recent extension of the SVR, the direct and indirect effects model of reading (DIER) (Kim, 2019), adds language and cognitive components skills underlying decoding and language comprehension. It introduces text-reading fluency as a factor mediating the effects of decoding and language comprehension and makes specific assumptions about the hierarchical, interactive and dynamic relations of these variables.

For the present study empirical findings on the (relative) contribution of visual word recognition and components that reflect language comprehension (e.g., sentence-level semantic integration) to the explanation of reading comprehension are of particular interest. There is a wealth of previous research showing positive relations of decoding and language comprehension to reading comprehension (e.g., Florit & Cain, 2011; García & Cain, 2014). The research findings in question are mainly based on samples of young readers (i.e., primary school students) and measurements without item-level time limits. The SVR predicts that for beginning readers decoding is particularly relevant for reading comprehension, whereas for advanced readers with efficient word recognition language comprehension becomes the limiting factor. Accordingly, studies using decoding and language comprehension jointly as predictors have shown that the relative contribution of decoding to reading comprehension is greater in early stages of reading, whereas in later stages the relative contribution of language comprehension is greater; for instance, Kim (2019) compares grade 2 vs. grade 4 (see also Foorman et al., 2018; Kim et al., 2012; Vellutino et al., 2007).

Efficiency in reading component skills

Cognitive theories of reading posit that reading comprehension relies on efficient component skills. The automaticity theory proposed by LaBerge and Samuels (1974) describes reading as a multi-stage process including visual, phonological, and episodic memory systems. In the beginning, attentional control and effort is needed for processing at various stages. However, when word meanings can be recognized automatically, attention can stay focused on the semantic and comprehension levels and does not need to be switched back and forth to decoding processes of reading. In continuation of Ehri's (2005a, 2005b) phases of learning to read words, ranging from the pre-alphabetic phase to the consolidated alphabetic phase, the consolidated phase is followed by automaticity, which enables skilled readers to automatically recognize the pronunciation and meaning of written words by seeing them.

Related to automaticity theory, the verbal efficiency theory (Perfetti, 1985) postulates that readers comprehend written text more easily if sub-processes that are amenable to automation can actually be performed efficiently. Specifically, verbal efficiency theory regards word identification and, to some extent, elementary propositional encoding (i.e., constructing a proposition based on several words) as candidates for automation, whereas others, such as inference processes, are not automated. Perfetti (2007) defines efficiency as the ratio of outcome to effort (i.e., time) and puts emphasis on the quality of the outcome as the source of efficiency, that is, the quality of lexical representations (knowledge about words). The lexical quality hypothesis (Perfetti, 2007; Perfetti & Hart, 2002) claims that reading comprehension is based on reliable and quickly retrievable lexical representations. The stability and precision of these representations determine the speed, accuracy, and ease with which a word can be retrieved and identified. Thus, the quality of lexical representations is a limiting factor in the reading comprehension process (Perfetti & Stafura, 2013).

Through practice, component skills can become increasingly automated, that is, accurate, quick, and effortless (Samuels & Flor, 1997). They place lower demands on the limited cognitive resources, which are then available for higher-order comprehension processes at the sentence level and text level (LaBerge & Samuels, 1974; Walczyk, 2000). Component skills at the sentence level are not assumed to become fully automated because syntactic parsing and semantic integration may require controlled processing to some extent, for instance, when inferring meaning by considering syntactical information. However, more efficient component skills at sentence level have been shown to foster comprehension at text level (Richter et al., 2012), and increasing syntactic and propositional complexity of sentences to have a weaker negative effect on reading time for fast (i.e., good) readers (Graesser et al., 1980). Although visual word recognition represents a component skill of semantic integration at sentence level, word recognition is assumed to also contribute uniquely to text-level comprehension, in that high-quality lexical representations support the integration of sentence sequences, situation updating, and inferencing (Perfetti & Stafura, 2013). Readers who are not in command of automated component skills (i.e., information processing is slow and of poor quality) are expected to accomplish reading comprehension tasks with higher demands of controlled and strategic processing. For example, they may rely on time-consuming compensatory behaviors, in particular, when completing difficult tasks (Walczyk, 2000; Walczyk et al., 2007).

Measuring word recognition and sentence-level semantic integration

Reading component skills, such as visual word recognition and semantic integration at sentence level, which are central to the present study, have been measured in different ways in previous research, using information about response accuracy, response time, or both. For measuring decoding, Hoover and Gough (1990) let primary school students complete a task to decode synthetic words and scored how accurately they pronounced these synthetic words. A

similar approach for scoring such decoding accuracy was used by Vellutino et al. (2007) in a context-free word identification task. For predicting reading comprehension, Richter et al. (2013) considered not only accuracy but also response time to represent the quality of access to lexical representations. The meta-analysis of Florit and Cain (2011) on the validity of SVR to different alphabetic orthographies considered different measures of decoding to investigate how measures of decoding fluency and decoding accuracy are related to reading comprehension for readers of transparent alphabetic orthographies. In the studies included in the meta-analysis, measures of decoding fluency were obtained, for instance, by the number of words read accurately divided by the time needed (Hagtvet, 2003) or as the number of correct answers achieved within an overall time limit (Droop & Verhoeven, 2003; see also, Kim et al., 2012; Kim, 2019; Kim, 2015; Language and Reading Research Consortium & Logan, 2017). In such efficiency or fluency tasks, respondents are usually instructed to answer as accurately and as quickly as possible.

Measurements of semantic integration at the level of single sentences typically require a respondent to read a sentence and verify its semantic content. Richter et al. (2012) used both response accuracy and response time in a sentence verification task to assess semantic integration at sentence level. A similar task was used by Kim et al. (2012) to measure silent reading fluency at the level of sentences (see also Johnson et al., 2011). The employed Test of Silent Reading Efficiency and Comprehension (TOSREC; Wagner et al., 2010) requires the reader to verify a number of sentences within an overall time limit. The fluency score is calculated as the number of correct responses minus the number of incorrect responses (to correct for guessing). Notably, in other studies the TOSREC task was used to assess reading comprehension (e.g., Ahmed et al., 2016; Lonigan et al., 2018), whereas the semantic integration task from Richter et al. (2012) used in the present study can be regarded as reading comprehension task at sentence level. Note that across these studies there are major differences in the administration procedure: Richter et al.

(2012) do not impose a time limit, the studies using TOSREC impose a time limit at the test level, and in the present study the time limit is implemented at item level.

Measuring efficiency of component skills considering the speed-ability tradeoff

As indicated by the review of measurements in the previous section, the efficiency of cognitive component skills is typically measured with performance tests showing a strong speed component, that is, the challenge is to complete as many items correctly as possible in a limited amount of time or to spend as little time as possible to complete a fixed number of items correctly (speed test, Gulliksen, 1950). Pure speed tests do not exist, so observations at the item level are response accuracy and response time (see Luce, 1986; van der Linden, 2009). Psychometrically, the variation in these manifest variables between persons can be explained by latent person variables of effective ability and effective speed and potential residual dependencies within items. "Effective" refers to the actual balance between speed and accuracy a person is using in the test, with consequences for ability estimates (based on accuracy) and speed estimates. A respondent may choose or change his or her speed and accuracy across situations or conditions. The *within*-person dependency of (effective) ability and (effective) speed is referred to as the speed-ability tradeoff (SAT) function (van der Linden, 2009), in accordance with the speed-accuracy tradeoff function investigated in experimental cognitive psychology (Luce, 1986; Wickelgren, 1977). Here, a person completing a test may increase (effective) speed at the cost of (effective) ability resulting in a negative within-person relation of ability and speed. Note that *between* persons the relation of ability and speed can be of any direction depending on characteristics of persons and items.

From the perspective of modelling latent person variables of ability and speed, multiple options to represent individual differences in efficiency can be derived. First, for each person an effective speed and an effective ability can be estimated for an unknown point on his or her

continuous SAT function (van der Linden, 2007). This corresponds to what can typically be done in testing without controlling a respondent's point on the function. The point on the tradeoff function depends on the respondent's choice of speed. Without controlling speed, the interpretation of observed differences in effective ability is ambiguous because they may be due to differences in the individual SAT function, the decision on speed, or both (Goldhammer, 2015; Goldhammer et al., 2017). Note that this concern also applies to speeded tests employing a global time limit.

Second, individuals can be compared on their within-person SAT function and the parameters describing this function, namely intercept, slope, and asymptote of the function of speed (x-axis) versus ability (y-axis) (e.g., Goldhammer et al., 2017; Lohman, 1989). In particular, the individual slope parameter of the tradeoff function would reflect the rate of information accumulation by using more time to respond. Representing the within-person SAT function requires implementing multiple speed conditions using an experimental within-subject design. Although this option provides the richest information, we did not include it into the present study because it requires the greatest effort in terms of testing material and time necessary for testing.

A third option, in the focus of the present study, is to measure ability conditional on a fixed speed level. This is done by implementing a single, medium-fast speed condition as opposed to the second option requiring multiple speed conditions ranging from slow to fast. The respondent's decision criterion on how much time to take to produce a correct response is controlled experimentally and, as a consequence, response speed is no longer a choice (Goldhammer & Kroehne, 2014; Lohman, 1989; Salthouse & Hedden, 2002). This seems particularly appropriate for speeded tests or cognitive efficiency measurements, where difficulty is supposed to be determined also by limited time and time pressure, respectively. The third

option provides ability estimates only for a selected speed condition, whereas the second option provides this information for various speed conditions to map the individual SAT function.

The present study

The measures obtained from the first and third options are focused on and compared in the present study. They can be conceptually decomposed as follows. When response speed is controlled (third option), the ability estimate captures individual efficiency. This is the case because the observed accuracy represents the work output in relation to the work input (time to work), which is fixed across individuals by the experimental control of response speed: *timed ability* = efficiency. However, when speed is not controlled (first option), the ability estimate is a function of both individual efficiency, as defined above, and the influence of the individual SAT, *untimed ability* = *f*(efficiency, SAT), and the speed estimate is also a function of efficiency and the individual SAT, *untimed speed* = *f*(efficiency, SAT). An SAT in favor of accuracy (e.g., the respondent is more concerned about the correctness of answers) has a positive effect on untimed ability and a negative effect on untimed speed reflected by working more slowly. Thus, favoring effective ability impairs effective speed and vice versa.

