

Vocabulary knowledge and reading comprehension in Arabic

Brahim Ait Hammou
Ministry of Education
Morocco
hammou76@gmail.com

ملخص:

بالرغم من وفرة الدراسات السابقة حول التغطية المعجمية وعلاقتها بفهم نصوص القراءة باللغة الإنجليزية، إلا أن هناك عجز علمي ومعرفي في فهم العلاقة بين التغطية وفهم النصوص في اللغة العربية. تسعى هذه الدراسة إلى سد هذه الفجوة والتحقيق في العلاقة بين التمكن من المفردات وفهم المقروء. ولهذه الغاية، قمنا بإجراء اختبارين على إجمالي 154 تلميذا مغربيا من المدارس الابتدائية والإعدادية والثانوية. في الأول تم تمرير رايث مرتبط بمعرفة مفردات النص. تم بعدها قمنا بتمرير اختبار آخر مرتبط بفهم النص القرائي. يتكون هذا الاختبار من 15 (خمسة عشر) سؤالاً للفهم يستهدف كل واحد منها مهارة قرائية مختلفة. أظهر تحليل البيانات أن المتعلمين في المستويات الدراسية الثلاث يتحكمون في المفردات اللغوية المرتبطة بالنص القرائي بشكل جيد مما يعني أن نسبة التغطية المعجمية للنص عالية (تلاميذ المستوى السادس ابتدائي = 55%؛ تلاميذ مستوى الجدد مشترك (الصف العاشر) = 74%) إلى نسب تغطية عالية جداً (تلاميذ المستوى التاسع = 91%). ومن جهة أخرى، تبين نتائج هذه الدراسة عن وجود علاقة ذات دلالة إحصائية بين فهم النص القرائي والتغطية المعجمية فقط لدى متعلمي المستوى الابتدائي. ورغم وجود درجات تغطية عالية للنص من حيث معرفة المفردات، حصل المتعلمون في المستويات الثلاثة فقط على نقطة أقل من عشرة من أصل 51 نقطة في فهم النص. يستنتج من هذا أن حجم المفردات اللغوية لدى المتعلمين قد يكون أقل مما هو ضروري لفهم النص بنسبة تمكن عالية. كما تشير نتائج هذه الدراسة إلى أنه حتى أثناء توفر مستويات تغطية معجمية في حدود 91%، لم يتمكن المتعلمون من الوصول إلى فهم كامل للنص.

Abstract

While there is a proliferation of studies on the relationship between vocabulary coverage and reading comprehension in English, this area has been under-researched in Arabic reading comprehension. This study sets out to fill this gap and to investigate the long-established connection between vocabulary knowledge and reading comprehension. To this purpose, the study administered two tests to a total number of 154 Moroccan learners of primary, middle and secondary high schools. First, a Yes/No vocabulary test targeting the words in the reading text was administered before learners were assigned a reading test composed of a text and 15 comprehension questions. The whole reading test was primarily used as a PISA training

test. The analysis of the data showed that the learners of the three groups reported high (6th graders = 55%; 10th graders = 74%) to very high (9th graders = 91%) coverage proportions. However, a significant correlation between reading comprehension and vocabulary coverage was reported only for primary school learners. Similarly, with these coverage levels, the learners of the three grade levels had only a score of less than 10 out of 15 points in reading comprehension. Also, the study showed that vocabulary coverage correlated significantly, though weakly, only with 6 reading comprehension sub-skills. The results suggest that the learners' vocabulary size might be lower than is required for successful text comprehension. Similarly, the results suggest that even at 91% coverage levels, successful comprehension cannot be taken for granted.

Keywords: Arabic L2, Reading comprehension, Coverage, Vocabulary

1 Introduction

The relationship between vocabulary knowledge and reading comprehension has been widely investigated, mainly in English language teaching EFL/ESL contexts. Various studies have examined the effect which knowledge of vocabulary might have on learners' performance in reading comprehension. These studies have established that a certain knowledge level of the text's vocabulary is necessary for successful comprehension (e.g., Hirsh and Nation, 1992; Karakoc and Kose, 2017; Laufer, 1992; Laufer and Ravenhorst-Kalovski, 2010; Nation, 2001, 2006; Qian, 1999, 2002; Stæhr, 2008). The amount of text vocabulary which a learner knows certainly affects the extent to which he/she comprehends the text. A reader who knows the meaning of a larger number of words in the text will certainly comprehend the text better than a reader who knows less words. Text coverage is defined as the proportion of words which the reader knows in a text (Hsueh-Chao and Nation, 2000; Ludewig, Hübner, and Schroeder, 2023).

As reviewed in the next section, a large bulk of research has examined coverage in English language reading comprehension. Although the studies which deal with this issue agree that vocabulary knowledge is a prerequisite for reading comprehension, there is little agreement about the threshold of vocabulary that is necessary for complete comprehension of texts (Schmitt, Jiang, & Grabe, 2011). The current study sets out to investigate the relationship between text vocabulary knowledge (i.e. coverage) and the comprehension of Arabic texts. To the best of our knowledge, no study has so far examined this area in relation to Arabic. Therefore, this study attempts to fill this gap by contributing to our understanding of the relationship between vocabulary knowledge and overall text comprehension as well as the relationship between coverage and specific reading comprehension skills. Similarly, this study sets out to highlight the reading sub-skills which correlate with knowledge of vocabulary. Based on the findings of this study, pedagogical recommendations are issued.

