Listening Comprehension Revisited: the Relationship between Lexical Predictors and Measures of Listening Comprehension

Vivian G. Trip

Faculty of Behavioural and Social Sciences, University of Groningen

Masterthesis Orthopedagogiek, PAMA5166

dr. Evelien M. B. Mulder and Erica Kamphorst, MSc.

prof. dr. Monika Smit (second assessor)

August 1, 2022

Abstract

The COVID-19 pandemic negatively impacted students' latest learning outcomes in Dutch primary schools. Reading comprehension (RC) had the largest learning outcome loss (-25%) (Haelermans et al., 2021). Several studies suggest that word decoding, vocabulary and listening comprehension (LC) are predictors in RC development (de Jong & van der Leij, 2002; Hoover & Gough, 1990; Verhoeven & Van Leeuwe, 2008). Apart from the input, the cognitive processes in RC and LC are similar (Hogan et al., 2011; Perfetti et al., 2005; Van den Broek, 2009). The present study examines whether there is a relationship between reading fluency and vocabulary as lexical predictors of measures of listening comprehension. The participants in this study were 47 students between the ages of 9 and 11 from two Dutch primary schools. In two classroom sessions and two individual sessions language tests were administered. Reading fluency was measured by the EMT (reading words aloud) and De Klepel (reading pseudowords) and the vocabulary has been assessed with a modified version of the PPVT-III-NL. LC has been tested with new measures; by multiple choice questions and by making a situation model. The situation models of the students were compared with the models of nine skilled adults. Multiple regression analyses showed that reading fluency was a significant predictor of the LC questions; vocabulary was not. For the LC situation model, both reading fluency and vocabulary had no predictive value. For future research, it is recommended to analyse the situation model in other ways.

Keywords: Listening Comprehension, Situation Model, Reading Fluency, Vocabulary

Samenvatting

De COVID-19 pandemie heeft een negatieve invloed gehad op de leerresultaten van Nederlandse basisschoolkinderen. Begrijpend lezen (RC) had het grootste verlies aan leerresultaten (-25%) (Haelermans et al., 2021). Verschillende onderzoeken suggereren dat woorddecodering, woordenschat en luistervaardigheid (LC) voorspellers zijn in de ontwikkeling van RC (de Jong & van der Leij, 2002; Hoover & Gough, 1990; Verhoeven & Van Leeuwe, 2008). De cognitieve processen in RC en LC zijn, afgezien van de input, vergelijkbaar (Hogan et al., 2011; Perfetti et al., 2005; Van den Broek, 2009). Het huidige onderzoek onderzoekt of er een verband is tussen leesvaardigheid en woordenschat als lexicale voorspellers van metingen van luistervaardigheid. De participanten van dit onderzoek waren 47 leerlingen in de leeftijd van 9 tot 11 jaar van twee Nederlandse basisscholen. In twee klassikale sessies en twee individuele sessies werden er taaltesten afgenomen. Leesvaardigheid is gemeten met de EMT (woorden voorlezen) en De Klepel (pseudowoorden voorlezen) en de woordenschat is beoordeeld met een aangepaste versie van de PPVT-III-NL. LC is getest met nieuwe, aangepaste testen; middels meerkeuzevragen en middels het maken van een situatiemodel. De situatiemodellen van de leerlingen werden vergeleken met de modellen van negen geschoolde volwassenen. Meervoudige regressieanalyses lieten zien dat leesvaardigheid een significante voorspeller is van de LC-meerkeuzevragen; woordenschat was dat niet. Voor het LC-situatiemodel hadden zowel leesvaardigheid als woordenschat geen voorspellende waarde. Voor toekomstig onderzoek is het aanbevolen het situatiemodel op andere manieren te analyseren.

Trefwoorden: Begrijpend luisteren, Situatiemodel, Leesvaardigheid, Woordenschat.

Listening Comprehension Revisited: the Relationship between Lexical Predictors and Measures of Listening Comprehension

Reading comprehension (RC) is an important prerequisite for learning in all subjects in school (Berends, 2011; de Jong, 2011; SLO, 2018) and it is also one of the important basic skills a student needs to participate properly in today's 21st-century society (Inspectie van het Onderwijs, 2022). Various studies indicate that a conditional and necessary language skill to learn and develop good RC is listening comprehension (LC) (Hogan et al., 2011; Mommers, 2003; Potocki et al., 2013). In the lower grades of primary schools there is considerable attention for developing the skills of LC through (interactive) reading by the teachers. In the upper grades, the focus in teaching shifts from LC to RC. Listening assignments and tests often only return in secondary education in the learning of foreign languages (Dunkel, 1991). The COVID-19 pandemic and school closures in school year 2020-2021 have had a negative impact on students' cognitive performances. Several studies have analysed the latest achievements of students in 4th to 7th grade of Dutch primary education (Engzell et al., 2020; Haelermans et al., 2021; Meshcheriakova et al., 2020) and the results suggest that learning outcomes in all grades were lower than before COVID-19. The loss in learning growth was the most profound in the students' RC skill, scoring 25% less than pre-COVID-19 (Haelermans et al., 2021), indicating that education is essential in the development of reading. The "Simple View of Reading" (Gough & Tunmer, 1986; Hoover & Gough, 1990) proposes that RC consists of two primary factors: word recognition or decoding and linguistic comprehension. Linguistic comprehension includes the concepts vocabulary and LC. According to this model, it is difficult to develop sufficient RC if a student is proficient in only one of the skills (Hoover & Gough, 1990). To improve RC skills in students, and specifically to try to compensate for the loss of learning growth due to COVID-19, it is necessary to evaluate and improve the current LC education in upper grades of primary schools. The goal of the current study was to explain, test and predict two new measures of LC in Dutch primary education.

Listening skills are the first language skills children acquire. After a few years, the development of other language skills, such as speaking, reading, and writing, follows (Heuvelman & Schreiner, 2010). Being able to hear well or being able to listen well are two different skills. The development of 'hearing' is a physical developmental process, while that of listening is a learned skill. Training LC can be started at a very young age, (picture) books can be read by parents to children from birth onwards. In pre-schoolers, teachers can stimulate listening skills in different ways, for example by asking questions at different levels, such as: “What animal is the story about?”, “Where is …?”, “Why did that happen?”, or “What would you do?”. To answer these questions, children need to listen actively and contribute ideas about a story or an informational text, give meaning to it and think about its content. In this process, children link their existing (vocabulary) knowledge to new information and thus expand their (vocabulary) knowledge further (Heuvelman & Schreiner, 2020). Of course, this is applied at an increasingly higher level and more independently in the higher grades of primary schools.

Listening comprehension can be defined as giving meaning to spoken language. It is an active and complex process for which skills such as vocabulary, prior knowledge, and the correct use of listening strategies are essential (Ahlers & Van de Mortel, 2009). For LC, an integral appeal is made to several essential language and thinking skills. Multiple studies show that there is a clear relationship between the processes of reading and listening (Danks & End, 1987; Kintsch & Kozminsky, 1977; Sinatra, 1990; Sticht et al., 1974; Sticht & James, 1984). In both reading and listening, largely the same cognitive processing mechanisms are involved; 1) perception of speech/decoding of written text, 2) giving meaning to the individual words (lexical access), 3) integration of sentences or phrases, 4) text comprehension and 5) evaluating and regulating understanding (Hogan et al., 2011; Perfetti et al., 2005; Van den Broek, 2009). Information enters through the auditory or visual memory and is stored briefly in the short-term memory. In the long-term memory, incoming information is combined, and the associated meaning is activated (Carpenter & Just, 2013; Cowan, 1996; Kidd, 2013). To improve higher cognitive processes as comprehension and syntactic processing, listening and reading must be trained.

However, teaching and assessing LC is often encountered as time consuming (Wolf et al., 2019) and therefore it hardly receives attention in the school curriculum (Mommers, 2020). In the higher grades of primary schools, LC is not part of the standard assessment. LC tests are only assessed if a student scores poorly on a RC exam and teachers wonder whether this is due to the level of technical reading or problems with comprehension in general. LC is necessary to understand a teacher's instruction and to gain knowledge in school, but it is also essential in everyday communication. In addition, as previously explained in the Simple View of Reading, it is an important predictor of reading success.