There are various (experimental) approaches to manipulate response speed by influencing the time spent on an item, such as verbal instructions or pay-offs, as well as response time deadlines or windows (Salthouse & Hedden, 2002; Wickelgren, 1977). In the present study, a timed condition was implemented by means of the response signal paradigm (Reed, 1973) requiring respondents to respond immediately once a response signal is given. A general prerequisite of this approach is that respondents are equally able to adjust their timing and response behavior according to the time-limits (Bolsinova & Tijmstra, 2015; Goldhammer, 2015). The untimed condition provides estimates of untimed ability and untimed speed, whereas, for the timed condition, only timed ability (efficiency) can be estimated because speed has

become an experimentally fixed variable. The latter assumption may be violated if respondents process the stimulus without having changed their decision criterion; that is, they may be in fact faster and wait for the signal to come, or they may run out of time and finally guess rapidly.

The overall goal of the present study was to investigate how the experimental control of response speed by moderate item-level time limits affects individual differences in measures of reading component skills. Two component skill measures were selected that tap central sub-processes of reading comprehension, namely visual word recognition and sentence-level semantic integration, which facilitate reading comprehension if they can be performed efficiently (Richter et al., 2012; Richter et al., 2013). To attain our goal, we investigated the relation between untimed and timed measures of the two reading component skills, as well as their relations to reading comprehension.

Hypotheses

Overall, we assume that untimed measures, as opposed to timed measures of reading component skills, are confounded with the SAT, expecting that the work input at item level (i.e., invested time) differs among individuals. As a consequence, timed and untimed measures of reading component skills are only moderately related (see Hypotheses 1.1, 2.2 below). Most importantly, timed measures have a clearer interpretation in terms of cognitive efficiency and are in turn assumed to be better predictors for reading comprehension at sentence level and text level than untimed measures (see Hypotheses 2.1, 3.1, 3.3 below). Furthermore, since efficiency in the untimed condition is represented by both untimed measures (e.g., respondents with the same untimed ability may differ in untimed speed and, therefore, in efficiency), we expect that the prediction by untimed measures can be improved when using untimed speed and untimed ability jointly as predictors (see Hypotheses 1.2, 2.3, 3.2 below).

In detail, we addressed the following hypotheses. The hypotheses 1.1 and 1.2 refer to the relation between timed and untimed measures, and both were tested separately for visual word recognition and semantic integration at sentence level.

Hypothesis 1.1: In untimed testing, respondents may trade speed for accuracy and vice versa, whereas this is prevented in timed testing by using item-level time limits. Thus, timed ability representing efficiency should only be moderately explained by untimed ability (e.g., Goldhammer & Kroehne, 2014) or untimed speed, as untimed measures represent both efficiency and inter- and intra-individual differences in the SAT.

Hypothesis 1.2: The explanation of timed ability by untimed measures is expected to become stronger by using untimed ability and untimed speed jointly as predictors. This is because in untimed testing, effective ability and effective speed together reflect how efficient the respondent was. The unique effects of untimed ability and untimed speed are assumed to be positive, that is, those respondents who were able to complete the items in the untimed condition both correctly and quickly were assumed to be most successful in the timed condition.

The hypotheses 2.1, 2.2, and 2.3 address the explanation of semantic integration at sentence level by visual word recognition (note that word recognition is always visual in the present study, even if we refer to only word recognition in the following).

Hypothesis 2.1: Word recognition represents a component skill of semantic integration at sentence level (Perfetti & Stafura, 2013), suggesting a positive relation to semantic integration. Since in the untimed condition, the SAT may vary inter- and intra-individually, sentence-level semantic integration ability is assumed to be more strongly explained by word recognition ability in the timed than in the untimed condition.

Hypothesis 2.2: As with Hypothesis 1.1, timed sentence-level semantic integration ability is expected to be only moderately explained by untimed ability in word recognition or untimed speed in word recognition.

Hypothesis 2.3: Similar to Hypothesis 1.2, the explanation of timed ability in sentence-level semantic integration is expected to be higher by using untimed ability and untimed speed in word recognition jointly as predictors. Again, the unique effects of untimed ability and untimed speed are assumed to be positive.

Finally, the hypotheses 3.1, 3.2, and 3.3 focus on the explanation of reading comprehension by sentence-level semantic integration and visual word recognition.

Hypothesis 3.1: Visual word recognition and sentence-level semantic integration are component skills of reading comprehension (Perfetti & Hart, 2002), suggesting that word recognition and sentence-level semantic integration positively explain reading comprehension. Given inter- and intra-individual differences in the SAT in the untimed condition, reading comprehension is assumed to be explained more strongly by timed ability in word recognition and sentence-level semantic integration than by untimed ability or untimed speed in word recognition and sentence-level semantic integration.

Hypothesis 3.2: As with Hypothesis 1.2 and 2.3, the explanation of reading comprehension by untimed ability and untimed speed in word recognition and sentence-level semantic integration is expected to be improved by using the four untimed measures jointly as predictors (compared with only untimed speed or untimed ability). Respondents are assumed to show better comprehension if they can demonstrate higher effective ability and effective speed in both word recognition and sentence-level semantic integration.

Hypothesis 3.3: As argued above, timed and untimed measures of reading component skills are not assumed to be completely independent. So, the unique effect of timed ability

measures is likely smaller when controlling for commonalities with the untimed measures. Nevertheless, we expect that the unique timed condition effects mainly explain reading comprehension and the untimed measures would not further account for much.

In addition to these hypotheses, we investigated how the experimental control of speed affects the precision of measurement and assumed that item-level time limits improve reliability (see Goldhammer & Kroehne, 2014). We modelled effective (untimed) ability and speed following the hierarchical model of van der Linden (2007); however, we also inspected residual dependencies between responses and response times within items (Bolsinova, Tijmstra, et al., 2017) and investigated whether the consideration of these dependencies affects the structural parameters of the latent regression models or not. Regarding the direction of the residual dependencies, we assumed that in the mode of controlled processing, being slower than expected is positively correlated with success and in the mode of automatic processing, being slower than expected is negatively correlated with success (see Goldhammer et al., 2014). For the visual word recognition task, controlled processing was expected for non-words, resulting in positive residual correlations; in contrast, words are likely to be processed automatically by experienced readers, suggesting rather negative residual correlations. The sentence-level semantic integration task requires the conscious inference of the meaningfulness of sentences, suggesting controlled processing to some extent, not only for false but also for correct sentences. Therefore, we expected positive residual correlations. We also inspected by stimulus type whether the residual correlation is related to item difficulty, as suggested by previous research (Bolsinova, De Boeck, et al., 2017; Goldhammer et al., 2014).

Method

Sample

A total sample of 888 students aged from 15.33 to 16.33 years (M = 15.82, SD = .29) participated in the study. The sample included 46.17% female and 49.21% male students, and 4.62% without indication. These students participated in both the PISA 2012 main study and a German add-on study. The sampling procedure for the main study consisted of two stages in which PISA 2012-eligible schools were first sampled, and then 25 students were drawn randomly from each selected school. For the add-on study, a subset of 77 schools were included with up to 14 students sampled from the group of 25 students. The test instruments were administered in groups using bring-in notebooks.

Instruments and booklet design

Visual word recognition. The lexical decision task (Balota et al., 2004; Richter et al., 2012) required the participants to distinguish between words and equivalent non-words (e.g., "Mele") by pressing the corresponding button on the keyboard. All the words were nouns, with their length varying between three and ten letters and between one and three syllables. To manipulate item difficulty for words, word frequency and the number of orthographic neighbors were varied. Non-words were obtained by manipulating words, and the item difficulty of a non-word was affected by the similarity with a word. Words and non-words were matched with respect to length and frequency, with non-words being based on the frequency of the original words (for further details on item development, see Richter et al., 2012).

Sentence-level semantic integration. Similarly, the sentence verification task (Richter et al., 2012) required the participants to distinguish between true sentences and false sentences (e.g., "Snails are fast") by pressing the corresponding response button. To manipulate item difficulty

for sentences, for instance, the semantic complexity was varied, as it is represented by the number of propositions (for further details on item development, see Richter et al., 2012).

Reading comprehension. The reading comprehension task was taken from the PISA 2009 reading assessment (OECD, 2009). Two intact clusters with non-overlapping items were selected for the add-on study to the PISA 2012 main study in Germany, and were computerized for this purpose with the CBA ItemBuilder (Roelke, 2012). Each of the two clusters included four reading units. A unit consists of a reading text with three to five subsequent items. The unit texts differ in format (e.g., continuous and non-continuous text), type (e.g., description, narration, argumentation), and in reading situation (e.g., personal, public, educational). The items require explicit and implicit information from the unit text and can also require the student to reflect on a text. Response formats included multiple choice as well as free text response formats. Overall, the two clusters comprised a total of five polytomously scored items (partial-credit) and 32 dichotomously scored items. Each cluster was designed to take 30 minutes; after this time the respondents were required by the test administrator to continue to the next part of the assessment. As it is done in PISA, omitted item responses were coded as incorrect, whereas for not reached items, the response information was regarded as not available (i.e., coded as missing).

Note that for another study on the equivalence between computer-based and paper-based assessment, the reading clusters were administered on computer and paper in a randomized within- and between-subject balanced design. Given empirical evidence for construct equivalence across modes, equivalence of discrimination for all items and of difficulty for a large majority of items (Kroehne et al., 2019), indicators of reading comprehension were not distinguished by mode in the present analyses.

Booklet design. The national add-on study of PISA 2012 included several other measures that were irrelevant for the present study. To accommodate all the measurements, a booklet

design that included 16 rotations was implemented with random assignment of participants to booklets. At the beginning the reading comprehension task, all booklets comprised of either one reading cluster (eight rotations) or two reading clusters (eight rotations). In 12 rotations, the participants also completed the visual word recognition task and the sentence-level semantic integration task in both an untimed and a timed conditions.

Experimental design and manipulation check

The within-subject design included an untimed condition followed by a timed condition for visual word recognition and then the same for sentence-level semantic integration; timed and untimed conditions used different item material.

Experimental conditions. Each trial began with a 500-ms presentation of a centered fixation cross. When it disappeared, the stimulus was presented. For the untimed condition, respondents decided individually when to respond. After responding, a blank screen appeared for 500 ms (interstimulus interval).

In the timed condition (see Fig. 1), the stimulus disappeared once the predefined presentation time elapsed, which was indicated by the response signal (see Reed, 1973). For visual word recognition, the stimulus presentation time was 741ms, and for sentence-level semantic integration, it was 1,500 ms. The participants did not see a timer but were familiar with the amount of available through timed practice trials. The participants had 300 ms to give a response as soon as they heard the signal via earphone (a beep). Feedback on timing was provided in all timed trials. If the response was given in time, a happy face was presented for 800 ms; if the response was too early or too late, an unhappy face was presented for 1,200 ms with the message “too early” or “too late.” After presenting the feedback, a blank screen was shown for 500 ms.