2 Review of Literature

Because of the scarcity of studies about coverage and the comprehension of Arabic texts, this section reviews mainly the studies which were conducted on this topic in English language teaching contexts. In this review, we are more interested in the vocabulary coverage proportions rather than any straightforward comparison between the two languages (i.e. English and Arabic), bearing in mind that the unit of measurement in both languages certainly yields different lemma/word family scores as a result of the inherent linguistic differences between the two language systems (Ech-Charfi, 2023; Ech-Charfi and Bouaddi, 2023).

One of the earliest studies on coverage and reading comprehension is Laufer (1989). The researcher used learners' self-reports to decide about the number of words learners don't know in the reading text. The results of this study indicated that for a successful comprehension of the text, learners should know around 95% of the words in the text. Laufer reported that 90% coverage did not show any difference between learners in comprehension. Also, Hirsh and Nation (1992) maintain that coverage of at least 98% tokens in the text is necessary for independent text comprehension. In a similar study, Feng and Webb (2020) argued that 95% knowledge of the word tokens in a text is a prerequisite for comprehension.

Hsueh-Chao and Nation (2000) examined coverage and reading comprehension of fiction texts at different coverage levels (80%, 90%, 95% and 100%) among students in a pre-university course. To show their comprehension, participants completed a multiple-choice (MC) and a cued written recall (WR) comprehension test. The results showed that none of the learners achieved adequate comprehension at 80% vocabulary coverage, a few did at 90%, and most did not even achieve adequate comprehension at 95%. The researchers concluded that the minimum amount of vocabulary coverage to make the text comprehensible is definitely above 80%, which is approximately 1 unknown word in 5. The study concluded that coverage of around 98% is necessary for unassisted reading for pleasure. Supporting these results, Nation (2001) maintains that "the probabilistic threshold is 98%" (p. 147). With this coverage, almost all learners have a chance of gaining adequate comprehension. Nation refers to the 95% coverage as only providing an acceptable level of comprehension.

In a more recent study, Laufer (2010) examined vocabulary knowledge, coverage and reading comprehension among EFL college students. The study concluded that for coverage of 98%, knowledge of 7000-8000 word families plus proper nouns is necessary. The study maintains that having more low frequent vocabulary in a text, reduces coverage rates. This indicates that the more difficult the text vocabulary is, the less likely learners will know it, and the more they find difficulties in comprehension. In a follow-up study, Laufer and Ravenhorst-Kalovski (2010) reported that 98% coverage was considered the threshold for unassisted reading, while 95% lexical coverage was considered as a coverage level for reading which required some guidance (on the part of the teacher or other resources such as dictionaries). The results reported by Laufer (2010) and Laufer and Ravenhorst-Kalovski (2010) are very similar to those of Hsueh-Chao and Nation (2000) and also those reported

by Nation (2001). In these three studies, a coverage of 98% is required for adequate text comprehension. A study by Schmitt, Jiang, and Grabe (2011) also reported a similar proportion. The researchers examined the relationship between vocabulary knowledge and reading comprehension among participants from 8 different countries. They reported that learners who know 95% to 98% of the words in a text are likely to understand it at 60% - 68%. The authors maintain that a 98% to 99% coverage is necessary for successful understanding of the text, with only one unknown word in each 100 tokens. The researchers concluded that “The total increase in comprehension went from about 50% comprehension at 90% vocabulary coverage to about 75% comprehension at 100% vocabulary coverage” (Schmitt, Jiang, and Grabe, 2011, p. 35). The researchers are skeptical about setting any vocabulary threshold for comprehension. Rather, as the study maintains, the required coverage level depends on the type of comprehension which is targeted (p. 39). What is striking in this study is that 90% vocabulary coverage could provide only comprehension of 50% of the information in the text, while 100% vocabulary coverage could lead only to the comprehension of 75% of the text. These results suggest that for a more successful comprehension of the text, coverage levels should be well-beyond a 90% coverage level.

Alavi and Akbarian (2012) examined the extent to which vocabulary knowledge is related to the performance on the five types of reading comprehension items tested in TOEFL (i.e., guessing vocabulary, main idea, inference, reference, and stated detail). The study reported a moderate, positive correlation between the vocabulary test (VLT) (Schmitt, Schmitt, and Clapham, 2001) and guessing vocabulary and understanding stated detail. The study also reported that VLT weakly correlates with main idea, inference and reference. Using a stepwise multiple regression, the study concludes that vocabulary knowledge accounts for up to 33% of unique variance in guessing vocabulary in reading comprehension texts.

To examine the relationship between coverage and reading comprehension among a sample of 178 Turkish EFL university learners, Güngör and Yaylı (2016) first tested the participants’ knowledge of the text’s vocabulary using a checklist. The findings of this study revealed that knowledge of vocabulary moderately correlated with reading comprehension with a relatively linear relationship between them. The study concluded that a level of 98% vocabulary coverage is necessary for learners to understand academic texts.