In the scientific literature, LC has often been examined as one of several lexical predictors of RC. No research has yet been done on lexical predictors of LC itself. Nevertheless, there are two Dutch studies with interesting outcomes. In the study of De Jong & van der Leij (2002), the influence of phonological skills, and linguistic comprehension on the development of RC was examined. Therefore, 141 Dutch children were assessed with various language tests at the end of grade 3 and grade 5. It came out that word decoding, vocabulary, and LC were found to influence the development of RC. Regression analyses showed that vocabulary was fully responsible for the additional effect that LC had on RC. Initially, this longitudinal study focused on RC, however, these results indicate a suspicion that vocabulary has an impact on LC. Verhoeven and Van Leeuwe (2008) set up a larger study by following 2143 Dutch children from 118 different schools from grade 3 to 8 of the

Dutch educational system. Again, word decoding, vocabulary, and LC were found to be significant predictors of reading comprehension. The effects of word decoding on RC decreased as students progressed to higher grades. The interesting question arises whether these predictors of RC are the same for LC since their cognitive processes are, apart from the input, very similar.

A considerable number of other studies have found vocabulary knowledge to be a significant predictor of reading success (Hirsh & Nation, 1992; Hu & Nation, 2000; Laufer, 1992; Sinatra et al., 2012). The larger the vocabulary, the greater the understanding of a text or story will be. The results of research by Sinatra (1990) suggests that the processes of listening and reading share the same mental lexicon. All word information is organized as a computer network in the brain. Linguistics uses the term mental lexicon for the systematic representation of word knowledge in long-term memory. Knowledge about multiple words form the vocabulary. In the literature a distinction is made between productive and receptive vocabulary. Productive (active) vocabulary refers to all the words that a person can use by himself through speech or writing. Receptive (passive) vocabulary refers to all words that a person understands when he sees or hears it. This will be focused on in this study, as children show their receptive knowledge earlier and better than their productive knowledge (Fan, 2000; Laufer & Paribakht, 1998; Zhou, 2010).

Having well developed word decoding skills (converting graphemes to phonemes) are important for good comprehension. Results from longitudinal studies indicate that, at least during primary school years, word recognition and word decoding (merged, the term reading fluency is integrated) also seem to predict RC (Adlof et al., 2006; Kendeou et al., 2009). In the present study it is examined whether this also applied to LC instead of RC, since it is argued from the Simple View of Reading that there is an interaction between LC and word decoding (i.e., they reinforce each other). Better decoding skills leaves more opportunity for language comprehension. Another explanation is that better decoders are also exposed to more language, which increases their vocabulary and improves comprehension (both during reading and listening). There is no answer to this yet, as it has not been investigated before. Thus, it is not yet clear what could be predictors of measures of LC; a convenient model such as a "Simple View of Listening" does not yet exist in the scientific literature. Moreover, in the Netherlands there are still few LC measures for whole-class administration. Creating manageable measures for LC, that can be administered in a robust and time-efficient manner, may help to gain insight into students' LC skills. In addition, the results of LC tests may be able to predict future problems with RC, and teachers may be able to anticipate by paying more attention.

The present study investigates the relationship between lexical predictors and measures of listening comprehension focusing on two main questions: 1) Are students' reading fluency skills and/or vocabulary predictors of listening comprehension measured with questions? and 2) Are students' reading fluency skills and/or vocabulary predictors of listening comprehension measured with a situation model?

For the first research question, the RC subtest of the Drempeltoets (Aarnoutse & Kapinga, 2006) was converted into a LC test. In the second research question, LC was assessed by creating a situation model (following Raudszus et al., 2019). LC has not been measured with this before. In the model a choice must be made at what distance certain concepts were placed in a framework based on the story that has been listened to.

Since various researchers showed evidence that word decoding (reading fluency) and vocabulary are predictors of RC, and since RC and LC largely consist of similar cognitive processes, this study examines whether these two lexical variables can also predict LC. It is expected that better reading fluency and larger vocabulary will improve LC performances.

Method

Participants

The participants of this study were 47 students (30 boys, 63.8%) between 9 and 11 years of age (*M* = 9.68; *SD* = .56). The students were recruited at two primary schools in Drenthe, a region in the north of The Netherlands. All participants needed to be in grade 6 (fourth year of formal schooling) or higher of Dutch primary schools and had to have a sufficient level of Dutch. Within the research group, 42 students were in the 6th grade and five were in grade 7. No students had impairments in hearing and/or sight that could not be compensated with hearing aids or glasses/lenses or delays in their (language) development.

Background data were requested from the parents. Students who doubled the school year, who were raised bilingually or participants who were diagnosed or suspected of having dyslexia, ADHD, or an autism spectrum disorder, were not excluded from this study. In the group of participants there were three students with dyslexia, in two other participants the parents suspected dyslexia. One student was diagnosed with an autism spectrum disorder. Three students had ADHD and three were suspected of having ADHD. Seven students were excluded due to no response from their parents to the consent forms and for one student the parents did not give permission for participation.

Materials

This study is part of a larger study into testing and predicting LC. Various instruments were administered to test the language abilities of the students. In addition, three questionnaires had to be completed by the parents, the teachers, and the students to provide more information about the background of the participant.

Reading fluency

Reading fluency was measured using the Eén-Minuut-Toets (EMT; Brus & Voeten, 2019) and De Klepel (van den Bos et al., 2019) respectively. The EMT is a list that contains

116 separate words placed one below the other in four columns. It is intended to determine a general level of technical reading. The student was instructed to read the words aloud as well and as quickly as possible in one minute. De Klepel consists of pseudo words and is therefore a decoding test. It examines the skill in converting letters and letter groups into sounds and assembling them into words. The student was given two minutes for this test. The pseudo words have no meaning but are pronounceable. The construction of De Klepel is based on the EMT. The 116 words of the EMT have been transformed into 116 pseudo words of the same length and difficulty in De Klepel. The pseudo words are also based on comparable syllables and sound clusters. The reliability and concept validity of the EMT were evaluated in 1981 by the "Commissie Testaangelegenheden Nederland" (COTAN) as 'good'. The reliability and concept validity of De Klepel were evaluated as 'sufficient' (COTAN, 1996). The criterion validity, for both the EMT and De Klepel, had not been assessed because this was not applicable according to the publishers. The performances of the EMT and De Klepel by each student were recorded with audio equipment. For the scoring, all audio recordings were listened to in order to accurately assess the pronunciation of the words for correctness. The final scores were the total number of words read minus the number of misread words. In case of doubts about the pronunciation of a sound, a colleague was asked for a second assessment. *Z*-scores were calculated to create a composite reading fluency score, based on the addition of both tests scores.

Vocabulary

All students completed a modified version of the Peabody Picture Vocabulary Test (Dutch version: PPVT-IV-NL; Dunn & Dunn, 2007). This is a multiple-choice test which gives an indication of the receptive vocabulary. The participant chooses out of four images the correct image for an orally presented word. A selection of 20 items from four sets of the original test was used to make whole-class administration feasible. The selection of words is added Table 4 in Appendix A. In this modified version, students marked their chosen picture with a pencil in a booklet instead of pointing to a card as is done during the original (individual) administration. Figure 1 in Appendix A shows the instruction of the modified test. COTAN (2006) evaluated the reliability of the PPVT-III-NL as 'good' and the concept validity as 'sufficient'. The criterion validity was evaluated as 'unsatisfactory', as this has not yet been investigated. The scoring was done according to the manual. The maximum possible score was 20, for each mistake one point was deducted. If none of the pictures or more than one was circled in an item, the child was assumed not to know the answer and a point was also deducted.

Listening comprehension

LC in this study was measured in two tasks developed by the project group.

Listening comprehension questions. First, a modified version of a subtest of the Drempeltoets (Aarnoutse & Kapinga, 2006) was used. The originally reading comprehension task was converted into a listening comprehension task for this study. Two texts were read aloud by the researcher. After listening to text one (narrative text) the student had to answer six questions about the text. After five minutes the researcher read aloud the same text again and the students were given another five minutes for the questions. Text two (informational text) proceeded similarly. For both texts three multiple choice questions (four answer options) and three true/false questions could be answered afterwards. The questions were asked in the same way as in the original test, in a response booklet the students could circle the correct answer (A/B/C/D) and write down 'true' or 'false'. The answers from the manual were consulted for the scoring. A total of 12 points could be obtained for this listening comprehension test. For an incorrect answer or no answer, one point was deducted.