For both visual word recognition and sentence-level semantic integration, items in the timed conditions with responses that were given before the onset of the stimulus were treated as not attempted items with a missing response and missing response time (2.57% of the expected total across items and respondents); there were no responses before the stimulus onset in the untimed conditions. In the timed conditions a small portion of the items showed omitted responses (3.76% of the expected total across items and respondents). For these items the response time was treated as not available (i.e., missing) and the response as incorrect. There were no omitted responses in the untimed condition.

Procedure. First, participants completed the visual word recognition test in the untimed condition, which consisted of 32 trials (16 words and 16 non-words). Then, they completed the visual word recognition test in a timed condition (again 16 words and 16 non-words). After that, the participants took the sentence-level semantic integration test in the untimed condition including 24 trials (12 true sentences, 12 false sentences), and finally, they completed the sentence-level semantic integration test in the timed condition (again 12 true sentences, 12 false sentences). In both reading component skill tasks, stimuli appeared in a random order, which was the same for all respondents. Also the order of conditions was the same for all participants. To avoid carryover effects by presenting a particular stimulus in both untimed and timed conditions, two parallel test forms were administered for both reading component skills tasks. The parallel forms are equalized in terms of stimulus properties used in the item design.

Each condition started with a block of practice trials to make participants familiar with the respective stimulus presentation and required method of responding. In the timed conditions, the participants learned in 12 practice trials to respond in time. The goal was to alter the response speed and the accuracy decision criterion needed to respond in time (Heitz, 2014). In the untimed conditions, the participants were instructed to work as quickly as possible and to avoid making

errors. In the timed condition, participants were required to press the response button of their choice once the response signal was presented and to give as many correct answers as possible.

Manipulation check. As a manipulation check, we inspected whether respondents could adapt their response speed in the timed condition as required by the response signal paradigm. The boxplots of item response times presented in Figure 2 show that for the visual word recognition test and the sentence-level semantic integration test, the majority of the responses were given within the response window, usually near the upper bound. For about only a third of the initial word recognition items were the whiskers clearly above and below the bounds of the response window. This indicates a learning effect that could be transferred to the sentence-level semantic integration test (except for the first item). For word recognition across all items, 66.11% of the observed responses were in time, 20.70% too early, and 13.20% too late. For sentence-level semantic integration, the results were similar; across all items 58.21% of the observations were in time, 17.35% too early, and 24.44% too late. The responses that were given before or after the response window were not excluded from data analysis, making our approach conservative.

For the timed condition a relative mild time pressure was intended which was empirically confirmed when comparing the response windows of the timed condition with the median response time of the untimed condition. The median of the median response time by item in the untimed condition was 1,266 ms for word recognition and 2,190 ms for sentence-level semantic integration. The 300-ms response window of the timed condition was 1,241 ms to 1,541 ms for word recognition and 2,000 ms to 2,300 ms for sentence-level semantic integration (note that the response times and the response window boundaries include the 500-ms presentation of the fixation cross). Thus, the respective response window included the median response time of the untimed condition.

Statistical Analyses

Modeling Approach. We employed a structural equation modelling approach that included measurement models with categorical (item responses) and continuous (log-transformed item response times) indicator variables and latent regression. The measurement models for effective ability and speed followed the hierarchical model of van der Linden (2007); however, the hierarchical part at the item-side was omitted, as proposed by Molenaar et al. (2015). For visual word recognition and sentence-level semantic integration, a logit link was used for item responses with equal loadings (one-parameter logistic model or 1-PL model), and a linear (identity) link for log-transformed item response times. For the more heterogeneous items assessing reading comprehension, a generalized partial credit model (GPCM) model was applied (as used in the recent PISA cycles).

The reliability of the ability measures was estimated as EAP reliability (Adams, 2005); standard errors were obtained via bootstrapping. EAP reliability is the ratio of the variance of theta estimates and the sum of the theta variance and the average of the posterior variances per theta score.

Statistical Software. For estimating measurement models and latent regression models Mplus version 8.2 was used (Muthén & Muthén, 1998-2017). We selected the (robust) maximum likelihood (MLR) estimator, since it can fully use the available information in the face of missing responses resulting from, for instance, the booklet design. Moreover, it can provide adjusted standard errors accounting for the non-independence of observations (i.e., students are nested into schools). However, we also used the Bayesian estimator (with probit link function) to estimate the more computationally demanding models, including residual correlations between response time and response accuracy within items. The minimum number of iterations was set to 25,000, and the Proportional Scale Reduction (PSR) factor used as the convergence criterion (Gelman &

Rubin, 1992). The first half of the iterations were considered as burn-in phase and discarded from parameter estimation.

For analyzing the reliability of ability measures and its standard errors, the R environment (R Core Team, 2016) with the packages TAM (Kiefer et al., 2016) and boot (Canty & Ripley, 2017) were used.

Results

Reliabilities, latent correlations, and residual dependencies

The reliability of the timed ability measure clearly exceeded the reliability of the corresponding untimed ability measure. For word recognition ability, the EAP reliability was .711 (*SE* = 0.016) for untimed and .850 (*SE* = 0.005) for timed, and for sentence-level semantic integration ability, it was .688 (*SE* = 0.014) for untimed and .811 (*SE* = 0.007) for timed (for reading comprehension, EAP reliability was .829, *SE* = 0.006).

Table 1 shows the latent correlations of untimed ability and speed measures, timed ability measures, and reading comprehension obtained from a multi-dimensional item response and response time model. The lower part shows the MLR estimates, and the upper part shows the Bayesian estimates (note that the latter model also included within-item residual correlations between response accuracy and response time indicators). The pattern of correlations from MLR estimation indicates that timed ability measures of word recognition and sentence-level semantic integration are more strongly correlated than the corresponding untimed ability measures. Most importantly, the correlations between reading comprehension with reading component skills are significantly stronger for the timed ability measures than for the untimed ones. As these are latent correlations, the difference in correlations between timed and untimed measures cannot merely be explained by differences in reliabilities. Correlations between untimed speed measures with reading comprehension were negligible. The latent speed-ability correlation for word recognition

was negative and of medium size, as it was for sentence-level semantic integration. Thus, in both reading component skill tasks, greater effective speed was associated with lower effective ability. The Bayesian estimation provided a very similar pattern of correlations.

To inspect residual correlations, we tested the hierarchical model separately for word recognition and sentence-level semantic integration (Bayesian estimation). Figure 3 shows the distribution for within-item residual correlations between response accuracy and response time by item type. For the word recognition test (upper panel), the residual correlations were positive for non-words and rather negative for words. That is, when judging non-words, taking more time on items than expected was associated with a greater probability of success, while when judging words, spending more time on items than expected was related to a lower probability of success. There was one outlier for non-words, which was the slightly misspelled German word for baby pig. This item was also very difficult as compared to the other non-words, suggesting that it was often understood as a word. For non-words the correlation between item difficulty and the residual correlation was negative (r = -.24); however, when excluding the outlier, it became clearly positive (r = .36). That is, in more difficult non-words, there was a greater benefit from extra time. For words, the correlation between item difficulty and the residual correlation was only small (r = .08). For the sentence-level semantic integration test (lower panel), the residual correlations were mostly positive for both true and false sentences. Notably, the variability of residual correlations was much greater for true sentences. For false sentences, the correlation between item difficulty and the residual correlation was positive (r = .48), and for true sentences, it was negative (r = -.38). That is, in more difficult “false” sentences there was a greater benefit from extra time whereas in more difficult “true” sentences, the size of the residual correlation was less positive and around zero.

Explaining timed by untimed measures

Table 2 shows how timed ability is explained by untimed ability and untimed speed for visual word recognition (models 1, 2, 3) and sentence-level semantic integration (models 4, 5, 6), respectively. Timed ability is first explained only by untimed ability (models 1, 4) or untimed speed (models 2, 5), and then jointly by both (models 3, 6) to obtain the effect of untimed ability when controlling for individual speed differences (and vice versa). Note that the standardized regression coefficients obtained from a model including residual correlations (Bayesian estimation) were very similar. As expected in Hypothesis 1.1, untimed ability and untimed speed only moderately explained timed ability. Regarding word recognition, the amount of explained variance in timed ability by untimed ability was $R^2 = .293$ (model 1), and by untimed speed it was negligible, $R^2 = .009$ (model 2). Regarding sentence-level semantic integration, the amount of explained variance in timed ability by untimed ability was moderate, $R^2 = .374$ (model 4), and by untimed speed, it was even zero, $R^2 = .000$ (model 5).

When using untimed ability and untimed speed jointly as predictors, the explanation could be substantially improved, although untimed speed did not show an effect in the simple regression. For word recognition, the proportion of explained variance increased to $R^2 = .437$ (model 3) and, for sentence-level semantic integration, to $R^2 = .517$ (model 6). Thus, Hypothesis 1.2 was clearly supported.

Notably, as shown in Table 2, the conditional effect of untimed ability when controlling for individual speed differences was significantly higher; for word recognition, the effect ranged from .542 to .732 and, for sentence-level semantic integration, from .612 to .855. The same was true for untimed speed, that is, when controlling for individual ability differences: The effect of untimed speed became significant and was significantly higher, from .096 to .424 for word

recognition and from .000 to .461 for sentence-level semantic integration. As shown above, untimed ability and speed are not independent but negatively correlated, that is, lower effective speed is associated with greater effective ability. Controlling for speed differences statistically means determining the effect of untimed ability for a constant level of speed, which substantially increases the explanation of timed ability (and vice versa).

Explaining sentence-level semantic integration by visual word recognition

Table 3 presents the results of the latent regressions of sentence-level semantic integration on visual word recognition. As expected in Hypothesis 2.1, sentence-level semantic integration ability was more (even twice as) strongly explained by word recognition ability in the timed condition (model 2: $R^2 = .683$) than in the untimed condition (model 1: $R^2 = .329$).

Next, we regressed timed sentence-level semantic integration ability separately on untimed word recognition ability and untimed word recognition speed. As assumed in Hypothesis 2.2, the explanation was only of moderate size when using untimed word recognition ability as a predictor (model 3: $R^2 = .293$) and very small when using untimed word recognition speed as a predictor (model 4: $R^2 = .006$).

Finally, we used untimed word recognition ability and untimed word recognition speed jointly as predictors. As shown in Table 3, the proportion of explained variance significantly increased by including both untimed measures (model 5: $R^2 = .423$) supporting Hypothesis 2.3. However, the proportion of explained variance was still lower than the proportion obtained with timed word recognition ability measure. Similarly for Hypothesis 1.2, we found that the conditional effects were larger than the effects from simple regression; for untimed ability, there was an increase from .541 to .726 and, for untimed speed from .075 to .406, making the effect significant.