Al-Khasawneh (2019) Alkhasawaneh (2019) also examined the effect of vocabulary coverage on reading comprehension among 64 university freshmen students enrolled in the Department of English Language at King Khalid University, Abha, Saudi Arabia. The study used the Vocabulary Size Test (Schmitt, Jiang, and Grabe, 2011) and reading comprehension tasks based on two TOEFL preparation manual texts. The study reported a linear relationship between vocabulary knowledge at various frequency levels and reading comprehension and also between the students’ overall vocabulary knowledge and reading comprehension.

Examining the relationship between vocabulary coverage and reading comprehension in an academic setting in Indonesia, Melani, William, Apgrianto and Lail (2020) tested the vo-

cabulary knowledge in the reading materials of university level students in Indonesia. The study reported that the participants are familiar with approximately 97.7% to 98.7% of the total words in all texts. That is to say, learners' vocabulary coverage estimation ranges from 97.7% to 98.7%. The researchers concluded that knowledge of most of the words used in the texts (98% coverage) enables learners to read the target texts without dictionary assistance.

Although Karakoc and Kose's (2017) study did not directly measure coverage, the study showed that there is a relationship between comprehension and vocabulary knowledge, measured in terms of size. Similarly, Gu (2017) examined the effect of both vocabulary breadth and depth on reading comprehension among EFL university students in China. The study used both multiple choice questions and summary writing to test comprehension. The results revealed that both vocabulary breadth and depth contribute significantly to reading comprehension. Similarly, in a recent study, Ludewig, Hübner, and Schroeder (2023) used a vocabulary test to estimate text coverage and text comprehension among German 4th grade learners of English. The results of the study indicated that there is a high correlation between vocabulary and text comprehension.

While there is abundance of research studies on the relationship between coverage and text comprehension in English, the topic is still under-researched in Arabic. Only a Master's thesis could be found about this topic. Salah, S. (2008) conducted a study about the relationship between coverage and the comprehension of Arabic texts among 23 Arabic (as a foreign language) students at Brigham Young University. The study reported a significant correlation of .6 to .7 for the two authentic media texts used in the study. Similarly, the study also concluded that the participants needed 90% coverage for adequate comprehension of one of the reading texts and 86.2% for adequate comprehension of the other text (defined in this study as 75% comprehension) and 94.2% for complete (100%) comprehension of the text. Two pitfalls seem to make the results of this study questionable. First, it used only 5 questions to test the comprehension of the whole text. This limited number of questions might have led to targeting the same information in more than one question, which would render the test results unreliable. While this study is of significance as it is a pioneer in examining the relationship between vocabulary knowledge and comprehension, the subjects are not speakers of Arabic or even students for whom Arabic is a second language. Therefore, there is a need for further research on the relationship between coverage and reading comprehension among learners of Arabic as a second language in a context where the language is studied from the early years of the primary school.

3 Methodology

3.1 Rationale

This study aims at investigating the relationship between vocabulary coverage and reading comprehension of Arabic texts. It also aims at highlighting the performance of Moroccan

learners in different reading comprehension skills. This will enable language teachers and materials writers to better design their reading materials, as they will have a clear estimation of the amount of vocabulary knowledge that should be familiar for overall comprehension of a reading text in Arabic. Also, because this study deals with specific reading sub-skills, it will enable Arabic language teachers to better identify the types of reading which are problematic for Arabic language learners, hence, re-directing and designing their instruction better. The study shall also provide other educational stakeholders with better knowledge about the role of vocabulary knowledge as one of the aspects which may facilitate or hinder text comprehension. This might also provide some explanation for the performance of Moroccan learners in international reading tests, mainly PISA. To the best knowledge of the researcher, no previous study has examined the issue at hand before. Hence, it is also hoped that this study will open doors to follow-up studies on the interconnection between coverage and reading comprehension.

This study has two main objectives. First, it aims at highlighting the level of vocabulary coverage that is necessary for overall successful comprehension of a reading text in Arabic. Second, it aims at determining the relationship between vocabulary coverage and learners' performance in specific reading skills. To achieve these two objectives, this study answers the following research questions:

1. Is there a relationship between the amount of vocabulary known in a text and the overall comprehension of the text?
2. Is there a relationship between vocabulary knowledge and performance in specific reading comprehension sub-skills?

3.2 Participants

To achieve the objectives of this study, the vocabulary and the reading comprehension tests were administered to Moroccan learners of Arabic of three different grade levels in public schools: primary, middle and high school classes. After cleaning the data, contributions from a total number of 154 participants were finally used as data for this study. During the time of data collection, the participants from primary school were at grade 6, the participants from middle school were at grade 9 and the participants from high school were at grade 10. This study collected data from participants from different school grades in order to ensure different proficiency levels in Arabic.

Table 1 presents the frequencies related to this study's demographics. The total number of participants in this study is 154 students of different grade levels. Among these are 68 males and 86 females. The participants are split over three different grade levels: 68 belonging to the 9th grade, 48 belonging to the 10th grade and 38 6th graders.