Listening comprehension situation model. For this task, a text from the Drempeltoets was used also. After listening to the third text (informational text, see Figure 2 in Appendix B) read aloud by the researcher, the students were asked to create a situation model on a laptop with the program jRateDrag v.2.0 (Schuelke, n.d.)[1]. The students were exposed to nine core terms (see Table 5 in Appendix B) that occurred in the text and were instructed to drag related terms closer together and less related terms further apart. After five minutes of creating, the text was read again, and students were allowed to modify their situation model in up to five more minutes. In Appendix C, four examples of the situation models, created by different participants, are shown. To analyze the situation models, the program Jpathfinder v 1.0 (Schvanefeldt, 1990) was used. Between each pair of terms, pixel distances were calculated in this program, from which individual matrices were created. With Minkowski's r set to 3 and q = n-1, the individual matrices were transformed into networks. To assess the quality of the networks, nine experts (skilled adult language comprehenders) were asked to listen to the same text and also create a situation model (following Fesel et al., 2015; Raudszus et al., 2019). As an indication of the similarity between the participant's network and the nine expert networks, the intersection of the networks was converted into a similarity score between 0 and .5, with a value indicating that the student's situation model was more similar to the expert models.

Research procedure

All data in this experimental study is quantitative and was collected by a master's student of the University of Groningen. A convenience sample was realized by approaching school principals and teachers of grade 6 of schools in the region of Groningen and Drenthe (The Netherlands) for participation via e-mail. When schools were willing to participate, parents from the potentially participating students received information about the project and an online consent form. Parents were asked to give permission for their child to participate in

[1] Before this was administered, the students first practiced with the program and the necessary technical skills with a sample text and a practice model.

the research as well as to grant permission for researchers to use the data anonymously. Children who had consent of their parents were asked to confirm they themselves also wanted to participate before the individual testing sessions started. Only students who agreed to participate were included in the study.

All participants completed four sessions: two plenary classroom sessions (both 20-30 minutes of administration) and two individual sessions (6 and 20 minutes). In between the plenary tests, breaks and/or movement exercises were organised. Compensation for the collaboration were stickers for the children and chocolates for the teachers.

Data analysis

All data were stored in IBM SPSS Statistics 26. The descriptive and statistical analyses were also performed in this program.

The research questions aimed to assess whether reading fluency and vocabulary can predict two different LC tasks in children in grade 6 and 7 of Dutch primary schools. Descriptive analyses for the four language tasks were performed for the mean, the standard deviation, and the range. Erroneous values and missing values were visually checked, and univariate outliers were tested. There was no case of erroneous or missing values in the data. A score was considered a univariate outlier if the *z*-score was above +3.29 or below -3.29 (Tabachnick & Fidell, 2013). No univariate outliers were found.

Multiple linear regressions were performed to statistically test if the two independent variables Reading Fluency and Vocabulary were predictors of the LC tasks. In the first analysis, LC measured by the situation model building ability was the dependent variable. In the second analysis, LC measured by multiple choice questions of the modified Drempeltoets was the dependent variable. In the multiple linear regression analyses a 5%-significance level was used in which a result of $p < .05$, the independent variable could be considered as a significant predictor. Before performing the analyses, it has been checked whether the data meets the assumptions for multiple linear regression.

Results

Descriptive Statistics

Table 1 summarizes the descriptive statistics (mean, standard deviation, and range) for vocabulary, reading fluency (before the two tests were merged into one value), and the two listening tasks. All students completed all tasks used in this study (N = 47).

Table 1

Descriptive Statistics of Raw scores Vocabulary, Reading Fluency, and Listening Comprehension (N = 47).

	M	*SD*	*Min*	*Max*
Vocabulary	12.91	2.73	8	18
Reading Fluency - EMT	62.09	15.61	24	100
Reading Fluency – De Klepel	48.57	16.33	11	100
Listening Comprehension Situation model	.25	.08	.08	.47
Listening Comprehension Questions	9.19	1.31	6	12

Assumption checks

The data was checked for multivariate outliers, normal distribution of the residuals, linearity, homoscedasticity, and multicollinearity. Multivariate outliers were investigated by calculating Mahalanobis distance. None of the respondents showed a divergent profile ($p < .001$), so there was no evidence of multivariate outliers. Histograms and P-P plots (Figures 7-10 in Appendix D) were used to visually check the normality of the residuals. To be sure, the Shapiro-Wilk test was performed, the residues were found to be normally distributed ($p = .141$). The homoscedasticity and linearity were checked using scatterplots of the standardized residuals and the standardized predicted values (see Figures 10 and 11 in Appendix D). Since random cloud patterns appears the assumptions of linearity and homoscedasticity do not seem to be violated. Finally, the Variance Inflation Factor (VIF) was calculated to determine whether the independent variables were related to each other (multicollinearity). A VIF higher than 4 or a tolerance lower than .25 indicates a high correlation between the independent variables (O'Brien, 2007), which turned out not to be the case (Reading Fluency, *Tolerance* = .94, *VIF* = 1.07; Vocabulary, *Tolerance* = .94, *VIF* = 1.07). None of the assumptions were violated.

Regression analyses

Multiple linear regression analysis was conducted in SPSS to answer the research question if reading fluency and/or vocabulary may be lexical predictors of listening comprehension questions. The results of the regression indicated the variables explained 32.4% of the variance (R^2 = .32, F (2,44) = 10.55, p <.001). The results of the multiple regression analysis are in Table 2. It was found that reading fluency significantly predicted the performances of the listening comprehension questions (β = .47, t = 3.70, p <.001). Vocabulary turned out not to be a significant predictor (β = .22, t = 1.72, p <.09). Students who read aloud faster and more accurately than their peers seem to achieve significantly higher scores on the questions.

Table 2

Multiple regression results for Listening Comprehension Questions

Listening Comprehension Questions	*B*	*95% CI for B*		*SE B*	β	R^2	ΔR^2	*p*
		LL	*UL*					
Model						.32	.29	
Constant	7.83	6.20	9.46	.81				<.001
Reading Fluency	.65	.29	1.00	.17	.47			<.001
Vocabulary	.11	-.02	.23	.06	.22			.09

Note. Statistical analysis: two-tailed, multiple regression. Model = "Enter" method in SPSS; *B* = unstandardized regression coefficient; *CI* = confidence interval; *LL* = lower limit; *UL* = upper limit; *SE B* = standard error of the coefficient; *β* = standardized coefficient; R^2 = coefficient of determination; ΔR^2 = adjusted R^2.

A second multiple regression analysis was done to test if reading fluency and/or vocabulary may be lexical predictors of the listening comprehension situation model (similarity score between the student model and the nine expert models). The results can be found in Table 3. The results of the regression showed that the two predictors did not explain the situation model scores ($R^2 = .00$, F (2,44) = .07, p <.931). The two lexical variables may not predict the listening comprehension situation model based on these data.

Table 3

Multiple regression results for Listening Comprehension Situation model

Listening Comprehension Situation model		*B*	*95% CI for B* *LL*	*UL*	*SE B*	*β*	R^2	ΔR^2	*p*
Model							.00	-.04	
	Constant	.26	.147	.377	.06				<.001
	Reading Fluency	.01	-.02	.03	.01	.06			.715
	Vocabulary	-.00	-.01	.01	.00	-.03			.857

Note. Statistical analysis: two-tailed, multiple regression. *B* = unstandardized regression coefficient; *CI* = confidence interval; *LL* = lower limit; *UL* = upper limit; *SE B* = standard error of the coefficient; *β* = standardized coefficient; R^2 = coefficient of determination; ΔR^2 = adjusted R^2.

Discussion

The aim of the present study was to examine the relationship between the lexical predictors reading fluency and vocabulary and outcomes of listening comprehension. The results show that reading fluency was a significant predictor of listening comprehension questions. The more words from the EMT and De Klepel the student read aloud accurately, the more comprehension questions he or she answered correctly. Vocabulary was not found to influence the scores obtained on the listening comprehension questions. For another measure of LC, the students created a situation model. The analyses showed that both reading fluency and vocabulary had no predictive value for the situation model.

In this study, LC was tested in two different tasks. After listening to informative and narrative texts (that were read aloud by the researcher twice), the students created a situation model in a computer program and answered questions in a response booklet. Interestingly, not only the administration, but also the difficulty of the tasks and the interpretation of the analyses differed. For the listening comprehension questions, the given answers could be right or wrong. The results showed that none of the students had less than half of the answers right and there was one participant who got 12/12 correct answers. The average score was relatively high for this task (M = 9.19; SD = 1.31). Possible explanations are that the questions were too easy for students of these grades, or guessing had played a role. Another plausible explanation may be that the students achieved higher scores because the texts were read twice by the researcher. Since the students had already read the questions before the second time of listening, they may had been able to listen more specifically to detect missed information and to identify the correct answers.