Explanation of reading comprehension by visual word recognition and sentence-level semantic integration

Table 4 shows the findings of the latent regressions of reading comprehension on visual word recognition and sentence-level semantic integration. Hypothesis 3.1 was supported in that reading comprehension was explained by word recognition and sentence-level semantic integration ability more strongly if the reading component skill measures from the timed condition were used (model 1: $R^2 = .554$) instead of word recognition and sentence-level semantic integration ability (model 2: $R^2 = .361$) or word recognition and sentence-level semantic integration speed (model 3: $R^2 = .006$) from the untimed condition.

As expected in Hypothesis 3.2, the proportion of explained variance could be increased by exploiting all information from the untimed condition, that is, by using untimed ability and untimed speed variables jointly as predictors (model 4: $R^2 = .450$). Again, the effects of untimed ability conditional on speed proved to be stronger; for word recognition, the effect increased from .377 to .480 and, for sentence-level semantic integration, slightly from .300 to .383. The same was true for the effects of untimed speed, which also became significant; for word recognition the effect changed from .051 to .242 and, for sentence-level semantic integration, from -.089 to .163.

Finally, to address Hypothesis 3.3 we investigated the timed condition effects of word recognition and sentence-level semantic integration by controlling timed abilities for the ability and speed variables from the untimed condition. As shown in Table 3 (model 5), although the effects of timed abilities became a bit smaller (i.e., .346 for word recognition and .264 for sentence-level semantic integration) as compared with model 1, they were still significant, and those two mainly contributed to the explanation of reading. From the untimed measures, only the effect of word recognition ability remained substantial (.264), while the sentence-level semantic

integration ability and the speed variables show negligible effect sizes. Thus, Hypothesis 3.3 was supported.

Discussion

Main findings

The substantive goal of the present study was to investigate how well the efficiency in reading component skills explains reading comprehension. This was done by developing a new type of efficiency measurement of reading component skills with a timed condition at item level. Based on the assumption that the compromise between speed and accuracy differs between individuals in an untimed condition, we proposed a conceptual decomposition of timed and untimed measures, that is, *timed ability* = efficiency, *untimed ability* = *f*(efficiency, SAT), and *untimed speed* = *f*(efficiency, SAT). In line with these conceptual relationships, we could demonstrate empirically that timed ability measures of reading component skills proved to be strong and even better predictors of reading comprehension at sentence level and text level than the corresponding untimed ability and speed measures. That is, sentence-level semantic integration ability was explained by visual word recognition ability in the timed condition to a greater extent than in the untimed condition, and reading comprehension was more strongly explained by timed ability of word recognition and sentence-level semantic integration than by the corresponding untimed measures of ability and speed.

The explanation for the strong positive effects of reading component skills on reading comprehension is that efficiency in these sub-processes frees limited cognitive resources (e.g., attention, working memory capacity) to be invested in the more complex aspects of reading comprehension. Comprehending text is a dynamic process with interactions across multiple levels of processing. Reading successfully requires the reader to continuously process

information at the word and sentence levels, enabling the construction of a text model that is extended to a situation model by integrating further inferences based on prior knowledge and experience (Kintsch, 1998). As postulated by the verbal efficiency theory (Perfetti, 1985), not only accurate but also efficient lexical access and integration of word meanings are needed for successful text comprehension. Efficiency in this sense is measured in the proposed timed condition, while efficiency measures from the untimed condition suffer from ambiguity due to the SAT. Thus, by controlling response speed and by imposing a mild time pressure, timed measures of reading component skills better represent how these skills are required in action when actually reading a text.

The stronger correlations between timed ability (i.e., efficiency) measures of word recognition and sentence-level semantic integration as compared to the corresponding untimed ability measures give rise to the question of whether there is an irrelevant source of variance shared by the two timed measures which can explain the stronger correlation with reading comprehension, for instance, an ability to deal successfully with the time-limit procedure. However, if this were the case, we probably would not have found greater correlations with reading comprehension because the reading comprehension test can be considered as a power test. Although there was a time limit for each reading cluster in the reading comprehension test, we nevertheless assume that the test represents a power test and does not include a strong speed component for the following reasons. First, the reading comprehension framework (OECD, 2013) does not define speed as part of the reading comprehension construct. Second, and in line with this, the average proportion of not reached items was small for the German sample in particular (see OECD, 2014, p. 238). Therefore, we conclude that the stronger relationship of the timed ability measures to reading comprehension cannot be explained by the fact that the PISA reading comprehension assessment was completed under a cluster-level time limit.

Explaining reading comprehension of adolescents

The present study extends previous research explaining reading comprehension by reading component skills in several respects. Building on past research, we developed new timed measurements to tap efficiency at word-level and sentence-level reading and compared their predictive power with traditional untimed measurements. The present study then widened the scope of current research by investigating the role of reading component skills in a sample of adolescent readers, whereas previous studies have focused mainly on students from primary school or the beginning of secondary school.

In the models explaining reading comprehension the untimed ability measures showed stronger effects than the corresponding untimed speed measures. This might have been unexpected in that for advanced readers decoding efficiency may be reflected in speed rather than in accuracy differences because accuracy is high or even perfect. Our findings speak against this expectation. Instead, they are consistent with the results of a meta-analysis by García and Cain (2014) on the relationship between word decoding and reading comprehension, which indicate that, on average, word reading accuracy is more strongly correlated with reading comprehension than word reading speed. This was even suggested for the group of older readers (over 10 years old). Moreover, the importance of decoding fluency was revealed also for the initial phase of reading development by several studies (Florit & Cain, 2011; Kim et al., 2012; Silverman et al., 2013). The meta-analysis by Florit and Cain (2011) showed for languages with transparent orthography (i.e., with high correspondence between graphemes and phonemes, as in German language), that for younger readers (years of schooling 1–2) decoding fluency is more predictive of reading comprehension than decoding accuracy, whereas for advanced readers (years of schooling 3–5) the influence is comparable. From our point of view, untimed speed measures are deficient measures of efficiency because they ignore the SAT as discussed in the introduction.

Instead, timed ability measures are expected to be more adequate measures of efficiency. Previous research has mainly used global time limits at the test level to tap fluency (see the studies in the meta-analysis by Florit & Cain, 2011), whereas we employed time limits at item level to control speed for each single work output. In line with this reasoning, word recognition ability from the timed condition (representing efficiency) proved to be a much stronger predictor of reading comprehension than word recognition speed (and ability) from the untimed condition.

Another noteworthy finding is that although we examined advanced readers, the effect of untimed word recognition ability remained significant and substantial when controlling for untimed word recognition speed and timed word recognition ability. Nevertheless, this effect was clearly smaller than the one of timed word recognition ability, which we assume to represent word recognition efficiency. Finally, on the background of previous SVR-related research (e.g., Kim et al., 2012), it might be unexpected for the sample of adolescents that reading comprehension was more strongly predicted by visual word recognition than by sentence-level semantic integration, regardless of whether untimed and timed measures were compared. The medium correlation of untimed word recognition ability with reading comprehension obtained in the present study was in the range of what could have been expected for this age group (see García & Cain, 2014). However, the relative weak effect of sentence-level semantic integration could be related to the employed sentence verification task. It probably does not fully capture the complexity of language comprehension, which includes higher mental processes of parsing, bridging inferences, and deriving discourse interpretations (Hoover & Gough, 1990).

Efficiency versus ability and speed

Based on the assumption that speed and ability in the untimed condition together represent cognitive efficiency better than either of them separately, we also used both measures jointly as predictors for timed ability and for reading comprehension. As expected, when

regressing timed ability jointly on ability and speed from the corresponding untimed condition, the proportion of explained variance was significantly higher. Furthermore, when explaining both sentence-level semantic integration by word recognition and reading comprehension by word recognition and sentence-level semantic integration, the untimed measures approached the timed ability measures in terms of explanatory power when using the full information (both untimed ability and untimed speed; for an explanation see also the following section). Nevertheless, the final model with the effects of timed abilities and controlling for commonalities with the untimed measures clearly showed strong and unique effects of timed abilities on reading comprehension. This suggests that controlling speed by means of item-level time limits makes for a better measurement of cognitive efficiency than an untimed measurement, as it integrates information that is otherwise (partly) spread across the variables of untimed speed and untimed ability.

We considered effective (untimed) ability and speed as substantive latent variables explaining individual differences in response accuracy and response time. However, as discussed by De Boeck et al. (2017) alternative latent variables could be constituted in terms of *capacity* combining accuracy and speed (or drift rate from diffusion models, Ratcliff & Smith, 2004) and the *speed-accuracy balance* representing response cautiousness. These alternative latent variables are viewed as equivalent to (untimed) ability and speed in that they can be located in the same two-dimensional space. Assuming ability to be the horizontal axis and speed the vertical axis, capacity is found as clock-wise rotated speed axis between speed and ability, and balance as clock-wise rotated ability axis between ability and the negative pole of speed (see Fig. 1, De Boeck et al., 2017). Thus, capacity is positively related to both ability and speed, and balance is positively related to ability and negatively to speed. From this perspective, the proposed new timed ability measure can be understood as a measure of capacity since it combines accuracy and speed, whereas the traditional untimed ability measure is a function of both capacity and balance,

which is equivalent to *untimed ability* = *f*(efficiency, SAT). That is, respondents showing a larger capacity are able to show at a fixed speed level a higher effective ability and can, in turn, produce more accurate results than respondents showing lower capacity. By using experimental time limits, the balance latent variable (i.e., how much accuracy is favored over speed or response cautiousness) is not a source of observed individual differences anymore.

Conditional effects of untimed ability and untimed speed

When explaining timed ability measures representing efficiency (i.e., visual word recognition, sentence-level semantic integration) or reading comprehension, we used ability and speed from the untimed condition. It was consistently shown that untimed speed positively and incrementally explains the dependent variable above and beyond untimed ability (and vice versa). That is, respondents who were able to show both high effective ability and high effective speed in the untimed condition were most successful in the timed condition and also in the reading comprehension task.

Most interestingly, when controlling statistically for untimed speed in all regression models, the explanatory power of untimed ability was clearly higher; this was even more pronounced for the explanatory power of untimed speed controlling for untimed ability. Thus, if only untimed ability or speed was included as a predictor without the other (untimed) measure, the effect of untimed ability or speed was underestimated because it was suppressed by the other (untimed speed or ability) measure.

The obtained empirical findings suggest a reciprocal suppression (Conger, 1974). That is, the two predictors—untimed ability and untimed speed—are negatively correlated; they both show a positive correlation with the criterion (which, however, was less clear for untimed speed), and the standardized regression weights for both predictors are significantly higher in the multiple regression compared to the simple regression. In line with the proposed conceptual

decomposition of untimed measures, we assume that the explanation by untimed ability is higher because untimed speed serves as a suppressor variable, removing irrelevant variance due to the SAT and vice versa. Put differently, the effect of untimed ability (or speed, respectively) is estimated with individual differences related to untimed speed (or ability) being removed. Note that the (negative) correlation between the predictors of untimed ability and speed is a correlation at the between-person level. This means that it does not reflect the SAT, which is a within-person relation obtained across multiple speed conditions. Although the SAT is a within-person phenomenon, individual differences in the speed-accuracy balance (response cautiousness) can lead to correlations between effective speed and ability.