Table 1: Study demographics

		Gender * Grade level			Total
		Grade level			
		9th grade	10th grade	6th grade	
Gender	Male	31	20	17	68
	Female	37	28	21	86
Total		68	48	38	154

3.3 Design and data collection procedures

To answer the stated research questions, this study used two tests: A vocabulary test and a reading comprehension test. For the vocabulary test, all the word tokens which appeared in the reading text were manually lemmatized based on the word frequency counts which are generated by the computer software AntConc (Anthony, 2023). This list of word lemmas is made as the basis for the development of a checklist Yes/No vocabulary test. The overwhelming majority of the word lemmas appeared only once in the text. To minimize the level of guessing, an Artificial Intelligence application (ChatGPT) was used to generate pseudo-Arabic words (nonce words) which were added to the items on the Yes/No test. These nonce words were further checked in multiple online Arabic dictionaries to make sure that each generated pseudo-word surely does not exist as an Arabic word. The total number of word lemmas in the vocabulary test is 250, including 20 nonce words (8%).

To avoid guessing the meaning of the items on the test from their context, learners were administered the Yes/No test prior to reading the text. The administration of the Yes/No test took 15 minutes during which learners were asked to work individually and to check Yes/No depending on whether they knew/didn't know each item. When the teachers collected the vocabulary test, they administered the reading comprehension test. During the scoring of the vocabulary test, learners were given one point for each correct word they checked as known. However, 8 point were subtracted from the total score for each 'Yes' tick on a nonce word.

A participant's vocabulary knowledge is calculated as 250 minus the number of real words which are indicated as unknown and also the number of nonce words which are indicated as known, bearing in mind that for each ticked nonce word, 8 points (i.e. 8 known words) are subtracted from the final score. In the results charts below, the total number of unknown words is used to calculate vocabulary knowledge. In other words, a participant's vocabulary knowledge is 250 words minus the number of mistakes as indicated in the results charts below.

The comprehension test is based on a text taken from PISA training documents which was issued and used by Saudi Ministry of Education to train students for the PISA test. The text is about the industrial revolution in Japan and it is composed of 601 word tokens (230 lemmas). Among these are 6 word types of proper nouns (Japan, Germany, Hamburg, Nakazaki, Europe, European). The reading text is of a narrative style as it tells the story of a Japanese student who emigrated to Germany and returned to his country with expertise in car making

industry.

To examine their comprehension of the reading text, learners were asked to respond to a set of 15 multiple choice (MC) questions which targeted different reading skills. Each reading comprehension question is associated with four options, and the test-taker is required to tick only the correct option. The administration of the reading comprehension test was done after the vocabulary test, and it took 45 minutes. Both tests were conducted during the participants' officially scheduled classes.

The reading skills targeted in the comprehension test are the following: skimming for the gist (Identifying main idea), recalling details from the text, understanding word meaning, understanding text structure (and event sequencing), inferring implicitly stated information, understanding cause-effect relationship, understanding the writer's purpose, understanding intended values, understanding feelings, drawing conclusions, distinguishing fact from opinion, distinguishing main idea from detail, distinguishing related from unrelated information, making judgements about an issue or about an idea in the text, and finally making a personal opinion.

3.4 Data analysis

The vocabulary test was scored by assigning learners one point for each real word they ticked as known. However, if a learner ticks a nonce word as known, one point is subtracted from the total score. For the fifteen reading comprehension questions, learners were given one point for each correct choice. Before administering the scores to further statistical analyses, the scores of the participants who didn't finish the whole comprehension or vocabulary tests were discarded. Similarly, we also discarded data from participants with too many wrong guesses on the vocabulary test.

After, tallying the vocabulary and the reading comprehension scores, all the scores were entered into SPSS (Statistical Package for Social Sciences; version 21). To answer our research questions, we first conducted descriptive statistics. Then, correlations were conducted between the vocabulary score as an independent variable and the total reading score at a first level and between the vocabulary score and the isolated reading sub-skill at a second level.

4 Results and discussion

This section provides the statistical analyses which are required to answer the research questions. The first research question examines the relationship between vocabulary knowledge and reading comprehension. In our answer to this question, we look at comprehension from the total score which learners obtained for their performance in reading the text. In other words, it is assumed that the overall score reflects the learners' comprehension of the text in its integrity.

4.1 The relationship between vocabulary coverage and reading comprehension

Before answering the first research question about the relationship between vocabulary coverage and text comprehension, we present the required descriptive statistics. In this section we look at comprehension from a holistic perspective by using the students' overall score which they obtained for answering the MC questions related to the reading text.

Table 2: Descriptive Statistics: Vocabulary score, text coverage and comprehension score

Grade level		N	Minimum	Maximum	Mean	Std. Dev.
9th grade	Vocabulary score	68	0	72	12,82	15,527
	Text coverage		52	100	91,45	10,352
	Total reading score		2	12	6,96	2,216
10th grade	Vocabulary score	48	5	88	38,13	18,726
	Text coverage		41	97	74,58	12,484
	Total reading score		3	13	9,31	2,477
6th grade	Vocabulary score	38	22	132	67,18	29,254
	Text coverage		12	85	55,21	19,502
	Total reading score		3	12	7,21	2,772

Table 2 presents the descriptive statistics which are related to the students' overall text vocabulary score as well as the overall reading comprehension score. The descriptive statistics indicate that the 6th graders don't know a mean number of 67.18 word lemmas (out of 150). This shows that students of this grade level have a text coverage of 55.21%. That is, almost half of the words in the reading text are unknown for this group of learners. For the 9th graders, the vocabulary test shows that they don't know a mean of 12.82 words. This indicates that they have a text coverage of 91.45%, suggesting that they have better scores in vocabulary compared to 6th graders. Finally, the results of the vocabulary test for the 10th graders indicated that they don't know a mean of 38.13 words in the reading text. In terms of text coverage, this means that their vocabulary knowledge covers 74.85% of the words in the text.

