

Measuring Academic Vocabulary Breadth and Depth in Upper-Elementary School Students

Vedrana Gnjiđić, Gordana Keresteš and Anita Peti-Stantić
Faculty of Humanities and Social Sciences, University of Zagreb

Abstract

This study analyzes a newly developed measure for assessing the knowledge of general academic vocabulary in upper-elementary school students in Croatia. This researcher-designed instrument assesses students' command of synonymy and the limits of collocational use as a proxy for academic vocabulary breadth and depth. Elementary school students in the 5th, 6th, and 7th grade participated in the study (N = 152, 69,7% girls). In addition to the newly developed measure for assessing general academic vocabulary, students completed the Mill Hill Vocabulary Scales, a standardized measure of general vocabulary knowledge. The results show that the newly developed instrument for measuring general academic vocabulary knowledge is reliable and valid—it has adequate internal consistency, is sensitive to the expected age and gender differences in vocabulary, and correlates highly with the results obtained by the Mill Hill Vocabulary Scales. This new instrument could fill the gap in academic vocabulary-related research since the currently available vocabulary assessment measures in Croatian do not focus on this particular register. More specifically, our new measure could be used to evaluate various types of academic vocabulary interventions in schools.

Key words: *collocations; general academic vocabulary; synonymy.*

Introduction

During elementary education, children are extensively exposed to new words. Clark (1995, p. 393) reported on estimates made for the English language—from the age of ten onwards, the total number of words children are exposed to in the school context reaches up to 10 000 new words during each school year.

In the context of education, texts are the main form of instruction and knowledge transfer, be it in the form of students' notes in notebooks, writings on the whiteboard,

or texts in textbooks. Upper-elementary school students are expected to use textbook texts more extensively and independently than lower-elementary school students. Also, they mostly learn from (often lengthy and relatively complex) expository texts (Williams, 2018, p. 1924). Starting from the assumption that textbook texts serve as building materials for developing knowledge about the world, it is clear that students can make sense of the content of the text they are reading only if they are able to understand most of the words in the text (Peti-Stantić, 2019, pp. 169, 198). According to the 2015 and 2018 PISA test results, the level of reading literacy of 15-year-old Croatian students is statistically significantly below OECD average (whereby girls outperform boys). In the 2015 cycle, Croatian students ranked 31st, and in 2018 they took the 29th place among the OECD countries (OECD 2016, p. 149, 2019, p. 97). Such results indicate that, until a systematic instructional change occurs, Croatian students could benefit from various types of reading literacy directed interventions. Interventions focused on vocabulary development are put forward as one of the crucial components of interventions directed at the development of reading literacy-related abilities (Hawkins et al., 2011, p. 61; Lesaux et al., 2010, p. 198; Proctor et al., p. 75).

Most of the content presented in textbooks is explained in a formal language register which is expected to be mastered by most students during their education. Besides specific vocabulary, this register is characterized by both syntactically and semantically demanding structures (Feilke, 2012, p. 8). According to Hiebert and Lubliner (2008, p. 110), there are four types of vocabulary used in school texts and tasks: content-specific academic vocabulary, general academic vocabulary, school-task vocabulary, and literary vocabulary. Words belonging to content-specific vocabulary belong to a specific knowledge area, such as chemistry, mathematics or language arts (e.g., *redox*, *multiple* (e.g. of ten), *adverb*), while school-task vocabulary is typically used in test instructions and practice tasks given to students in school (*underline*, *categorize*). Literary vocabulary is specific to literary texts and includes words such as *aghost* or *gingerly*, while words such as *partial*, *suggest* or *cause*, which are used across school subjects and when talking about the abstract dimensions of life, belong to the realm of general academic vocabulary.

The cross-domain presence of words belonging to general academic vocabulary points to the importance of mastering them, i.e. if students do not understand those words and the potential of their use, then they cannot understand the texts in textbooks from which they are supposed to study (Peti-Stantić, 2019, pp. 169, 198). In the literature, words belonging to academic vocabulary are characterized as non-concrete, low in imagery and relatively low in frequency (Corson, 1997, p. 696). Since the process of vocabulary development in higher grades of elementary schools is predominantly seen as automatic, academic vocabulary is not explicitly taught to students in Croatia. However, that goes against the findings that the word-learning process is not always automatic. Research shows that, on average, students in grades 4 to 11 incidentally