The created situation models could not be evaluated “right” or “wrong”. Each student model was compared with the models of nine skilled adult language comprehenders to manufacture a similarity score. It was notable that the models, by both the students and the adults, were created in very diverse ways. Some made a horizontal or vertical enumeration of the words; others made a circle or chose some core terms around which they placed other words. It also varied from person to person which terms were placed closer or further away from each other. The question arises whether the nine adults could make better models than children. Creative thinking and making associations seemed to have the biggest role. The students and adults were not asked which strategies were used in making the situation models. The above emphasizes that the similarity scores in retrospect may have provided less interesting information, because there may be no good sample models to accompany a text or story. Every situation model is and will be unique.

There are no known researchers that have investigated potential predictors of LC before. It is known from the literature that reading fluency (Adlof et al., 2006; Kendeou et al., 2009) and vocabulary (Hirsh & Nation, 1992; Hu & Nation, 2000; Laufer, 1992; Sinatra et al., 2012) are predictors of RC. The same has been found in longitudinal studies among Dutch students by De Jong & Van der Leij (2002) and Verhoeven & Van Leeuwe (2008). In addition, except for the input, the cognitive processes in RC and LC are very alike (Carpenter & Just, 2013; Cowan, 1996; Hogan et al., 2011; Kidd, 2013; Perfetti et al., 2005; Van den Broek, 2009). Therefore, for the research questions of this study, it was expected that reading fluency and vocabulary could also be lexical predictors of LC. Reading fluency was indeed a significant predictor of the listening questions. The rest of the outcomes were not significant. However, the power was low with a value of .627. The original aim was to recruit 150 participants. This goal was not met due to several setbacks in the process. A total sample size of 47 students was insufficient for a medium effect size of .15 and .80 power. A sample size of 68 or more participants would have been the desired size for drawing conclusions. In this case, non-significant results may be due to insufficient power since there is a greater chance of Type II errors (i.e., no effect was found while effects could be found in the population).

An interesting part of the study was the implementation and testing of newly developed or modified LC tasks. It is known that LC is an important early predictor of the development of reading (comprehension) (Heuvelman & Schreiner, 2010) and that in the upper grades of primary school there is less focus on teaching LC (Mommers, 2013). Good and efficient measures are useful for the future. Answering questions after listening to texts seems appropriate for measuring LC. However, the current analysis of the situation model is strongly questioned, as mentioned earlier.

A recommendation for further research is, instead of calculating a similarity score (to what extent the student models are comparable to adult models), to calculate the distances between all words and to calculate whether the students (and adults) used certain words together significantly more often. Latent semantic analysis (LSA; Landauer & Dumais, 1997) can be used to investigate the semantic overlaps between words. According to Zwaan & Madden (2004) semantic overlap among words in a situation model may affect the process of comprehension. Another recommendation for the situation models is the examination of word frequency. This may be an interesting predictor for creating situation models. Students may shift the high-frequency words first and then place the low-frequency words around them. In second language learners, the ability to recognize high frequency words in orally presented texts is a predictor of successful L2 listening comprehension (Matthews & Cheng, 2015). It is possible that this also applies to LC in the first language.

Thus, in follow-up research, more students should be included for a higher power. It is recommended to test as many 7th grade students as 6th grade students and a more comparative number of boys and girls for better comparison. In addition, it can be interesting to investigate whether students listen differently to informative and narrative texts or to look at answers in open questions in addition to multiple choice questions. Children may reproduce information better if they are asked multiple-choice questions since they can test whether they can remember one of the answer options.

LC is an important language skill on which the focus should not be lost in the Dutch curriculum. The present study has given a head start for new assessments of LC; reading fluency came out to be a significant predictor of one of the two tasks but there is still plenty to explore about LC!

References

Aarnoutse, C., & Kapinga, T. (2006). *Het Drempelonderzoek - Begrijpend lezen.* 678 Onderwijs Advisering.

Ahlers, L. & Mortel, K. van de (2009). *Begrijpend lezen. Een doorgaande lijn van groep 1 tot en met 8.* CPS Onderwijsontwikkeling en advies.

Adlof, S. M., Catts, H. W., & Little, T. (2006). Should the simple view of reading include a fluency component? *Reading and Writing*, 19(9), 933–958.

Berends, R. (2011). Begrijpend leesonderwijs: zin en onzin. *Tijdschrift Taal*, 2(3), 22-29.

Bos, K.P., van den, Lutje Spelberg, H.C., Scheepstra, A.J.M., & Vries, J.R., de (1994). *De Klepel: Een test voor de leesvaardigheid van pseudowoorden.* Pearson.

Brus, B.T., & Voeten, M.J.M. (2019). *Een-minuut-test, vorm A en B: Verantwoording en handleiding.* Pearson.

Carpenter, P. A., & Just, M. A. (2013). The role of working memory in language comprehension. In *Complex information processing* (pp. 51-88). Psychology Press.

Clariana, R. B., & Wallace, P. E. (2009). A comparison of pair-wise, list-wise, and clustering approaches for eliciting structural knowledge. *International Journal of Instructional Media*, 36(3), 287–302.

Commissie Testaangelegenheden Nederland (1981). *Een-Minuut-Test.* Retrieved from http://www.cotandocumentatie.nl/beoordelingen

Commissie Testaangelegenheden Nederland (1996). *De Klepel.* Retrieved from http://www.cotandocumentatie.nl/beoordelingen

Commissie Testaangelegenheden Nederland (2006). *Peabody Picture Vocabulary Test-III-NL.* Retrieved from http://www.cotandocumentatie.nl/beoordelingen

Cowan, N. (1996). Short-term memory, working memory, and their importance in language processing. *Topics in Language Disorders,* 17(1), 1-18.

Danks, J. H., & End, L. J. (1987). Processing strategies for reading and listening. In R. Horowitz & S. J. Samuels (Eds.), *Comprehending oral and written language* (pp. 271-294). American Press.

de Jong, P. F. (2011). Begrijpend lezen. In P. F. de Jong & H. M. Y. Koomen (Eds.), *Interventie bij onderwijsleerproblemen (pp. 25-39)*. Garant.

de Jong, P. F., & van der Leij, A. (2002). Effects of phonological abilities and linguistic comprehension on the development of reading. *Scientific Studies of Reading*, 6(1), 51-77.

Dunkel, P. (1991). Listening in the native and second/foreign language: Toward an integration of research and practice. *TESOL Quarterly*, *25*(3), 431-457.

Dunn, L.M., & Dunn, D.M. (2007). *Peabody Picture Vocabulary Test - fourth edition.* Pearson.

Engzell, P., Frey, A., & Verhagen, M. D. (2020). *Learning inequality during the COVID-19 pandemic* (No. ve4z7). Center for Open Science.

Fan, M. (2000). How big is the gap and how to narrow it? An investigation into the active and passive vocabulary knowledge of L2 learners. *RELC Journal*, 31(2), 105-119.

Fesel, S. S., Segers, E., Clariana, R. B., & Verhoeven, L. (2015). Quality of children's knowledge representations in digital text comprehension: Evidence from pathfinder networks. *Computers in Human Behavior*, 48, 135-146.

Gough, P. B., & Tunmer, W. (1986). Decoding, reading, and reading disability. *Remedial and Special Education*, 7, 6–10.

Haelermans, C., van der Velden, R., Aarts, B., Bijlsma, I., Huijgen, T., Jacobs, M., van Vught, L., & van Wetten, S. (2021). *Door eerste schoolsluiting minder leergroei, vooral bij begrijpend lezen.* Nationaal Regieorgaan Onderwijsonderzoek

Heuvelman R., & Schreiner, G. (2020). *Hoe stimuleer je het begrijpend luisteren al bij jonge kinderen?* Early Years Blog. Universiteit Utrecht. Retrieved from https://earlyyearsblog.nl/2020/01/17/hoe-stimuleer-je-het-begrijpend-luisteren-al-bij-jonge-kinderen/

Hirsh, D., & Nation, I. S. P. (1992). What vocabulary size is needed to read unsimplified texts for pleasure? *Reading in a Foreign Language*, 8, 689–696.

Hogan, T., Bridges, M., Justice, L., & Cain, K. (2011). Increasing higher level language skills to improve reading comprehension. *Focus on Exceptional Children*, 44(3), 1–20.