Using a one-way between-groups analysis of variance (ANOVA) to check the significance of the differences in text vocabulary knowledge between the three groups, the results indicate that the difference between the groups is statistically significant ($F = 85.89$, $df = 2$, $p < .001$). Using Tukey HSD as a post hoc test, it was observed that the between-group differences are statistically significant for the three comparisons ($p < .001$). The differences between the three groups in terms of vocabulary coverage are also statistically significant between the three groups: ($F = 85.89$, $df = 2$, $p < .001$). These results suggest that the observed differences between the learners of the three groups in terms of the knowledge of the vocabulary of the text are meaningful. It is, however, observed that the differences are not linear as we move from lower to upper grade levels. 9th graders scored better than 10th graders in terms of text coverage/vocabulary. It might be the case that some type of vocabulary knowledge helped

9th graders to demonstrate better vocabulary knowledge compared to their higher group, 10th graders. We are, however, more concerned in this study with the relationship between this reported vocabulary knowledge and reading comprehension. By looking at these coverage scores as a whole, we would expect learners to have a weak comprehension of the text as they are far behind the coverage thresholds which are reported for successful text comprehension in the studies conducted on English language. The highest coverage level is reported for 9th graders (91%) which would make text comprehension almost impossible for the three groups following previous studies which required higher levels of text coverage for comprehension (e.g. Hirsh and Nation, 1992; Laufer, 1989; Nation, 2001, 2006; Schmitt, Jiang, and Grabe, 2011).

Using data on Table 2 above, it can also be observed that the three groups scored just around the average mean in the reading comprehension test. For the 6th graders, they obtained a mean score of 7.21. Similarly, the 9th graders had a mean score of 6.96 and the 10th graders had a score of 9.31. The standard deviation scores are small, which indicates that the performance of learners within the three groups is homogenous. Using an analysis of variance, the results indicate that the difference in the comprehension score between the three groups is significant ($F = 14.33$, $df = 2$, $p < .001$). The results of Tukey post-hoc test showed that the difference between the 9th graders and the 10th graders and also between the 6th graders and the 10th graders were significant. However, the difference between the 6th graders and the 9th graders was not significant. This may suggest that although the descriptive means look not far spaced and are all below 10/15, 10th graders seem to significantly perform better in their reading comprehension compared to both 9th and 6th graders. This better performance seems to oppose our vocabulary results above. The between-group comparisons showed that the 10th graders had a significantly lower text coverage (i.e. more unknown words) at least compared to 9th graders; however, they performed better when it comes to text comprehension. This suggests that comprehension may not be a by-product of better text coverage. This is further explored in the following correlational analysis.

Table 3: Correlations between vocabulary score and comprehension score

Grade level			Text coverage	Total reading score
9th grade	Text coverage	Pearson Correlation	1	,071
		Sig. (2-tailed)		,568
		N	68	68
	Total reading score	Pearson Correlation	,071	1
		Sig. (2-tailed)	,568	
		N	68	68
10th grade	Text coverage	Pearson Correlation	1	,220
		Sig. (2-tailed)		,133
		N	48	48
	Total reading score	Pearson Correlation	,220	1
		Sig. (2-tailed)		,133
		N	48	48
6th grade	Text coverage	Pearson Correlation	1	,474**
		Sig. (2-tailed)		,003
		N	38	38
	Total reading score	Pearson Correlation	,474**	1
		Sig. (2-tailed)	,003	
		N	38	38

As Table 3 indicates, a significantly correlation ($r = .474$, $p < .05$, $R^2 = .22$) between text coverage and text comprehension is observed only for the 6th graders. This suggests that at least for this grade level text coverage is to some extent important for text comprehension. However, it should be noted that the correlation is not strong enough. Although 9th and 10th graders showed excellent knowledge of text vocabulary, the correlational analysis did not show that this vocabulary knowledge translates into better text comprehension as we did not observe any significant correlations for these two groups. It can be understood from these results that better text comprehension may not always be related to better vocabulary knowledge, in its coverage aspect. This might suggest that mobilizing other skills and strategies might be involved in text comprehension. The absence of a (clear) correlation between text vocabulary knowledge and text comprehension is also reflected in the link between coverage and comprehension. This is shown through the following coverage-comprehension chart.