learn only 15% of the unknown words encountered while reading texts on their own (Swanborn & de Glopper, 1999, p. 272). This is the reason why we have decided to analyze upper-elementary students' knowledge of general academic vocabulary and to develop an instrument which could be used in evaluating future interventions aimed at advancing the understanding and use of general academic vocabulary in Croatian students. One of the conclusions of the National Reading Panel of the USA (2000, pp. 4-24) is that researcher-based vocabulary measures tailored according to the specific focus of individual interventions are more sensitive to detecting the effects of vocabulary training than standardized vocabulary measures. In accordance with that conclusion, we aimed to develop an instrument which would sensitively and reliably measure the development of general academic vocabulary, i.e., the breadth and depth of academic vocabulary knowledge for the purpose of measuring the effectiveness of school interventions focused on general academic vocabulary. We use the terms vocabulary breadth and depth even though there is no final consensus on their definitions in the relevant literature. Anderson and Freebody (1979, p. 22) define the concept of vocabulary breadth as the number of words we know (this knowledge only includes the basic levels of meaning), while vocabulary depth refers to the quality of knowledge about words. We accept these definitions while further specifying them in line with Killian's interpretation of vocabulary breadth and depth (2016, p. 133) where vocabulary breadth is perceived as encompassing the knowledge about the prototypical context and the structure of words which are formed from the same word root, and vocabulary depth as including the denotative and connotative meanings, associative semantic potentials, and metonymical and metaphorical extensions.

Building upon Read's (2000, p. 185) componential, i.e. multidimensional approach to vocabulary depth, Schmitt (2014, p. 915) describes vocabulary depth as the knowledge of several lexical dimensions of words, some of which include knowing the multiple meanings of polysemous words or the existence of rich associative word networks, which are a part of the holistic approach to word meaning. He operationalizes vocabulary breadth as the sum of all lexical units we are familiar with, i.e., which we have mastered to some extent (Schmitt, 2014, p. 942). The ongoing discussion among scientists about the details regarding the conceptualization of vocabulary breadth and depth indicates that these two dimensions most likely overlap to some extent, just as it is the case with other psycholinguistic constructs such as word abstractness, concreteness, and imageability. For example, Schmitt (2014, p. 942) says that vocabulary breadth always partially encompasses vocabulary depth, and that vocabulary depth can only be defined as an interwoven network of various distinct aspects of word meaning.

These unclear boundaries are also reflected in the use and interpretation of different vocabulary measures. Vocabulary measures are most frequently divided into those assessing receptive vocabulary and those assessing productive vocabulary (Pearson et al., p. 284). Productive vocabulary knowledge is usually assessed through speaking or writing skills, i.e., through tasks which include spontaneous and active use of

words (Pignot-Shahov, 2012, p. 39). Traditionally, receptive vocabulary measures rely on tasks such as providing a definition for a word (e.g., Mill Hill Vocabulary Scales, Raven, 2000), choosing the correct definition of a word (e.g., The Vocabulary Size Test, Nation & Beglar, 2007), or matching the correct word with a definition (e.g., Inter-American Series, Manuel, 1966). Alongside word definitions, Cronbach (1942) and Richards (1976) proposed several components of word knowledge to be used in diagnostic vocabulary testing, such as recognizing the multiple meanings of a word and the limitations connected to its use. Cronbach defines knowing a word as complete only if one is able to both understand and use a word appropriately, while Richards underlines the importance of developing rich associative links among words in our mental lexicons. These associative links include relations of synonymy, antonymy, supra-, co-, or subordination, as well as collocations. Nation (2001, p. 27) summarizes the components of word knowledge and groups them into three dimensions: knowledge of word form (spoken, written, morphological), meaning (the connection between form and meaning, concept and referent, associations) and use (semantic, syntactic and pragmatic limitations of use).

The way in which certain task types are going to be used and interpreted in different vocabulary studies depends on the way in which the authors conceptualize vocabulary knowledge, i.e., on the definitions of vocabulary breadth and depth they choose to accept. For example, Quellette (2006, p. 558) defines vocabulary breadth as the total number of words which exist in our mental lexicon and operationalizes it through a picture naming task (by using two sub-scales of The Test of Word Knowledge), as it is most frequently the case (e.g., The Peabody Picture Vocabulary Test used in Logan & Kieffer, 2017). Even though this type of task is suitable for testing vocabulary knowledge of many different words, academic vocabulary mostly consists of abstract words which makes picture naming inappropriate for operationalizing vocabulary breadth in this register.