Hoover, W. A., & Gough, P. B. (1990). The simple view of reading. *Reading and writing,* 2(2), 127-160.

Hu, M., & Nation, I. S. P. (2000). Unknown vocabulary density and reading comprehension. *Reading in a Foreign Language*, 13, 403–430.

Inspectie van het Onderwijs. (2022). *De Staat van het Onderwijs 2022*. Utrecht: Inspectie van het Onderwijs.

Kendeou, P., van den Broek, P., White, M. J., & Lynch, J. S. (2009). Predicting reading comprehension in early elementary school. *Journal of Educational Psychology*, 101(4), 765–778.

Kidd, E. (2013). The role of verbal working memory in children's sentence comprehension: A critical review. *Topics in Language Disorders*, 33(3), 208-223.

Kintsch, W., & Kozminsky, E. (1977). Summarizing stories after reading and listening. *Journal of Educational Psychology*, 69(5), 491.

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological review*, 104(2), 211.

Laufer, B. (1992). How much lexis is necessary for reading comprehension? In P. J. L. Arnaud, & H. Béjoint (Eds.), *Vocabulary and applied linguistics* (pp. 126-132). Palgrave Macmillan.

Laufer, B., & Paribakht, T. S. (1998). The relationship between passive and active vocabularies: Effects of language learning context. *Language Learning*, 48(3), 365-391.

Matthews, J., & Cheng, J. (2015). Recognition of high frequency words from speech as a predictor of L2 listening comprehension. *System*, 52, 1-13.

Meshcheriakova, O., Borghans, L., Haenbeukers, R., & Schils, T. (2020). Scholen dicht, prestaties gedaald. *Didactief*, 50(10), 25-26

Mommers, F. C. (2003). Wat onderzoek leert over begrijpend lezen. *Jeugd in School En Wereld*, 87(5), 11–14.

Mommers, J. C. (2020). The Relationship Between Decoding Skills, Reading Comprehension and Spelling Skills in the First Three Years of Primary School. In M. Young, M. Thomas, & P. Munn (Eds.), *Reading and writing skills in primary education* (pp. 63-75). CRC Press.

Oakhill, J., & Cain, K. (2007). Introduction to Comprehension Development. In K. Cain & J. Oakhill (Eds.), *Children's comprehension problems in oral and written language: A cognitive perspective* (pp. 3–40). Guilford Press.

O'brien, R. M. (2007). A caution regarding rules of thumb for variance inflation factors. *Quality & quantity*, 41(5), 673-690.

Oostdijk, N.H.J., Hoste, V., De Jong, F., Reynaert, M., De Clercq, O., Van den Heuvel, H., & Desmet, B. (2014). *SoNaR-500.*

Perfetti, C. A., Landi, N., & Oakhill, J. (2005). The acquisition of reading comprehension skill. In M. J. Snowling & C. Hulme (Eds.), *The science of reading: A handbook* (pp. 227–247). Oxford: Blackwell Publishing.

Potocki, A., Ecalle, J., & Magnan, A. (2013). Narrative comprehension skills in 5-year-old children: Correlational analysis and comprehender profiles. *The Journal of Educational Research*, 106(1), 14–26.

Raudszus, H., Segers, E., & Verhoeven, L. (2019). Situation model building ability uniquely predicts first and second language reading comprehension. *Journal of Neurolinguistics*, 50, 106-119.

Schuelke, M. (n.d.). *JrateDrag, v2.0.* Retrieved from: http://endaemon.com/.

Schvaneveldt, R. W. (1990). *Pathfinder associative networks: Studies in knowledge organization*. Ablex Publishing.

Sinatra, G. M. (1990). Convergence of listening and reading processing. *Reading Research Quarterly*, 115-130.

SLO. (2018). *Begrijpend lezen: wat is dat? De componenten die een rol spelen bij begrijpend lezen.* Nationaal expertisecentrum leerplanontwikkeling.

Sticht, T. G., Beck, L. J., Hauke, R. N., Kleinman, G. M. & James, J. H. (1974). *Auding and reading: A developmental model.* Human Resources Research Organization.

Sticht, T. G., & James, J. H. (1984). Listening and reading. *Handbook of reading research*, 1, 293-317.

Tabachnick, B. G., & Fidell, L. S. (2013). *Using multivariate statistics* (6th edition). Pearson.

Van den Broek, P. W. (2009). *Cognitieve en neurologische processen tijdens begrijpend lezen: Fundamenteel onderzoek en onderwijskundige toepassing.* Universiteit Leiden.

Verhoeven, L., & Leeuwe, J. van. (2008). Prediction of the development of reading comprehension. A longitudinal study. *Applied Cognitive Psychology*, 22(3), 407-423.

Wiig, E. H., Semel, E., & Secord, W. A. (2013). *Clinical Evaluation of Language Fundamentals–Fifth Edition (CELF-5)*. Pearson.

Wolf, M. C., Muijselaar, M. M., Boonstra, A. M., & de Bree, E. H. (2019). The relationship between reading and listening comprehension: shared and modality-specific components. *Reading and Writing*, 32(7), 1747-1767.

Zhou, S. (2010). Comparing receptive and productive academic vocabulary knowledge of Chinese EFL learners. *Asian Social Science*, 6(10), 14.

Zwaan, R. A., & Madden, C. J. (2004). Updating situation models. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 30(1), 283-288.

Appendix A

The selected items of set 8, 9, 10 and 11 of the PPVT-III-NL and the instruction of the modified version used in this study

Table 4

The 20 selected items from the PPVT-III-NL for the modified version of this study. The items that were used from each set are marked in bold. The English translations of the used words are displayed in the last columns of the tables.

Item	New item	Dutch concept	English translation
Set 8 (entry 8;0-9;11)			
85	**1**	**Venster**	*Window*
~~86~~		~~Ploegen~~	
87	**2**	**Van leer**	*Leathern*
~~88~~		~~Dam~~	
~~89~~		~~Omhelzen~~	
90	**3**	**Vitrine**	*(Glass) showcase*
~~91~~		~~Woud~~	
92	**4**	**Geketend**	*Chained*
~~93~~		~~Autoriteit~~	
~~94~~		~~Haspel~~	
95	**5**	**Schuren**	*Sanding*
~~96~~		~~Prooi~~	
Set 9			
97	**6**	**Voertuig**	*Vehicle*
~~98~~		~~Onverwacht~~	
99	**7**	**Burcht**	*(specific sort of a) castle*
~~100~~		~~Vergiet~~	
~~101~~		~~Dakkapel~~	
102	**8**	**Vergezellen**	*(To) accompany*
~~103~~		~~Verstelbaar~~	
104	**9**	**Pelikaan**	*Pelican*
~~105~~		~~Klarinet~~	
~~106~~		~~Pedaal~~	
107	**10**	**Bankbiljet**	*Banknote*
~~108~~		~~Hiel~~	
Set 10			
109	**11**	**Kandelaar**	*Candle holder*
~~110~~		~~Naar het oosten~~	
111	**12**	**Balken**	*Braying*
~~112~~		~~Parallel~~	
~~113~~		~~Openbaar vervoer~~	
114	**13**	**Vloeibaar**	*Liquid*
~~115~~		~~Competitief~~	
116	**14**	**Signaal**	*Signal*
~~117~~		~~Lozen~~	
~~118~~		~~Garde~~	
119	**15**	**Sanitair**	*Sanitary*
~~120~~		~~Gefrankeerd~~	
Set 11			
121	**16**	**Peulvrucht**	*Legume*
~~122~~		~~Rund~~	
123	**17**	**Valuta**	*Valuta*
~~124~~		~~Agrarisch~~	
~~125~~		~~Identiek~~	
126	**18**	**Oase**	*Oasis*
~~127~~		~~Porselein~~	
128	**19**	**Doceren**	*Teaching*
~~129~~		~~Karaf~~	
~~130~~		~~Silhouet~~	
131	**20**	**Duet**	*Duet*
~~132~~		~~Transparent~~	

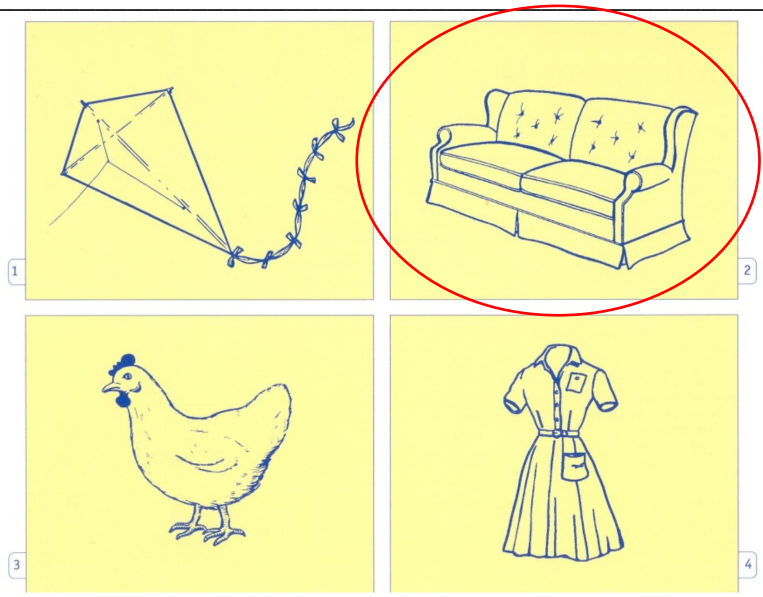
Figure 1

Instruction of the modified version of the PPVT-III-NL as was shown to the students in the response booklet.