Table 4: Text vocabulary coverage and text comprehension

Grade	Vocabulary score (unknown words)	Text coverage	Reading comprehension score
9th graders	12.82	91%	6.96 (out of 15) (46%)
10th graders	38.13	74%	9.31(62%)
6th graders	67.18	55%	7.21 (48%)

By analyzing the vocabulary-comprehension relationship in terms of text coverage, the mean comprehension scores suggest that all the three groups achieved or approximated only 50% of

text comprehension, considering the 15 reading sub-skills targeted through the 15 MC questions. Even for the 6th graders who showed a significant correlation of $r = .47$ between vocabulary knowledge and reading comprehension score with a coverage of approximately 55% of text vocabulary, their comprehension score is just above average in the reading score (7/15 points), with a text comprehension of only 48%. Similarly, we noted earlier that the differences between the 9th and 10th graders are statistically significant both in their knowledge of the text vocabulary and also in their comprehension of the text. The results on Table 2 above indicate that 10th graders have better comprehension of the text compared to 9th graders although they showed better vocabulary knowledge: 62% comprehension for 10th graders compared to only 46% for 9th graders. A coverage of approximately 91% of the vocabulary has allowed 9th graders to comprehend less than 50% of the text. However, with coverage of approximately 74% of the vocabulary in the text, 10th graders approximated 10 points out of 15 in comprehension (mean = 9.31) and managed to understand 62% of the text. Similar to 9th graders, 6th graders showed a comprehension level of 48% with vocabulary coverage of 55%. These results indicate that none of the three groups achieved complete comprehension of the text. These results are consistent with previous findings about the coverage threshold levels which are required for successful text comprehension. For instance, Hirsh and Nation suggested that 98% as the necessary coverage level at which most learners can successfully read the text. They also reported that 95% is the coverage which is required only for an acceptable comprehension level. Quite similar to the results of our study, Hsueh-Chao and Nation (2000) also reported that at a coverage level of 80% no learner achieved text comprehension while 90% provided only a few learners with adequate comprehension. The study concluded that even a coverage level of 95% is not enough for successful comprehension. Although these results were reported for English reading texts, they seem to be consistent for Arabic reading comprehension. The results of the current study indicate that even at 91% text coverage, learners did not achieve adequate text comprehension as they reached only around 50% of text comprehension. For Laufer and Ravenhorst-Kalovski (2010), even 55% is considered as an unsatisfactory level of comprehension. It might be possible that our learners have weak vocabulary size repertoires, which might have led to inadequate text comprehension even with levels of text coverage as great as 91%.

To conclude this section, it can be maintained that learners' knowledge of text vocabulary does not always follow a linear, progressive pattern as students move from one grade level to the next. It is observed here that although 9th grade learners showed better text vocabulary coverage compared to 6th graders, the two groups' comprehension of the text can be described as below what is expected as they both achieved less than 50% of text comprehension. On the other side, although 10th graders showed weaker vocabulary coverage compared to 9th graders, they achieved better text comprehension (62%). However, their performance too is far from what might be expected of a student who knows approximately all the words in the reading text (74%). This coverage level has allowed them only to comprehend almost two thirds of the text. It might be possible that the 10th graders' better reading performance is the result of better overall vocabulary size. This remains a hypothesis to be explored.

4.2 The relationship between vocabulary knowledge and the reading sub-skills

In this section we look more closely into the relationship between text vocabulary knowledge and the performance of students in fifteen different reading tasks which target different comprehension sub-skills. This section attempts to answer the second research question of this paper:

Research question 2: Is there a relationship between vocabulary knowledge and performance in specific reading comprehension skills?

Similar to what we have done for the first research question, we look at the correlations by considering each group separately. Because of the large number of reading sub-skills targeted in this study, it is practically not appropriate to add a large correlational table for the three grade levels within this paper. Therefore, we will report only the correlating variables without re-producing the tables here.

For the 6th grade, among the fifteen target reading sub-skills, only four showed significant correlations with the overall text vocabulary score. First, ‘understanding text structure and events sequence’, which is also concerned with restructuring the main events as they appeared in the text, correlated significantly with the vocabulary score ($r = .483$, $p < .01$, $r^2 = .23$). Second, ‘drawing conclusions’ also significantly correlated with vocabulary score ($r = .331$, $p < .05$, $r^2 = .10$). The third variable which showed significant correlations with the vocabulary score is ‘distinguishing fact from opinion’ ($r = .453$, $p < .01$, $r^2 = .20$). Lastly, ‘understanding cause-effect relationships’ also showed a significant correlation with the vocabulary score ($r = .516$, $p < .01$, $r^2 = .26$). It is important to highlight here that vocabulary coverage (i.e. proportion of known words in the text) has a significant relationship with the performance of learners in some higher-order thinking reading skills. For this grade level, none of the skills which require comprehension of superficial information such as skimming or recalling text details correlated with knowledge of text vocabulary. Out of 15 reading sub-skills, only four showed weak correlations with text coverage.

For the 9th graders, only two variables (i.e. two reading sub-skills) showed significant correlations with text coverage scores, indicating that probably for this grade level, vocabulary is an important factor in performing the reading sub-skills involved in the tasks. Similar to 6th graders, the variable of ‘understanding text structure and events sequence’ correlated significantly with the vocabulary score ($r = .264$, $p < .05$, $r^2 = .06$). It seems that to attain an understanding of text structure requires knowing the lexical items which signpost transitions and how the writer moves from one main story detail to the next. Second, the variable of ‘understanding intended values’ also significantly correlated with the vocabulary score ($r = .289$, $p < .05$, $r^2 = .07$). This reading sub-skill requires some inferential knowledge on the part of the learners and usually this is built on understanding the meaning of certain key vocabulary items in the text.