For assessing vocabulary depth, Quellette (2006, p. 558) used synonym recognition tasks and word definition tasks because he believes that these tasks enable measurement of the quality of word knowledge. The word definition task is also used as a measure of vocabulary depth by Li and Kirby (2014), while Read (1993) decided to use the synonym recognition task (paradigmatic relations) for the same purpose, alongside tasks which include choosing correct collocations (syntagmatic relations) and recognizing semantic associations (analytical relations).

Even though both Quellette and Read categorize the task of synonym recognition as a measure of vocabulary depth, in some recent studies this task is used as a measure of vocabulary breadth, or the measure of basic word knowledge (e.g., The Vocabulary Subtest, a part of the Gates-MacGinitie Reading Test in Townsend et al., 2012; Li & Kirby, 2014).

Schmitt (2014, p. 941) concludes that the operationalization of the construct of vocabulary depth is always incomplete, regardless of the tasks used to measure it, because

every instrument will always measure only a limited number of segments which form the whole, i.e., the quality of word knowledge, and never the totality of that quality.

This is why, in this study, we chose those tasks for which we believe encompass all the important dimensions of vocabulary breadth and depth. We accept the claim that all tasks which measure vocabulary breadth partially capture some level of vocabulary depth. However, we think that individual tasks are still dominantly oriented towards one of the two dimensions, i.e., vocabulary breadth or depth. Therefore, we chose the synonym recognition task as a measure of (predominantly) vocabulary breadth because this task requires the students to compare words on the level of the wholeness of their individual meanings. The task of recognizing plausible collocations was taken as a measure of vocabulary depth because it requires the students to observe the words from the perspective of limitations of their interactions, both semantic and syntactic.

A dimension often disregarded in vocabulary assessment is the differentiation of vocabulary knowledge in different word classes. The importance of addressing different word classes in vocabulary testing is supported by the findings of psycholinguistic studies. For example, nouns are recognized faster than verbs in lexical decision tasks (e.g., Cordier et al., 2013, p. 28). This is why we put emphasis on this distinction in our study. Also, nouns belonging to general vocabulary are rated as more concrete than verbs from the same register, and verbs as more concrete than adjectives (Peti-Stantić et al., 2018, p. 104).

In this context it is important to emphasize the differences in conceptualization of concrete and abstract words. Wiemer-Hastings and Xu (2005, p. 724) describe abstract concepts as relational because they are mostly observed in relation to and comparison with other concepts, while concrete concepts are usually conceptualized through their intrinsic properties or characteristic features. Also, abstract concepts are relatively unspecified in comparison with concrete concepts because their meaning is often molded according to the situation, i.e., the context in which they are used (Wiemer-Hastings & Xu, 2005, p. 721).

Vocabulary belonging to the academic register is characterized by a higher level of abstractness when compared to the general language register. We can assume that the difference in concreteness between different word classes in the general language register is present in the academic register as well, although not necessarily in the same ratios. More specifically, if concreteness ratings in the general language register are higher for nouns than for verbs, and higher for verbs than for adjectives, it is possible that, precisely due to the abstractness of the academic register and the relational nature of verbs as a word class, verbs belonging to the academic register are more abstract than adjectives, and adjectives more abstract than nouns in that register.

Besides the abstractness of the register, which additionally emphasizes the differences between word classes, one must also take into account their different linking potentials at the level of deep meaning structures. Those linking potentials expose the difference in their complexity - nouns and adjectives connect on a linear level, while verbs connect

according to their valencies. The relational nature of verbs additionally separates them from nouns and adjectives. Therefore, the types of relations which are assumed by nouns, adjectives and verbs in accordance with their semantic, syntactic and pragmatic limitations are the factor which makes the use of identical task types for the assessment of vocabulary breadth and depth for each word class impossible.

Even though the tasks in instruments assessing general vocabulary knowledge, such as the Mill Hill Vocabulary Scales, contain several word classes (nouns, verbs, adjectives and adverbs), those differences are not addressed in any way. This is why we constructed an instrument for measuring general academic vocabulary breadth and depth in which we specifically focus on different word classes despite the limitations of their use which are a result of their internal semantic and syntactic composition. To this end, nouns are assessed through the relation of synonymy, adjectives and verbs through the limitations of collocational use, i.e., strong and weak collocations.

The National Reading Panel of the USA (2000, pp. 3-4) summarized the most frequent vocabulary instruction methods. Those methods include explicit vocabulary instruction (oriented towards definitions), but also implicit vocabulary instruction relying on extensive reading, multimedia methods (focused on multisensory learning), capacity methods for younger learners which are focused on the automation of the reading process, and association methods which encourage students to create associations between new and existing knowledge. One of the conclusions presented in the report of the National Reading Panel of the USA (2000, pp. 4-27) is that students' age and abilities influence the effectiveness of any of those individual vocabulary instruction methods. Therefore, all interventions using any of these methods of vocabulary instruction should employ instruments allowing them to assess the effects of the instructional process in different age groups and with students of different abilities.