Task instruction: “Je hoort straks 20 woorden en jij omcirkelt welk plaatje erbij hoort. Hieronder zie je twee voorbeelden voor het woord ‘bank’ en het woord ‘fietsen’. Succes!”
Translation: *You will soon hear 20 words. You may draw a circle around the picture that belongs to that word. Below you see two examples for the word 'sofa' and the word 'cycling'. Good luck!*

Voorbeeld *(example)* 1:

Bank *(Sofa)*



Voorbeeld *(example)* 2:

Fietsen *(Cycling)*

Appendix B

Listening Comprehension Situation Model

Figure 2

Text "Sundials" of the Drempeltoets that was read aloud by the researcher before students started creating a situation model.

Zonnewijzers

We weten dat de aarde om haar as draait. De zon komt 's morgens op en gaat 's avonds weer onder. 's Nachts zien we de zon niet, dan is het donker. 's Middags staat de zon het hoogst. In de zomer staat hij dan bijna recht boven ons. De stand van de zon heeft dus veel met de tijd te maken. Vroeger hadden de mensen geen horloges. Toch wisten ze wel ongeveer hoe laat het was. Dat zagen ze aan de stand van de zon. Als de zon hoog aan de hemel stond, was het ongeveer twaalf uur 's middags. Erg nauwkeurig was dat niet, maar de mensen hoefden vroeger ook niet precies op de minuut ergens te zijn.

Toch was er een manier om nauwkeurig de tijd te meten. Dat deed men met zonnewijzers. Een zonnewijzer bestaat uit een ijzeren staaf die schaduw op een wijzerplaat laat vallen. De wijzerplaat kan op een muur geschilderd worden of op een smeedijzeren ring zijn aangebracht. Door te kijken waar de schaduw valt, kun je zien hoe laat het is. Helaas is de wijzerplaat niet in alle landen even goed bruikbaar. In Nederland hebben we bijvoorbeeld niet zoveel zon als in Italië, Frankrijk of Spanje.

Soms lijkt de zonnewijzer van slag. Als je de tijd op de zonnewijzer vergelijkt met die op je horloge, dan is er een groot verschil. Ga er dan maar rustig van uit dat de zonnewijzer goed is en je horloge fout. De tijd op je horloge is de tijd die we samen in Europa hebben afgesproken en dat is lang niet altijd de zonnetijd of de echte tijd. De verschillen worden nog groter, omdat we met de zomer- en wintertijd het horloge ook nog eens een uur verzetten.

Table 5

The nine Dutch concepts used in the situation model and their English translation.

Item	Dutch concept	English translation
1	Zonnewijzer	Sundial
2	Zon	Sun
3	Tijd	Time
4	Wijzerplaat	Dial
5	Schaduw	Shadow
6	Nauwkeurig	Accurately
7	Horloges	Watches
8	Stand	Position
9	Zonnetijd	Solar time

Appendix C

Four (randomly chosen) examples of situation models made in jRateDrag by students of this study.