For the 10th graders, it is observed that only one reading sub-skill variable has a significant relationship with students' knowledge of text vocabulary. The variable of 'recalling details from the text' correlated significantly with vocabulary knowledge ($r = .359$, $p < .05$, $r^2 = .12$). This significant relationship suggests that at least for this grade level vocabulary is important for understanding details from the text. It appears that vocabulary might be helpful in understanding the details which are stated in the text.

As we noted earlier, the limited number of reading sub-skills which showed a significant relationship with the vocabulary score for each grade level further supports the result we outlined for the first research question. The absence of any correlation with most reading sub-skills and the reported weak correlations (Cohen, 1988) with very few reading sub-skills further reinforces our analysis of the relationship between vocabulary knowledge and students' performance on independent reading sub-skills: Only 7 variables (i.e. reading sub-skills) correlated significantly with the vocabulary knowledge for the three groups (4 for the 6th graders, 2 for the 9th graders and 1 for the 10th graders). These results suggest that the performance of learners in reading comprehension is not strongly related to their knowledge of text vocabulary. Similar to what we have reported for the relationship between text coverage and text comprehension, the lack of a (strong) relationship between text coverage and specific reading sub-skills might indicate that the threshold levels are not high enough to demonstrate any relationship. It might be the case that higher coverage levels would lead to better performance in different reading sub-skills.

To further examine students' performance in reading comprehension, we plotted their performance in the 15 different reading tasks on a Guttman chart. In a desirable performance, we would expect each student to score 1 point for each question related to each reading sub-skill, which would render a score of 154 if we look at each variable from the top to the bottom of the students' list. The analysis of the participants' results indicates that while more than 50% of the students dealt successfully with the skills of 'recalling details from the text', 'distinguishing main idea from detail', 'understanding word meaning', 'understanding text structure and events sequencing', 'understanding cause-effect relationships' and 'making judgements about an issue/idea in the text', less than 50% performed well in the other sub-skills which require the use of inferencing, higher-order thinking skills and also some of those which invite learners to form their own opinions or make personal judgements/evaluations, namely: 'drawing conclusions', 'inferring implicitly stated information', 'explaining feelings and emotions', 'understanding intended values', 'distinguishing fact from opinion', 'understanding the writer's purpose' and 'forming an opinion about the topic of the text'. It is also important to note that although 'skimming' seems to be a skill which doesn't require the use of inferring skills as it rather it depends more on understanding the gist of the text, only 45% of the students obtained a score of 1 for the related question item in the reading test. These results are consistent with Herman (2019), who reported that the correlation between text coverage and comprehension was large only for literal comprehension. This study reported that even for the group of learners who showed text coverage of 90-94% correlations existed only with literal questions. It can be concluded, therefore, that, in general, the per-

formance of learners in reading comprehension sub-skills is also skewed towards the skills which require literal understanding of text information while those which require the use of other inferential skills show weak performance.

5 Conclusion and limitations

This study set out to examine the relationship between text coverage and the comprehension of a written text as well as the relationship between text coverage and specific reading comprehension skills. The results showed that primary, middle and high school learners showed very moderate to high coverage levels. The highest coverage level was 91% for 9th graders. The results of the study showed that none of the groups reached complete or even adequate comprehension of the text in the sense that their comprehension was just around 50 to 60% of the text. These results are consistent with those which were reported for the relationship between coverage and text comprehension in English language. As reviewed in this article, most studies have set a threshold of 98% coverage as the level which might lead to successful text comprehension. Below this level, learners struggled with the text and they showed only unsatisfactory to adequate text comprehension.

The analysis of the relationship between text coverage and performance in specific reading comprehension skills also showed that only a few skills correlated for each group. It was observed that the relationship was weak and that most skills which require inferencing and forming personal opinions or evaluations about the text were the skills which showed the weakest performance among the three groups. It is reported that the coverage levels might also be weak to the extent that they didn't reflect any relationship with the performance in specific reading comprehension skills.

One shortcoming of the current study is that it did not measure the target learners' vocabulary size. It might be possible that the grade levels which are targeted in this study have very low vocabulary size levels, and this has led to the low coverage levels. It is, therefore, important to examine the connection between both vocabulary size, coverage and performance in reading comprehension in future research. Similarly, because most of the reading sub-skills which are targeted in this study are not emphasized in the Moroccan curricula, it is important to examine if this might have affected comprehension, besides the weak coverage levels. The current study relied on a narrative text in its analysis of the relationship between vocabulary knowledge and comprehension. As previous research established the effect of genre in general, it is also important to examine the relationship between coverage and comprehension using different genres such as descriptive, argumentative and expository texts.