There are several diagnostic instruments measuring general vocabulary knowledge available in Croatian – the Croatian version of the Peabody Picture Vocabulary Test (PPVT-III-HR, Dunn et al., 2009), the Croatian version of the Mill Hill Vocabulary Scales (MHV, Raven et al., 1994, in Križan & Matešić, 2001) and a translation of Manuel's test from the Inter-American series, which is used by some researchers (e.g., Didović & Kolić-Vehovec, 2009; Pečjak et al., 2009). Although the constructs of general vocabulary knowledge and the knowledge of general academic vocabulary are related and the measures of both constructs correlate highly (e.g., Masrai & Milton, 2018, p. 50, $r = .78$), research shows they are distinct. For example, Townsend et al. (2012, p. 513) show that the results on a measure of general academic word knowledge explain an additional portion of variance in the results on standardized measures of academic achievement, even when general vocabulary breadth of the participants was controlled for. Since items for measures of general academic vocabulary are pooled from textbook texts and academic texts and articles, such instruments can be used for measuring the specific effects of interventions focused on developing general academic vocabulary comprehension and use. Concerning the fact that the abovementioned vocabulary

measurement instruments available in Croatian language measure general vocabulary knowledge, we have identified the need to develop and pilot a new instrument which could be used to measure general academic vocabulary in the context of Croatian upper-elementary school education. Based on reading development theories (Indrisano & Chall, 1995), we conclude that in upper-elementary school grades it is crucial to develop the students' lexical resources in terms of academic vocabulary, while in lower-elementary school grades the focus should be on the automation of the reading skill itself. During high school education, the focus can shift towards enriching the fundamental academic vocabulary knowledge in the mother tongue which was consolidated during elementary education and move towards more complex strategies of independent vocabulary enrichment and text comprehension strategies.

The Study

As already mentioned, Croatian students' results of international reading literacy assessments indicate a need for various types of reading literacy directed interventions. To our knowledge, there are no systematic interventions conducted in Croatian schools which focus on developing students' (academic) vocabulary or reading literacy, just as there are no instruments in Croatian language measuring academic vocabulary knowledge. Since a valid instrument measuring academic vocabulary knowledge is a prerequisite for assessing the effect of any future school intervention targeting academic vocabulary development in students, we decided to design one such measure of general academic vocabulary knowledge within our MEGHR project. We created a preliminary version of an instrument which would measure our specific areas of interest—breadth and depth of vocabulary knowledge of nouns, verbs and adjectives belonging to general academic vocabulary. The results of testing this preliminary version are presented in this paper. If that version of the instrument proves to be valid and reliable, the instrument can be developed and used in future research as a measure of general academic vocabulary knowledge and for assessing the effects of different approaches to teaching academic vocabulary.

We expect our measurement instrument to be sensitive to age, i.e., school grade, because, as their education progresses, students gather more experience with more complex concepts and words used to express those concepts. Based on the aforementioned assumption that verbs belonging to general academic vocabulary are more abstract than adjectives and nouns belonging to that same register, we expect the effect of word class. More specifically, we expect that students will be most successful in solving the subscale with nouns, and least successful in solving the subscale with verbs on our measure of general academic vocabulary knowledge.

In accordance with the maturation-related development of cognitive abilities, it is also possible to expect an interaction of those two factors: the effect of word class can be different at different ages, i.e., in different school grades, and the effect of grade can be different for each word class.

Additionally, we expect that our instrument will be sensitive to gender differences which have often been reported in vocabulary studies, with girls more frequently having higher results on different vocabulary measures than boys (e.g., Bornstein et al., 2004, p. 296; Eriksson et al., 2012, pp. 337-338; Kaushanskaya et al., 2013, p. 86).

Students' results on our new instrument measuring general academic vocabulary knowledge will be compared with their results on the Mill Hill Vocabulary Scales, a standardized measure of general vocabulary knowledge. That way we will inspect the convergent validity of the measure of general academic vocabulary knowledge (GAVM) that we propose in this paper. We expect that students' GAVM results will moderately to highly positively correlate with their results on the Mill Hill Vocabulary Scales.

Finally, we expect that GAVM will demonstrate adequate internal consistency, indexed by Cronbach's alpha coefficient.