Taken together, if untimed speed and untimed ability jointly explain the respective criteria, the information about individual differences captured by timed and untimed measures becomes more similar. However, the obtained result pattern does not suggest that the experimental control and statistical control of speed converge completely, with respect to the amount of explained variance.

Construct interpretation of efficiency measures

From a validation perspective the construct interpretation of a test score is threatened by construct-irrelevant sources of variance (AERA et al., 2014). If effective ability from the untimed condition is used as a measure of efficiency, a potentially confounding variable here would be the decision on the speed-accuracy balance. Individual differences in this decision would compromise the interpretation (and use) of effective ability as an efficiency score, as discussed in the introduction section. Controlling speed experimentally aims at removing this confounding factor by task design. Furthermore, the pattern of empirical relations obtained in the present study provides strong convergent evidence for a (more) valid construct interpretation of timed ability measures compared to untimed measures. Specifically, we could demonstrate that the relationship

to other variables (word recognition to sentence-level semantic integration as well as word recognition and sentence-level semantic integration to reading comprehension) was stronger for the timed ability measures than for the untimed measures including both speed and ability, as hypothesized. Related to this, and given the consistent pattern of results, there was no evidence that a new confound was introduced by using item-level time limits (e.g., as discussed above the ability to deal successfully with the time-limit procedure).

Regarding the measurement of efficiency, using traditional approaches (see option 1, described in the introduction), we conclude that considering only untimed ability or untimed speed is insufficient, even if the importance of each of the two variables may differ depending on the developmental phase, for instance, in reading. Thus, when measuring reading component skills using an untimed condition, both speed and ability should be considered and included as predictors in explanatory models, in that they can mutually remove irrelevant variance suppressing their effects. However, regarding the assessment of individual differences, pairs of speed and ability scores are difficult to interpret and to compare across individuals. Therefore, we recommend using measurements with time limits at item level, providing a single score with a clear interpretation in terms of efficiency, which is also reflected in the present study by the superior prediction of reading comprehension through item-level time limits. Timed conditions require the choice of time limits, which is ideally based on response time distributions available from an untimed condition or at least based on some small-scale piloting. The technical implementation can be done, for instance, by using software for the design of psychological experiments as it allows for an exact control of presentation times and recording of response times.

Limitations and outlook

The timed condition was implemented in a way to avoid confounding sources of variance (e.g., test anxiety, perceived strain); the procedure was transparent by presenting practice trials, and the time limit was defined to induce only moderate speededness. Nevertheless, it would be relevant to show empirically that the timed measure is not more confounded than the untimed measure. This could be done by comparing the correlation of untimed and timed measures with self-reports about how the assessment was experienced (e.g., affective or motivational state variables).

In the present study the order of conditions and the presented item material were fixed. Thus, the experimental design could be further improved by balancing the position of the tests, the position of conditions, and the test content. Although the experimental timed conditions were carefully designed and implemented, some respondents may not have changed or were unable to change their decision criterion in the timed condition. In particular, we cannot distinguish from the data between individuals who responded too early and those who responded on time after having waited for the signal. Therefore, response speed is not perfectly controlled. Another limitation is that, given the lack of more specific hypotheses on the shape of the relationships, we only tested linear effects. Thus, in the case of non-linear effects in the true regression, we have only approximated the effects in a linear way.

To further evaluate the approach of measuring cognitive efficiency by means of item-level time limits, the generalizability of the findings needs to be investigated. This does not apply only to other efficiency constructs in the domain of reading (e.g., phonological recoding, word meaning activation) and other domains (e.g., perceptual speed) but also to other target populations as long as the efficiency concept applies (e.g., children, elderly people). The relation of reading component skills to reading comprehension has been investigated in numerous studies

usually with younger participants and by means of observed variables based of traditional measurements where individual differences in the SAT are typically ignored. Therefore, it would be interesting to investigate whether previous result patterns can be replicated when using measurements with item-level time limits to avoid confounding with SAT differences. Given the present study, one could expect that the effects of reading component skills become stronger.

The experimental condition aimed at controlling individual differences in response speed by implementing medium-fast time limits at item level. Future studies should add a number of more liberal or strict speed conditions to investigate how this affects the findings from the regression models. In a similar vein, Kendall (1964) explored the predictive validity of an intelligence test and showed that for test-level time limits ranging from 15 to 30 minutes, the medium limit of 22 minutes instead of the most liberal time limit of 30 minutes yielded the highest correlation with the criterion (see also Baxter, 1941).

The criterion measure (i.e., reading comprehension in the present study) also needs further consideration in future research, as it may also be affected by individual differences in the SAT. In a first step, this could be done by taking estimated speed differences into account. Moreover, to ensure that respondents take their time needed, one could consider providing feedback to the respondent to support individual time management (Goldhammer, 2015).

Finally, it would be valuable to challenge the present approach by comparing it with alternative approaches of measuring efficiency without controlling response speed. For instance, the diffusion model (Ratcliff & Smith, 2004) can separate efficiency of responding from response caution (Schmitz & Wilhelm, 2015). The drift-rate parameter describes the amount of evidence accumulated over time and, thus, individual differences in the efficiency of information processing. However, the diffusion model seems only appropriate for very simple cognitive tasks solved by continuous information accumulation (e.g., processes of visual word recognition, but

not necessarily of sentence-level semantic integration), and it is unclear whether it works for cognitive tests in general (De Boeck & Jeon, 2019). In their research, van der Maas et al. (2011) have demonstrated that their adaptation of the diffusion model to a latent variable model is suitable for cognitive tests assessing simple abilities. Another relevant approach of measuring efficiency is to account for response times when scoring responses. The Signed Residual Time (SRT) scoring rule rewards fast correct responses and penalizes fast incorrect responses (Maris & van der Maas, 2012). Since the rule is transparent for the respondents, it motivates them to work at an optimal speed-accuracy balance and to respond quickly and correctly at the same time. Other unidimensional latent variable models of efficiency based on the scoring of correct response times, or models of automaticity based on the scoring of correct response times and responses, have been discussed recently and compared by Su and Davison (2019).

Thus, there are several remaining questions to be addressed in future research on the measurement of cognitive efficiency. Nevertheless, given the obtained clear-cut findings for the domain of reading, we conclude that measurements controlling response speed by item-level time limits enable a clearer interpretation of performance in terms of cognitive efficiency than traditional measurements where the speed-accuracy balance is up to the respondent. As a consequence, individual differences of efficiency in reading component skills, as assessed in the item-level timed condition, better inform us about how strongly lower level processes support or hamper higher level comprehension processes of reading.

References

Ackerman, P. L., & Beier, M. E. (2007). Further explorations of perceptual speed abilities in the context of assessment methods, cognitive abilities, and individual differences during skill acquisition. *Journal of Experimental Psychology: Applied, 13*(4), 249-272.

Adams, R. J. (2005, 2005/01/01). Reliability as a measurement design effect. *Studies in Educational Evaluation, 31*(2), 162-172. https://doi.org/10.1016/j.stueduc.2005.05.008

AERA, APA, NCME, & Joint Committee on Standards for Educational Psychological Testing. (2014). *Standards for educational and psychological testing*.

Ahmed, Y., Francis, D. J., York, M., Fletcher, J. M., Barnes, M., & Kulesz, P. (2016, 2016/01/01/). Validation of the direct and inferential mediation (DIME) model of reading comprehension in grades 7 through 12. *Contemporary Educational Psychology, 44-45*, 68-82. https://doi.org/https://doi.org/10.1016/j.cedpsych.2016.02.002

Balota, D. A., Cortese, M. J., Sergent-Marshall, S. D., Spieler, D. H., & Yap, M. (2004, Jun). Visual word recognition of single-syllable words. *Journal of Experimental Psychology: General, 133*(2), 283-316. https://doi.org/10.1037/0096-3445.133.2.283

Barnes, M. A., Ahmed, Y., Barth, A., & Francis, D. J. (2015). The Relation of Knowledge-Text Integration Processes and Reading Comprehension in 7th- to 12th-Grade Students.

Scientific Studies of Reading, 19(4), 253-272. https://doi.org/10.1080/10888438.2015.1022650

Baxter, B. (1941). An experimental analysis of the contributions of speed and level in an intelligence test. *Journal of Educational Psychology, 32*(4), 285-296. https://doi.org/10.1037/h0061115

Bolsinova, M., De Boeck, P., & Tijmstra, J. (2017, 2017/12/01). Modelling Conditional Dependence Between Response Time and Accuracy. *Psychometrika, 82*(4), 1126-1148. https://doi.org/10.1007/s11336-016-9537-6

Bolsinova, M., & Tijmstra, J. (2015). Can response speed be fixed experimentally, and does this lead to unconfounded measurement of ability? *Measurement: Interdisciplinary Research and Perspectives, 13*(3-4), 165-168. https://doi.org/10.1080/15366367.2015.1105080

Bolsinova, M., Tijmstra, J., Molenaar, D., & De Boeck, P. (2017). Conditional Dependence between Response Time and Accuracy: An Overview of its Possible Sources and Directions for Distinguishing between Them. *Frontiers in psychology, 8*, 202-202. https://doi.org/10.3389/fpsyg.2017.00202

Canty, A., & Ripley, B. (2017). *boot: Bootstrap R (S-Plus) Functions. R package version 1.3-20.* In https://cran.r-project.org/web/packages/boot/index.html

Carlson, R. A., Khoo, B. H., Yaure, R. G., & Schneider, W. (1990). Acquisition of a problem-solving skill: Levels of organization and use of working memory. *Journal of Experimental Psychology: General, 119*(2), 193-214. https://doi.org/10.1037/0096-3445.119.2.193

Conger, A. J. (1974, 1974/04/01). A Revised Definition for Suppressor Variables: a Guide To Their Identification and Interpretation. *Educational and Psychological Measurement, 34*(1), 35-46. https://doi.org/10.1177/001316447403400105

Cunningham, A., & Stanovich, K. (1990, 12/01). Assessing Print Exposure and Orthographic Processing Skill in Children: A Quick Measure of Reading Experience. *Journal of Educational Psychology, 82*, 733-740. https://doi.org/10.1037/0022-0663.82.4.733

De Boeck, P., Chen, H., & Davison, M. (2017). Spontaneous and imposed speed of cognitive test responses. *British Journal of Mathematical and Statistical Psychology, 70*(2), 225-237. https://doi.org/10.1111/bmsp.12094