References

- Al-Khasawneh, F. (2019). “The impact of vocabulary knowledge on the reading comprehension of Saudi EFL learners”. In: *Journal of Language and Education* 5.3, pp. 24–34. DOI: <http://10.17323/jle.2019.8822> (cit. on p. 39).
- Alavi, M. and I. Akbarian (2012). “The role of vocabulary size in predicting performance on TOEFL reading item types”. In: *System* 40, pp. 376–385 (cit. on p. 39).
- Anthony, L. (2023). *AntConc (Version 4.2.4)*. Tokyo. URL: <https://www.laurenceanthony.net/software> (cit. on p. 42).
- Ech-Charfi, A. (2023). “Word frequency and lexical coverage in English and Arabic”. In: *Journal of Applied Language and Culture Studies* 6.3, pp. 1–19 (cit. on p. 38).
- Ech-Charfi, A. and Y. Bouaddi (2023). “Receptive vocabulary size of Moroccan learners of Modern SA”. In: *International Journal of Arabic Linguistics* 9, pp. 111–130 (cit. on p. 38).
- Feng, Y. and S. Webb (2020). “Learning vocabulary through reading, listening, and viewing: Which mode of input is most effective?” In: *Studies in Second Language Acquisition* 42.3, pp. 499–523. DOI: <http://10.1017/S0272263119000494> (cit. on p. 38).
- Gu, T. (2017). “The effect of vocabulary knowledge on Chinese English learners’ reading comprehension”. In: *International Journal of English Linguistics* 7.4, pp. 45–55. DOI: <https://doi.org/10.5539/ijel.v7n4p45> (cit. on p. 40).
- Güngör, F. and D. Yaylı (2016). “The interplay between text-based vocabulary size and reading comprehension of Turkish EFL learners”. In: *Educational Sciences: Theory & Practice* 16, pp. 1171–1188 (cit. on p. 39).
- Hirsh, D. and P. Nation (1992). “What vocabulary size is needed to read unsimplified texts for pleasure?” In: *Reading in a Foreign Language* 8, pp. 689–696 (cit. on pp. 37, 38, 45).
- Hsueh-Chao, M. H. and P. Nation (2000). “Unknown vocabulary density and reading comprehension”. In: *Reading in a Foreign Language* 13.1, pp. 403–430. DOI: <https://doi.org/10.26686/wgtn.12560354> (cit. on pp. 37, 38, 47).
- Karakoc, D. and G. Kose (2017). “The impact of vocabulary knowledge on reading, writing and proficiency scores of EFL learners”. In: *Journal of Language and Linguistic Studies* 13.1, pp. 352–378 (cit. on pp. 37, 40).
- Laufer, B. (1989). “What percentage of text lexis is essential for comprehension?” In: *Special language: From humans thinking to thinking machines*. Ed. by C. Lauren and M. Nordman. Multilingual Matters, pp. 316–323 (cit. on pp. 38, 45).
- (1992). “How much lexis is necessary for reading comprehension?” In: *Vocabulary and Applied Linguistics*. Ed. by P. Arnaud and H. Béjoint. McMillan, pp. 126–132 (cit. on p. 37).
- Laufer, B. and G. C. Ravenhorst-Kalovski (2010). “Lexical threshold revisited: Lexical text coverage, learners’ vocabulary size and reading comprehension”. In: *Reading in a Foreign Language* 22.1, pp. 15–30 (cit. on pp. 37, 38, 47).
- Ludewig, U., N. Hübner, and S. Schroeder (2023). “Vocabulary, text coverage, word frequency and the lexical threshold in elementary school reading comprehension”. In: *Read*

- Writ* 36, pp. 2409–2431. DOI: <https://doi.org/10.1007/s11145-022-10385-0> (cit. on pp. 37, 40).
- Nation, P. (2001). *Learning Vocabulary in Another Language*. Second edi. Cambridge University Press (cit. on pp. 37–39, 45).
- (2006). “How large a vocabulary is needed for reading and listening?” In: *Canadian Modern Language Review* 63, pp. 59–82 (cit. on pp. 37, 45).
- Qian, D. D. (1999). “Assessing the roles of depth and breadth of vocabulary knowledge in reading comprehension”. In: *The Canadian Modern Language Review* 56, pp. 282–307 (cit. on p. 37).
- (2002). “Investigating the relationship between vocabulary knowledge and academic reading performance: an assessment perspective”. In: *Language Learning* 52, pp. 513–536 (cit. on p. 37).
- Salah, S., M. (2008). “The Relationship Between Vocabulary Knowledge and Reading Comprehension of Authentic Arabic Texts”. Ph.D. dissertation. Brigham Young University (cit. on p. 40).
- Schmitt, N., X. Jiang, and W. Grabe (2011). “The percentage of words known in a text and reading comprehension”. In: *The Modern Language Journal* 95.1, pp. 26–43. DOI: <https://doi.org/10.1111/j.1540-4781.2011.01146.x> (cit. on pp. 39, 45).
- Schmitt, N., D. Schmitt, and C. Clapham (2001). “Developing and exploring the behaviour of two new versions of the Vocabulary Levels Test”. In: *Language Testing* 18.1, pp. 55–88. DOI: <https://doi.org/10.1177/026553220101800103> (cit. on p. 39).
- Stæhr, L. S. (2008). “Vocabulary size and the skills of listening, reading and writing”. In: *Language Learning Journal* 63.2, pp. 139–152 (cit. on p. 37).