Figure 3

Situation Model A

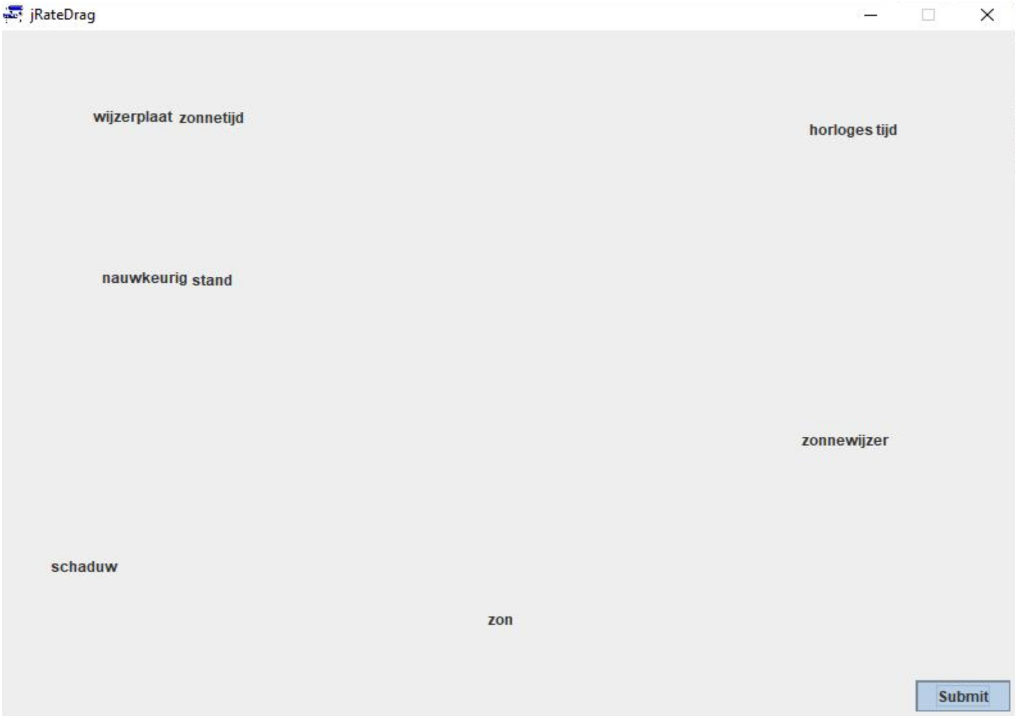


Figure 4

Situation Model B

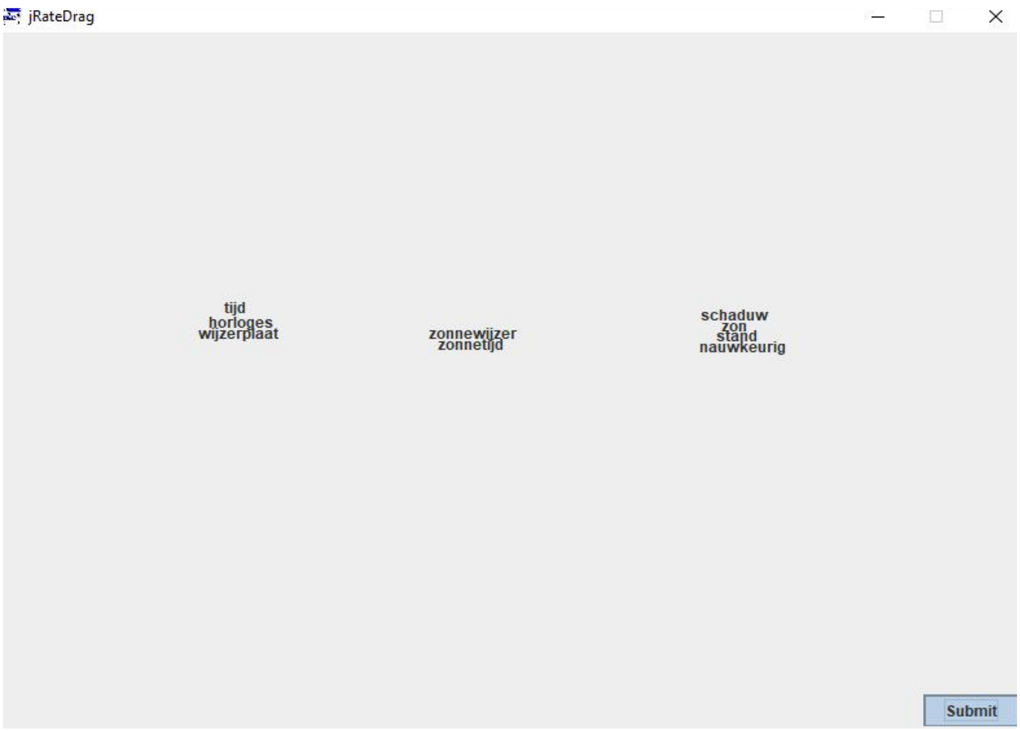


Figure 5

Situation Model C

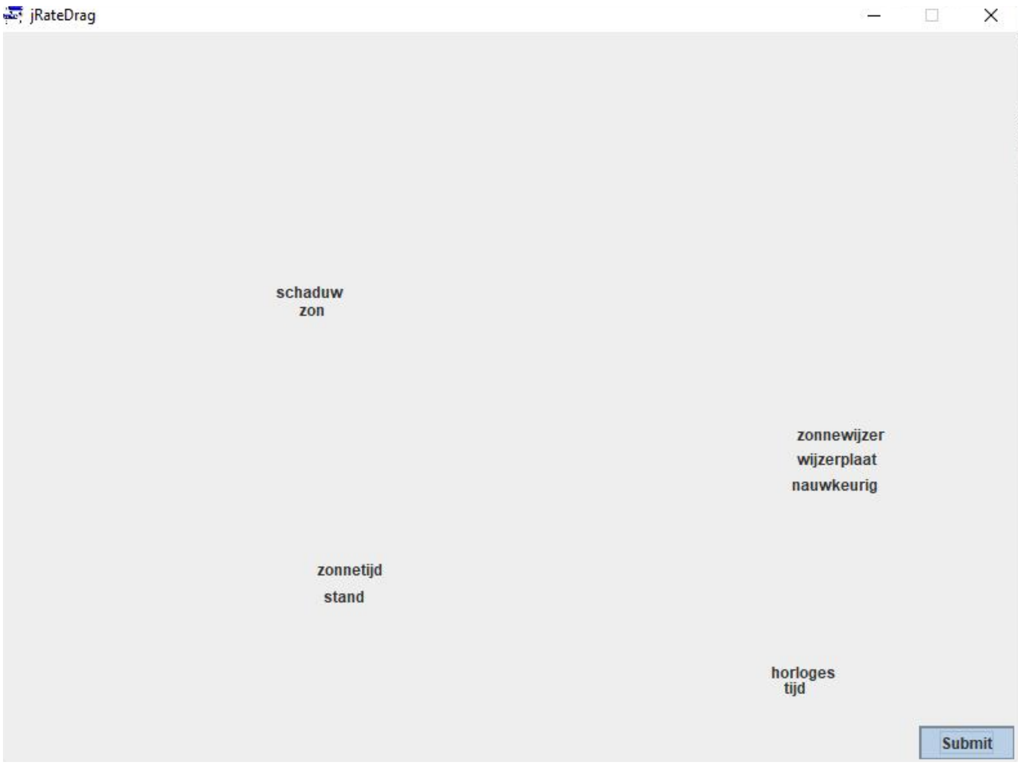
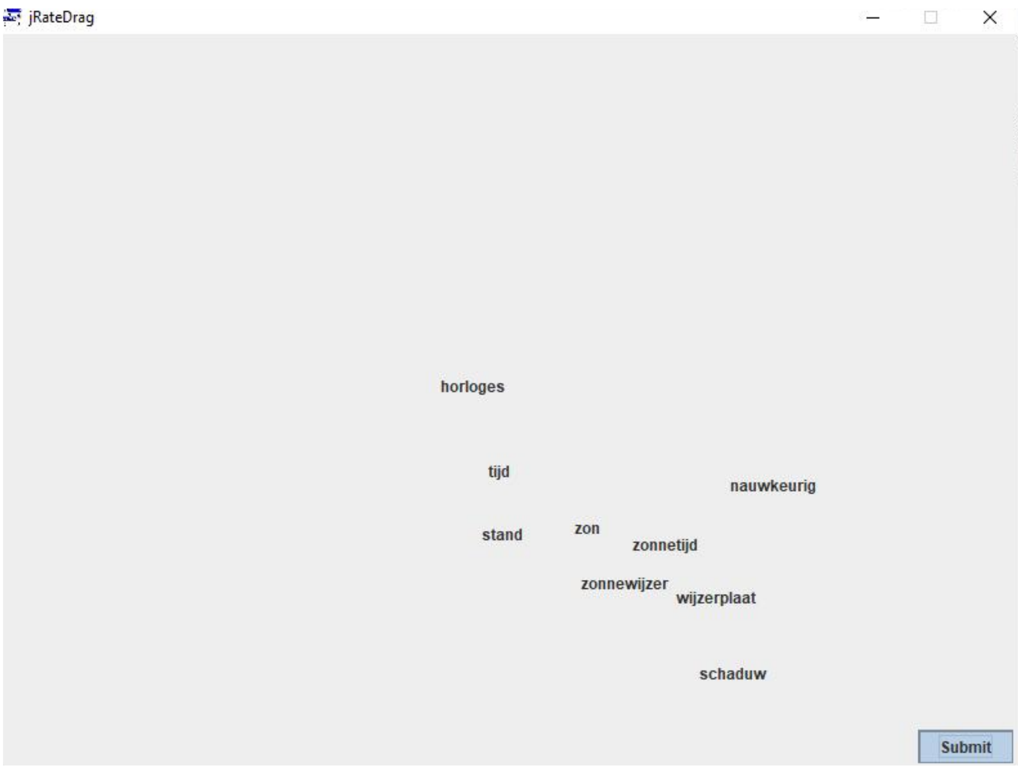


Figure 6

Situation Model D



Appendix D

Assumption checks

Figure 7

Histogram for the normality check of the listening comprehension questions.

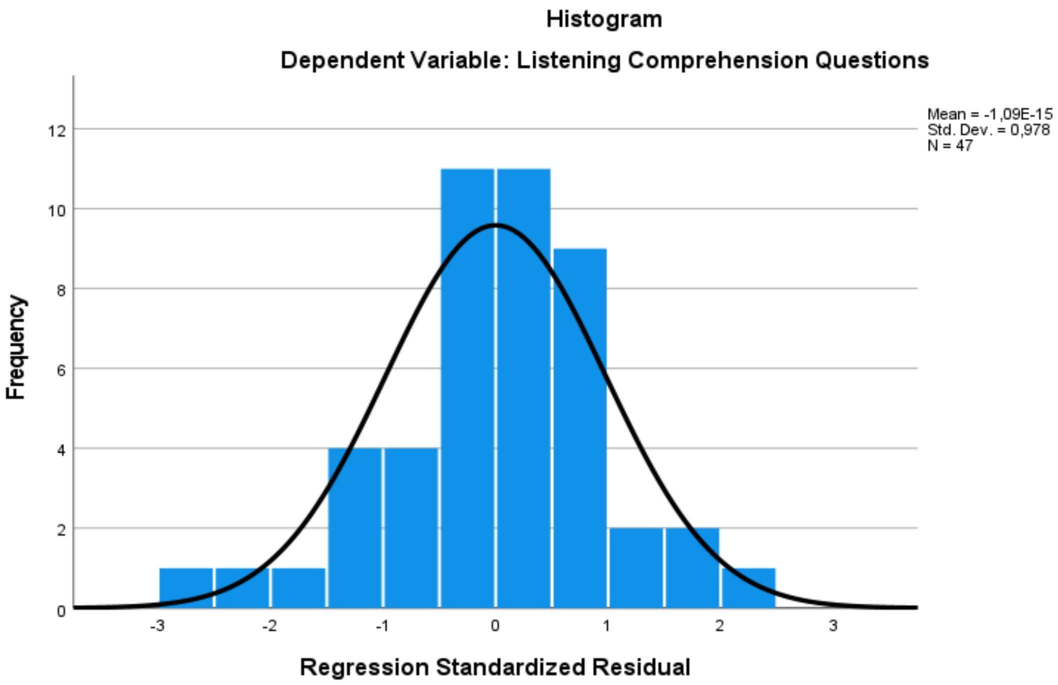


Figure 8

P-P Plot for the normality check of the listening comprehension questions.