De Boeck, P., & Jeon, M. (2019, 2019-February-06). An Overview of Models for Response Times and Processes in Cognitive Tests [Systematic Review]. *Frontiers in psychology, 10*(102). https://doi.org/10.3389/fpsyg.2019.00102

Droop, M., & Verhoeven, L. (2003). Language proficiency and reading ability in first- and second-language learners [Article]. *Reading Research Quarterly, 38*(1), 78-103. https://doi.org/10.1598/RRQ.38.1.4

Ehri, L. C. (2005a). Development of Sight Word Reading: Phases and Findings. In *The science of reading: A handbook.* (pp. 135-154). Blackwell Publishing. https://doi.org/10.1002/9780470757642.ch8

Ehri, L. C. (2005b). Learning to Read Words: Theory, Findings, and Issues. *Scientific Studies of Reading, 9*(2), 167-188. https://doi.org/10.1207/s1532799xssr0902_4

Florit, E., & Cain, K. (2011, 2011/12/01). The Simple View of Reading: Is It Valid for Different Types of Alphabetic Orthographies? *Educational Psychology Review, 23*(4), 553-576. https://doi.org/10.1007/s10648-011-9175-6

Foorman, B. R., Petscher, Y., & Herrera, S. (2018, 2018/04/01/). Unique and common effects of decoding and language factors in predicting reading comprehension in grades 1–10. *Learning and Individual Differences, 63*, 12-23. https://doi.org/https://doi.org/10.1016/j.lindif.2018.02.011

García, J. R., & Cain, K. (2014). Decoding and Reading Comprehension: A Meta-Analysis to Identify Which Reader and Assessment Characteristics Influence the Strength of the

Relationship in English. *Review of Educational Research, 84*(1), 74-111. www.jstor.org/stable/24434229

Gelman, A., & Rubin, D. B. (1992, 1992/11). Inference from Iterative Simulation Using Multiple Sequences. *Statistical Science, 7*(4), 457-472. https://doi.org/10.1214/ss/1177011136

Goldhammer, F. (2015). Measuring Ability, Speed, or Both? Challenges, Psychometric Solutions, and What Can Be Gained From Experimental Control. *Measurement: Interdisciplinary Research and Perspectives, 13*(3-4), 133-164. https://doi.org/10.1080/15366367.2015.1100020

Goldhammer, F., & Kroehne, U. (2014). Controlling Individuals' Time Spent on Task in Speeded Performance Measures: Experimental Time Limits, Posterior Time Limits, and Response Time Modeling. *Applied Psychological Measurement, 38*, 255–267. https://doi.org/10.1177/0146621613517164

Goldhammer, F., Naumann, J., Stelter, A., Tóth, K., Rölke, H., & Klieme, E. (2014). The Time on Task Effect in Reading and Problem Solving Is Moderated by Task Difficulty and Skill: Insights From a Computer-Based Large-Scale Assessment. *Journal of Educational Psychology, 106*, 608–626. https://doi.org/10.1037/a0034716

Goldhammer, F., Steinwascher, M. A., Kroehne, U., & Naumann, J. (2017). Modelling individual response time effects between and within experimental speed conditions: A GLMM

approach for speeded tests. *British Journal of Mathematical and Statistical Psychology, 70*(2), 238-256. https://doi.org/doi:10.1111/bmsp.12099

Graesser, A. C., Hoffman, N. L., & Clark, L. F. (1980, 1980/04/01/). Structural components of reading time. *Journal of Verbal Learning and Verbal Behavior, 19*(2), 135-151. https://doi.org/https://doi.org/10.1016/S0022-5371(80)90132-2

Gulliksen, H. (1950). *Theory of mental tests*. Wiley.

Hagtvet, B. E. (2003, 2003/09/01). Listening comprehension and reading comprehension in poor decoders: Evidence for the importance of syntactic and semantic skills as well as phonological skills. *Reading and Writing, 16*(6), 505-539. https://doi.org/10.1023/A:1025521722900

Heitz, R. P. (2014, 2014-June-11). The Speed-Accuracy Tradeoff: History, Physiology, Methodology, and Behavior [Review]. *Frontiers in Neuroscience, 8*. https://doi.org/10.3389/fnins.2014.00150

Hoffman, B. (2012, 2012/04/01/). Cognitive efficiency: A conceptual and methodological comparison. *Learning and Instruction, 22*(2), 133-144. https://doi.org/https://doi.org/10.1016/j.learninstruc.2011.09.001

Hoover, W. A., & Gough, P. B. (1990, 1990/06/01). The simple view of reading. *Reading and Writing, 2*(2), 127-160. https://doi.org/10.1007/BF00401799

Hoover, W. A., & Tunmer, W. E. (2018, 2018/09/01). The Simple View of Reading: Three Assessments of Its Adequacy. *Remedial and Special Education, 39*(5), 304-312. https://doi.org/10.1177/0741932518773154

Hunt, E. (1978). Mechanics of verbal ability. *Psychological Review, 85*(2), 109-130. https://doi.org/10.1037/0033-295X.85.2.109

Johnson, E. S., Pool, J. L., & Carter, D. R. (2011, 2011/12/01). Validity Evidence for the Test of Silent Reading Efficiency and Comprehension (TOSREC). *Assessment for Effective Intervention, 37*(1), 50-57. https://doi.org/10.1177/1534508411395556

Kendall, L. M. (1964). The Effects of Varying Time Limits on Test Validity. *Educational and Psychological Measurement, 24*(4), 789-800. https://doi.org/10.1177/001316446402400406

Kiefer, T., Robitzsch, A., & Wu, M. (2016). *TAM: Test Analysis Modules. R package version 1.99-6.* In http://CRAN.R-project.org/package=TAM

Kim, Y.-S., Wagner, R. K., & Lopez, D. (2012). Developmental relations between reading fluency and reading comprehension: a longitudinal study from Grade 1 to Grade 2. *J Exp Child Psychol, 113*(1), 93-111. https://doi.org/10.1016/j.jecp.2012.03.002

Kim, Y.-S. G. (2019). Hierarchical and dynamic relations of language and cognitive skills to reading comprehension: Testing the direct and indirect effects model of reading (DIER). *Journal of Educational Psychology*, No Pagination Specified-No Pagination Specified. https://doi.org/10.1037/edu0000407

Kim, Y. S. (2015, Jan-Feb). Language and cognitive predictors of text comprehension: evidence from multivariate analysis. *Child Dev, 86*(1), 128-144. https://doi.org/10.1111/cdev.12293

Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. Cambridge University Press.

Kintsch, W., & van Dijk, T. A. (1978). Toward a model of text comprehension and production. *Psychological Review, 85*(5), 363-394. https://doi.org/10.1037/0033-295X.85.5.363

Kroehne, U., Buerger, S., Hahnel, C., & Goldhammer, F. (2019). Construct Equivalence of PISA Reading Comprehension Measured With Paper-Based and Computer-Based Assessments. *Educational Measurement: Issues and Practice, 38*(3), 97-111. https://doi.org/10.1111/emip.12280

LaBerge, D., & Samuels, S. J. (1974, 1974/04/01/). Toward a theory of automatic information processing in reading. *Cogn Psychol, 6*(2), 293-323. https://doi.org/https://doi.org/10.1016/0010-0285(74)90015-2

Language and Reading Research Consortium, & Logan, J. (2017). Pressure points in reading comprehension: A quantile multiple regression analysis. *Journal of Educational Psychology, 109*(4), 451-464. https://doi.org/10.1037/edu0000150

Lohman, D. F. (1989, Sum 1989). Individual differences in errors and latencies on cognitive tasks. *Learning and Individual Differences, 1*(2), 179-202. https://doi.org/10.1016/1041-6080(89)90002-2 (Methodology and individual differences)

Lonigan, C. J., Burgess, S. R., & Schatschneider, C. (2018, 2018/09/01). Examining the Simple View of Reading With Elementary School Children: Still Simple After All These Years. *Remedial and Special Education, 39*(5), 260-273. https://doi.org/10.1177/0741932518764833

Luce, R. D. (1986). *Response times: Their roles in inferring elementary mental organization*. Oxford University Press.

Maris, G., & van der Maas, H. (2012). Speed-accuracy response models: Scoring rules based on response time and accuracy. *Psychometrika, 77*(4), 615-633. https://doi.org/10.1007/s11336-012-9288-y

Molenaar, D., Tuerlinckx, F., & van der Maas, H. L. (2015, May). A generalized linear factor model approach to the hierarchical framework for responses and response times. *British Journal of Mathematical and Statistical Psychology, 68*(2), 197-219. https://doi.org/10.1111/bmsp.12042

Muthén, L. K., & Muthén, B. O. (1998-2017). *Mplus User's Guide. Eighth Edition*. Muthén & Muthén.

OECD. (2009). *PISA 2009 Assessment Framework*. OECD Publishing. https://www.oecd.org/pisa/pisaproducts/44455820.pdf

OECD. (2013). *PISA 2012 Assessment and Analytical Framework*. OECD Publishing. https://doi.org/doi:https://doi.org/10.1787/9789264190511-en

OECD. (2014). *PISA 2012 Technical Report*. OECD Publishing.

Perfetti, C. A. (1985). *Reading ability*. Oxford University Press.

Perfetti, C. A. (2007). Reading ability: Lexical quality to comprehension. *Scientific Studies of Reading, 11*(4), 357-383. https://doi.org/10.1080/10888430701530730 (What should the scientific study of reading be now and in the near future?)

Perfetti, C. A., & Hart, L. (2002). The lexical quality hypothesis. In L. Verhoeven, C. Elbro, & P. Reitsma (Eds.), *Precursors of functional literacy* (pp. 189-213). John Benjamin.

Perfetti, C. A., & Stafura, J. (2013). Word Knowledge in a Theory of Reading Comprehension. *Scientific Studies of Reading, 18*(1), 22-37. https://doi.org/10.1080/10888438.2013.827687

R Core Team. (2016). *R: A language and environment for statistical computing.* In (Version 3.1.3) R Foundation for Statistical Computing. http://www.R-project.org/

Ratcliff, R., & Smith, P. L. (2004). A Comparison of Sequential Sampling Models for Two-Choice Reaction Time. *Psychological Review, 111*(2), 333-367. https://doi.org/10.1037/0033-295X.111.2.333

Reed, A. V. (1973). Speed-accuracy trade-off in recognition memory. *Science, 181*(4099), 574-576. https://doi.org/10.1126/science.181.4099.574

Richter, T., Isberner, M.-B., Naumann, J., & Kutzner, Y. (2012). Prozessbezogene Diagnostik von Lesefähigkeiten bei Grundschulkindern [Process-based measurement of reading skills in primary school children]. *Zeitschrift für Pädagogische Psychologie, 26*, 313-313. https://doi.org/10.1024/1010-0652/a000079

Richter, T., Isberner, M.-B., Naumann, J., & Neeb, Y. (2013, 03/22). Lexical Quality and Reading Comprehension in Primary School Children. *Scientific Studies of Reading, 17*. https://doi.org/10.1080/10888438.2013.764879

Roelke, H. (2012). The ItemBuilder: A Graphical Authoring System for Complex Item Development. World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education, Montréal, Quebec, Canada.

Salthouse, T. A. (1996, Jul). The processing-speed theory of adult age differences in cognition. *Psychol Rev, 103*(3), 403-428.

Salthouse, T. A., & Hedden, T. (2002). Interpreting reaction time measures in between-group comparisons. *Journal of Clinical and Experimental Neuropsychology, 24*(7), 858-872. https://doi.org/10.1076/jcen.24.7.858.8392 (Contributions of cognitive neuroscience to clinical neuropsychology: The role of reaction time-based studies)

Samuels, S. J., & Flor, R. F. (1997, 1997/04/01). The importance of automaticity for developing expertise in reading. *Reading & Writing Quarterly, 13*(2), 107-121. https://doi.org/10.1080/1057356970130202

Schmitz, F., & Wilhelm, O. (2015, 2015/10/02). Item-Level Time Limits Are Not a Panacea. *Measurement: Interdisciplinary Research and Perspectives, 13*(3-4), 182-185. https://doi.org/10.1080/15366367.2015.1115300

Silverman, R. D., Speece, D. L., Harring, J. R., & Ritchey, K. D. (2013). Fluency has a role in the simple view of reading. *Scientific Studies of Reading, 17*(2), 108-133. https://doi.org/10.1080/10888438.2011.618153

Su, S., & Davison, M. L. (2019, 2019/04/03). Improving the Predictive Validity of Reading Comprehension Using Response Times of Correct Item Responses. *Applied Measurement in Education, 32*(2), 166-182. https://doi.org/10.1080/08957347.2019.1577247

van der Linden, W. J. (2007). A Hierarchical Framework for Modeling Speed and Accuracy on Test Items. *Psychometrika, 72*(3), 287-308. https://doi.org/10.1007/s11336-006-1478-z

van der Linden, W. J. (2009). Conceptual Issues in Response-Time Modeling. *Journal of Educational Measurement, 46*(3), 247–272. https://doi.org/10.1111/j.1745-3984.2009.00080.x

van der Maas, H. L. J., Molenaar, D., Maris, G., Kievit, R. A., & Borsboom, D. (2011). Cognitive psychology meets psychometric theory: On the relation between process models for decision making and latent variable models for individual differences. *Psychological Review, 118*(2), 339-356. https://doi.org/10.1037/a0022749

Vellutino, F. R., Tunmer, W. E., Jaccard, J. J., & Chen, R. (2007, 2007/01/01). Components of Reading Ability: Multivariate Evidence for a Convergent Skills Model of Reading Development. *Scientific Studies of Reading, 11*(1), 3-32. https://doi.org/10.1080/10888430709336632

Wagner, R. K., Torgesen, J., Rashotte, C. A., & Pearson, N. (2010). *Test of sentence reading efficiency and comprehension*. Pro-Ed.

Walczyk, J. J. (2000). The interplay between automatic and control processes in reading. *Reading Research Quarterly, 35*(4), 554-566. https://doi.org/10.1598/RRQ.35.4.7

Walczyk, J. J., Wei, M., Grifith-Ross, D. A., Goubert, S. E., Cooper, A. L., & Zha, P. (2007). Development of the interplay between automatic processes and cognitive resources in reading. *Journal of Educational Psychology, 99*(4), 867-887. https://doi.org/10.1037/0022-0663.99.4.867

Wickelgren, W. A. (1977). Speed-accuracy tradeoff and information processing dynamics. *Acta Psychologica, 41*, 67-85. https://doi.org/10.1016/0001-6918(77)90012-9

Table 1

Latent correlations of untimed ability and speed measures, timed ability measures, and reading comprehension.

Variable		1.	2.	3.	4.	5.	6.	7.
1. Word recognition ability	untimed	1.00	-.421 (0.039)	.522 (0.039)	.566 (0.042)	-.251 (0.043)	.520 (0.040)	.539 (0.040)
2. Word recognition speed	untimed	-.445 (0.045)	1.00	.097 (0.042)	-.227 (0.044)	.460 (0.033)	.085 (0.044)	.011 (0.046)
3. Word recognition ability	timed	.522 (0.047)	.105 (0.051)	1.00	.523 (0.041)	.008 (0.043)	.811 (0.021)	.711 (0.027)
4. Semantic integration ability	untimed	.655 (0.036)	-.207 (0.057)	.622 (0.039)	1.00	-.536 (0.036)	.574 (0.040)	.501 (0.043)
5. Semantic integration speed	untimed	-.259 (0.057)	.475 (0.047)	.001 (0.051)	-.502 (0.057)	1.00	-.001 (0.044)	-.061 (0.045)
6. Semantic integration ability	timed	.540 (0.040)	.078 (0.047)	.836 (0.022)	.638 (0.037)	-.001 (0.056)	1.00	.680 (0.030)
7. Reading comprehension		.565 (0.033)	.010 (0.054)	.723 (0.027)	.565 (0.039)	-.058 (0.052)	.712 (0.029)	1.00

Note. Values below the diagonal are MLR estimates (with standard error in brackets), and values above the diagonal are Bayesian estimates (with posterior standard deviation in brackets).

Table 2

Latent regression of timed on untimed condition for word recognition and sentence-level semantic integration.

Model	Criterion		Predictors		$\beta_{j.MLR}$ (SE)	R^2 (SE)	$\beta_{j.Bayes}$ $(Posterior\ SD)$
1	Word recognition ability	timed	Word recognition ability	untimed	.542*** (0.042)	.293 (0.045)	.538 (0.039)
2	Word recognition ability	timed	Word recognition speed	untimed	.096 (0.051)	.009 (0.010)	.095 (0.043)
3	Word recognition ability	timed	Word recognition ability	untimed	.732*** (0.049)	.437 (0.047)	.718 (0.041)
			Word recognition speed	untimed	.424*** (0.044)		.421 (0.042)
4	Semantic integration ability	timed	Semantic integration ability	untimed	. 612*** (0.038)	.374 (0.047)	. 599 (0.040)
5	Semantic integration ability	timed	Semantic integration speed	untimed	.000 (0.060)	.000 (0.000)	.001 (0.044)
6	Semantic integration ability	timed	Semantic integration ability	untimed	.855*** (0.054)	.517 (0.054)	.837 (0.052)
			Semantic integration speed	untimed	.461*** (0.049)		.462 (0.053)

Note. * $p < .05$, ** $p < .01$, *** $p < .001$; *SE* = Standard Error; *SD* = Standard Deviation; all regression coefficients are standardized.

Table 3

Latent regression of sentence-level semantic integration on word recognition.

Model	Criterion		Predictors		$\beta_{j.MLR}$ (SE)	R^2 (SE)	$\beta_{j.Bayes}$ $(Posterior\ SD)$
1	Semantic integration ability	untimed	Word recognition ability	untimed	.574*** (0.048)	.329 (0.055)	.569 (0.041)
2	Semantic integration ability	timed	Word recognition ability	timed	.826*** (0.028)	.683 (0.046)	.826 (0.021)
3	Semantic integration ability	timed	Word recognition ability	untimed	.541*** (0.040)	.293 (0.043)	.536 (0.040)
4	Semantic integration ability	timed	Word recognition speed	untimed	.075 (0.049)	.006 (0.007)	.076 (0.044)
5	Semantic integration ability	timed	Word recognition ability	untimed	.726*** (0.046)	.423 (0.048)	.711 (0.043)
			Word recognition speed	untimed	.406*** (0.049)		.403 (0.045)

Note. * $p < .05$, ** $p < .01$, *** $p < .001$; *SE* = Standard Error; *SD* = Standard Deviation; all regression coefficients are standardized.

Table 4

Latent regression of reading comprehension on word recognition and sentence-level semantic integration.

Model	Criterion	Predictors		$\beta_{j.MLR}$ (SE)	R^2 (SE)	$\beta_{j.Bayes}$ $(Posterior\ SD)$
1	Reading comprehension	Word recognition ability	timed	.476*** (0.075)	.554 (0.037)	.466 (0.074)
		Semantic integration ability	timed	.302*** (0.079)		.306 (0.077)
2	Reading comprehension	Word recognition ability	untimed	.377*** (0.057)	.361 (0.037)	.379 (0.059)
		Semantic integration ability	untimed	.300*** (0.055)		.292 (0.059)
3	Reading comprehension	Word recognition speed	untimed	.051 (0.057)	.006 (0.008)	.053 (0.053)
		Semantic integration speed	untimed	-.089 (0.055)		-.092 (0.052)
4	Reading comprehension	Word recognition ability	untimed	.480*** (0.078)	.450 (0.044)	.480 (0.070)
		Semantic integration ability	untimed	.383*** (0.085)		.373 (0.082)
		Word recognition speed	untimed	.242*** (0.064)		.244 (0.057)
		Semantic integration speed	untimed	.163* (0.065)		.159 (0.064)
5	Reading comprehension	Word recognition ability	untimed	.264** (0.079)	. 597 (0.036)	.226 (0.072)
		Semantic integration ability	untimed	.014 (0.124)		.059 (0.093)

	Word recognition speed	untimed	.084 (0.067)	.078 (0.054)
	Semantic integration speed	untimed	-.022 (0.079)	-.009 (0.062)
	Word recognition ability	timed	.346*** (0.085)	.384 (0.080)
	Semantic integration ability	timed	.264** (0.098)	.207 (0.091)