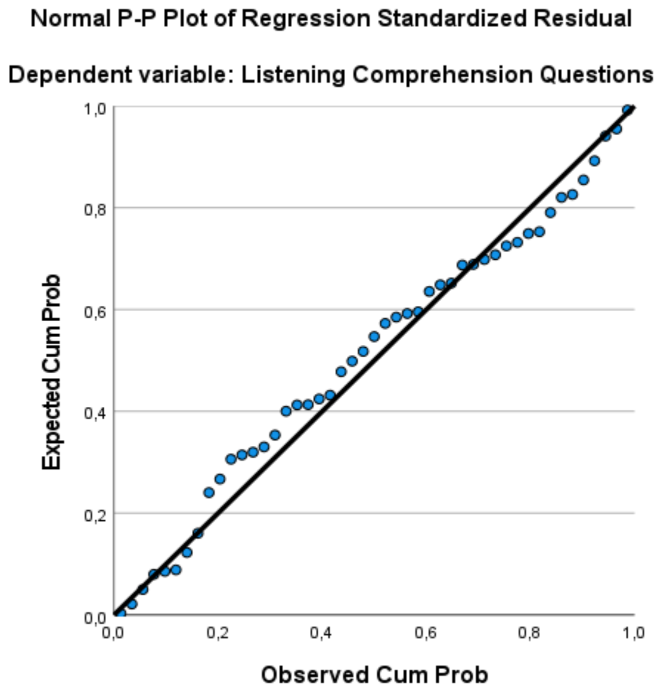


Figure 9

Histogram for the normality check of the listening comprehension situation model.

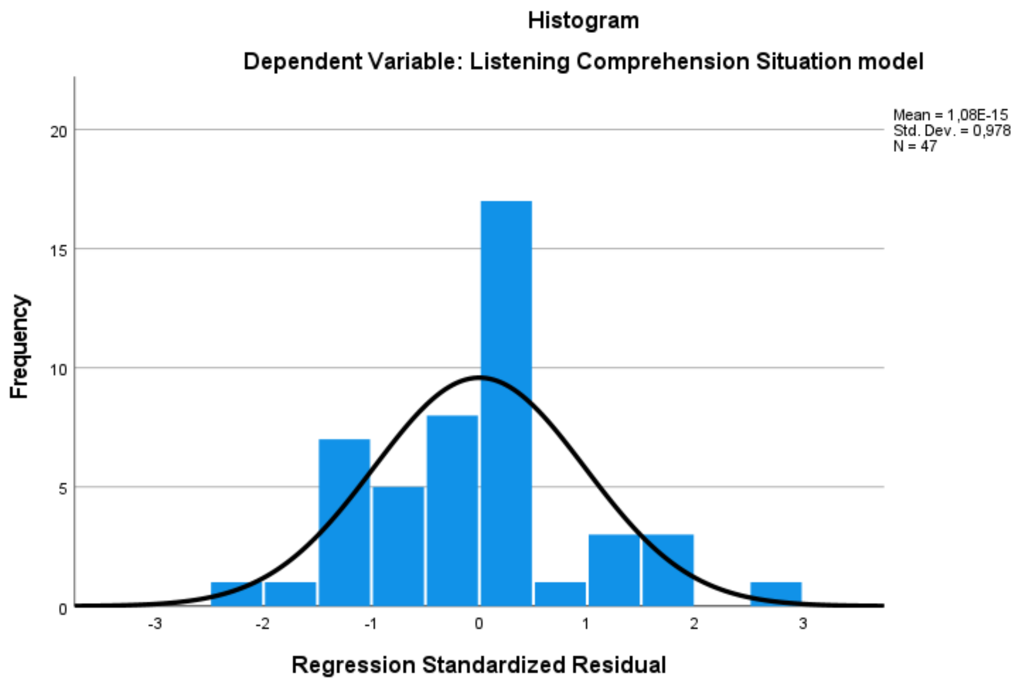


Figure 10

P-P Plot for the normality check of the listening comprehension situation model.

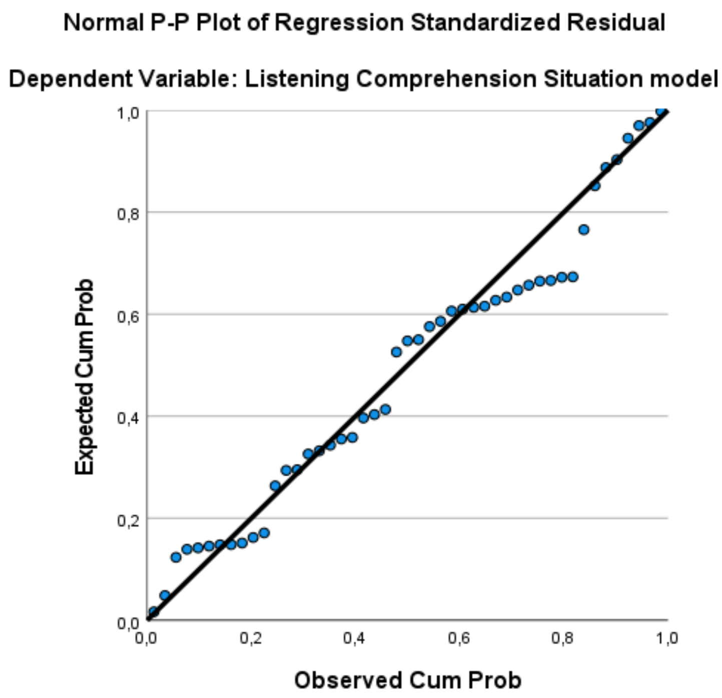


Figure 11

Scatterplot to check the assumptions for linearity and homoscedasticity for the residuals of the listening comprehension questions.

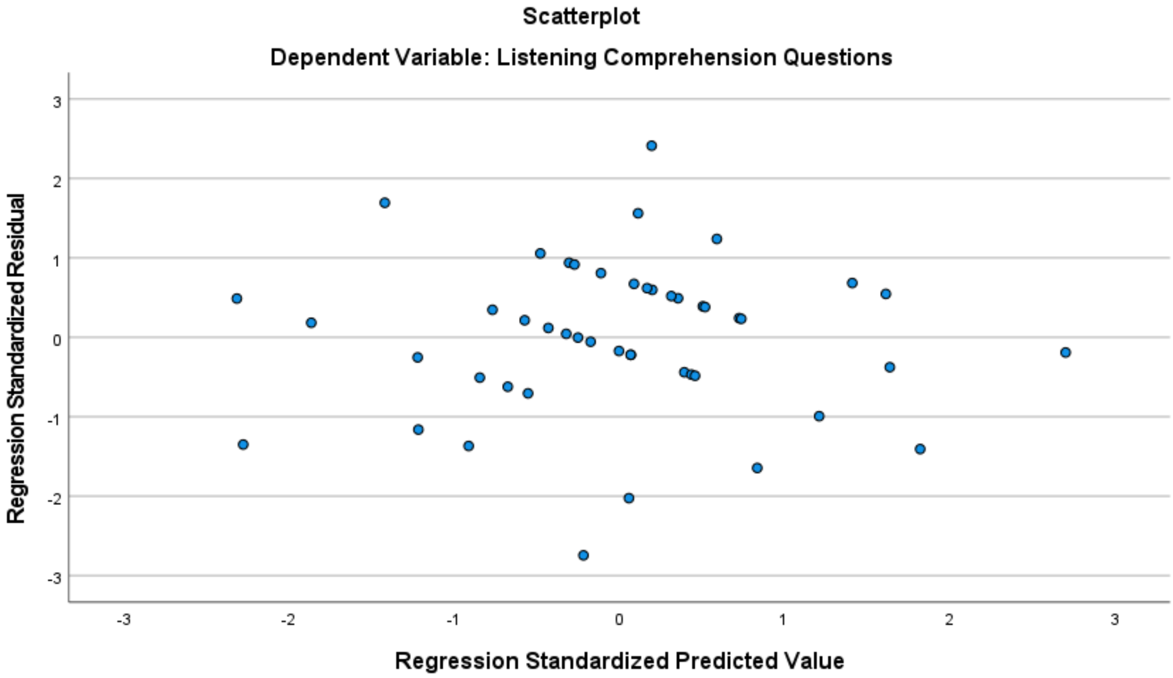


Figure 12

Scatterplot to check the assumptions for linearity and homoscedasticity for the residuals of the listening comprehension situation model.