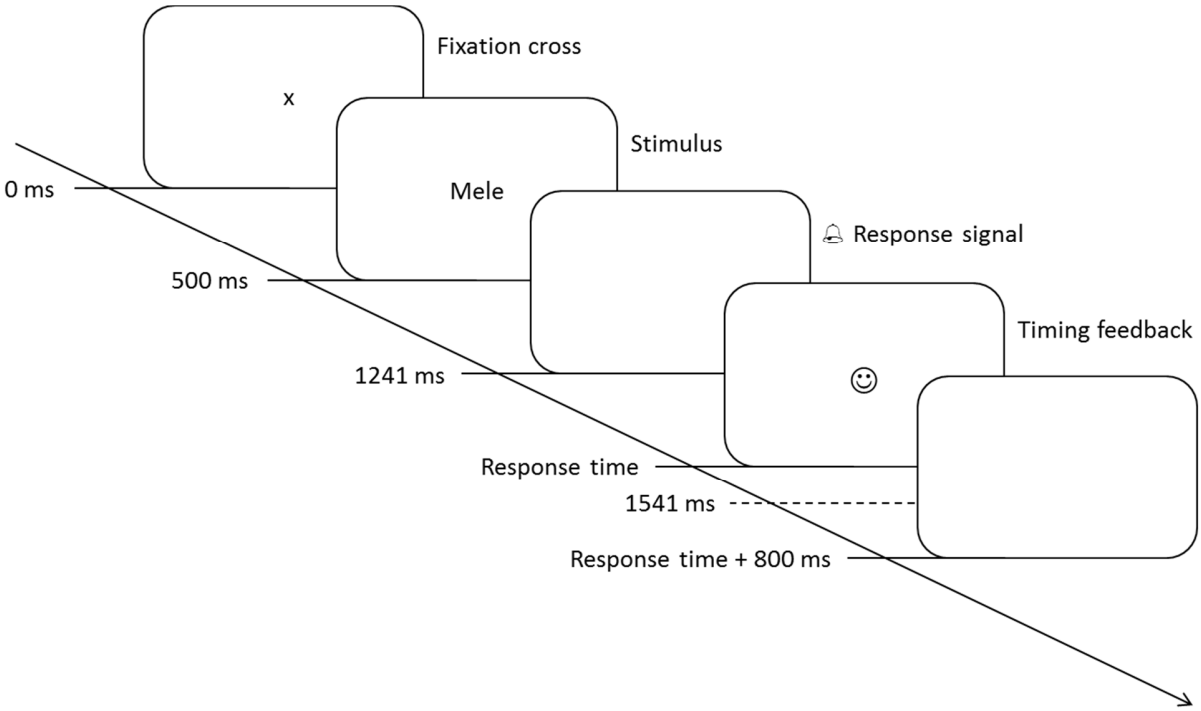
Note. * $p < .05$, ** $p < .01$, *** $p < .001$; *SE* = Standard Error; *SD* = Standard Deviation; all regression coefficients are standardized.

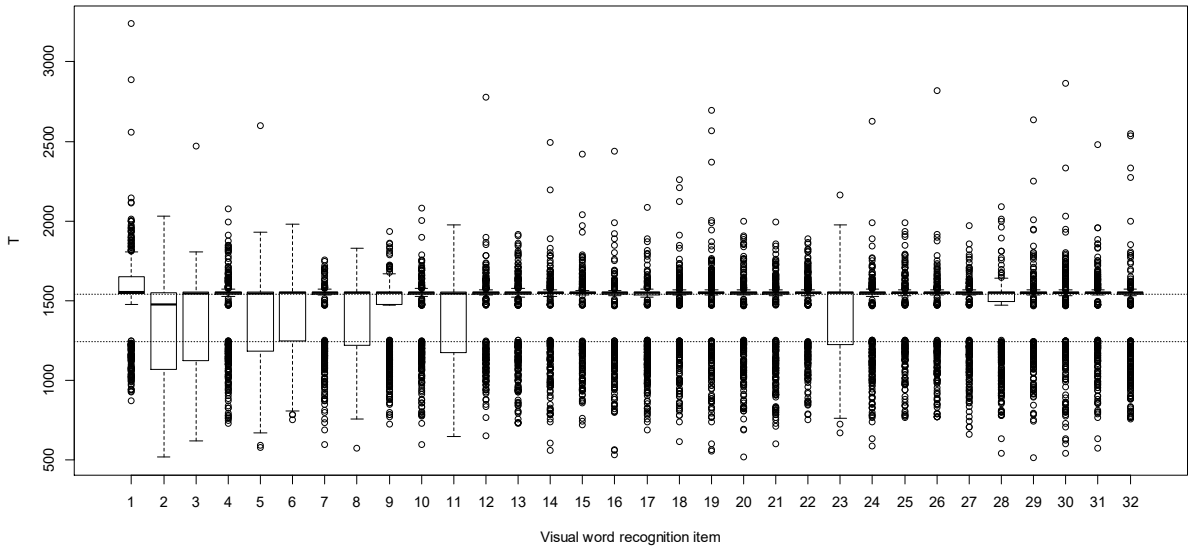
Figure Captions

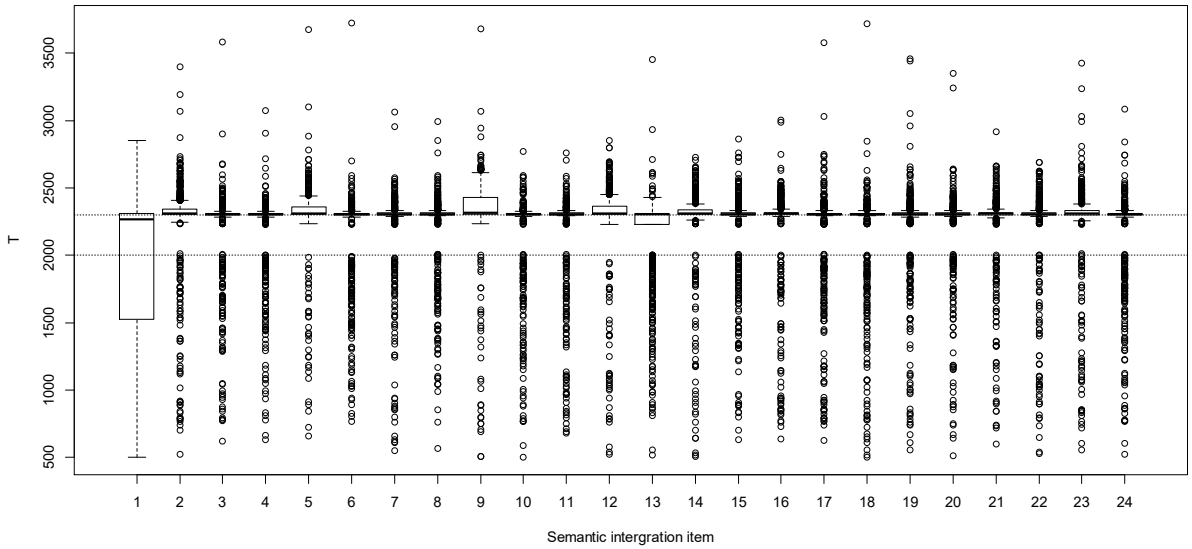
Figure 1. Trial of the word recognition task in the timed condition with a stimulus presentation time of 741ms. In the example the response was given after the response signal and before the 300-ms response window ended (dashed line). Therefore, the feedback was positive (smiling face).

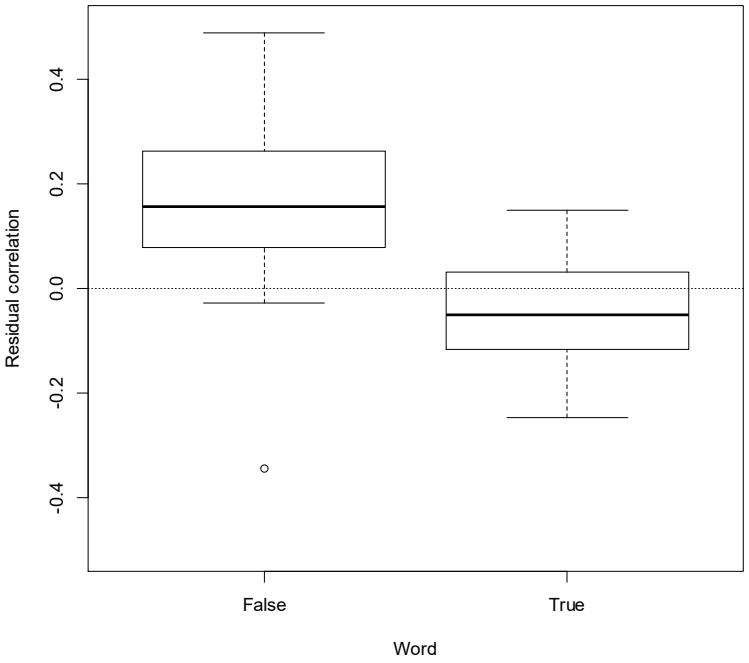
Figure 2. Boxplots of response time by item observed in the timed condition for the word recognition test (upper panel) and the sentence-level semantic integration test (lower panel). The dashed horizontal lines indicate the 300-ms response window beginning when the stimulus disappeared, and the response signal was presented, respectively.

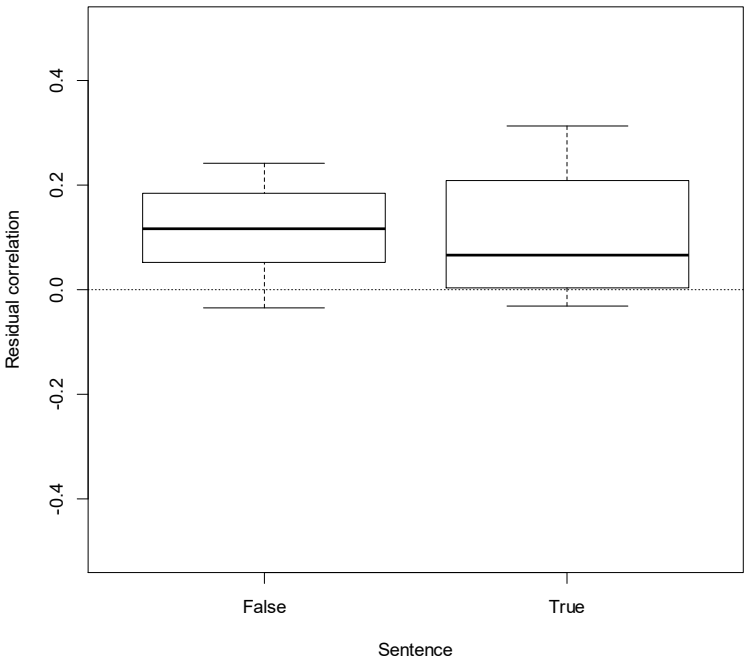
Figure 3. Boxplots of within-item residual correlations between response accuracy and response time by item type for the word recognition test (upper panel) and the sentence-level semantic integration test (lower panel). The item type “true” means that a word and a correct sentence, respectively, are evaluated, whereas “false” means that a non-word and an incorrect sentence, respectively, are evaluated.


Fixation cross
x
Stimulus
Mele
0 ms
Response signal
500 ms
Timing feedback
1241 ms
Response time
1541 ms
Response time + 800 ms


T
500
1000
1500
2000
2500
3000
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32
Visual word recognition item


T
500
1000
1500
2000
2500
3000
3500
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24
Semantic intergration item


Residual correlation
0.4
0.2
0.0
-0.2
-0.4
False
True
Word


Residual correlation
0.4
0.2
0.0
-0.2
-0.4
False
True
Sentence