Method

Participants

Our participants were upper-elementary school students (5th, 6th and 7th grade) from four different schools in the Varazdin, Medjimurje and Koprivnica-Krizevci counties (N=152). The number of boys and girls participating in the study was not balanced (Table 1).

Table 1
Sampling structure by grade and gender

	Girls	Boys	Missing data	Total
5th grade	29	26	2	57
6th grade	38	10	3	51
7th grade	35	8	1	44
Total	102	44	6	152

One aspect influencing the number of girls and boys participating in this study might be the traditional perception of reading and reading-related activities as more appropriate for girls than boys. Since participation in the study was voluntary, fewer boys were motivated to participate owing to the topic of the study. Additionally, research shows that girls usually have a higher level of reading motivation (McGeown et al., 2015, p. 561), while boys tend to assess their reading skills as average or below average and spend less time reading (Usen & Mürsepp, 2012, p. 1799). Based on such reading-related attitudes, one can expect that boys will have a reduced level of motivation for participating in a study in which researchers will measure vocabulary knowledge. Since vocabulary measurement instruments are similar to a school exam, boys might a priori perceive participating in vocabulary testing as a situation in which they will not be successful and therefore be more reluctant to participate even though they know that their results will not be connected with their identity and will have no effect on their school success.

Instruments

Researcher-designed general academic vocabulary measure, GAVM

The general academic vocabulary measure (GAVM) analyzed in this study was constructed by the last author of this paper. The first phase of construction consisted of pooling the lexemes from the most frequently used textbooks in Croatian schools in Croatian language, mathematics, history, geography, and science classes for the 4th, 5th, and 6th grade. Several university teachers and students of linguistics and psychology identified 3000 highly frequent general academic vocabulary words from those textbooks which were deemed more difficult to understand or less familiar to the age group for which the textbooks were written. In the second phase of construction, this list was reduced to 1500 words. The third phase included further narrowing-down with the help of elementary school teachers – core lists containing 100 most frequently used words were formed for every grade level from which the textbooks were consulted. Thereby a total of 300 words was selected. The fourth stage consisted of choosing 20 nouns, 20 adjectives and 20 verbs from that abridged list. Everybody included in the final selection agreed that all students should know the meaning and use of those 300 words after finishing elementary school because those words are broadly used both in public and in academic discourse. Those words were used to form the tasks tapping into the breadth and depth of general academic vocabulary knowledge which form the GAVM.

In its preliminary form GAVM consists of 60 items which constitute the tasks for measuring vocabulary breadth and depth. A correct response to each item is scored with one point. The first part of the GAVM focuses on vocabulary breadth, i.e., on assessing the ability to find synonyms to individual nouns (20 items) among five response options. Special attention was paid to the choice of response distractors on the basis of word formation, semantic or phonological similarities with each initial noun. Table 2 shows an example item from that subscale.

Table 2
Example item: synonym identification

Example item— nouns (translation)	pothvat (endeavor)	rukohvat (handrail)	akcija (action)	prihvat (shelter)	učinak (effect)	neuspjeh (failure)
---	-----------------------	------------------------	--------------------	----------------------	--------------------	-----------------------

The words “rukohvat” (Eng. handrail) and “prihvat” (Eng. shelter) were chosen based on word formation and phonological similarity with the initial noun. The Croatian word for failure can be semantically connected with the initial noun because an endeavor can be both successful and unsuccessful. Also, an endeavor of some kind can cause a certain effect (“učinak”) but those two words are still not synonyms. The correct answer for this item is the Croatian word for action (“akcija”).

The second part of the GAVM consists of two subscales of the same type—identifying unlikely (or impossible) collocations for adjectives and verbs. Each subscale consists of 20 items. Five response options are offered for each item and only one of those

response options can be identified as the unlikely/impossible collocation. This task type was chosen as a measure of vocabulary depth because it required students to recognize the limitations of word use. Table 3 shows example items from this task type.

Table 3

Example item: adjective and verb collocations

Example item— adjectives (translation)	neznatna (negligible)	razlika (difference)	pogreška (mistake)	količina (amount)	udaljenost (distance)	kombinacija (combination)
Example item— verbs (translation)	ugroziti (to jeopardize)	život (life)	bolest (illness)	prijateljstvo (friendship)	zdravlje (health)	sigurnost (safety)

In these two tasks, correct responses and distractors were chosen in such a way to represent the semantic nuances of word use as diversely as possible, whereby both the denotative and the connotative level of meaning were taken into consideration.

The adjective “neznatan” (Eng. negligible) refers to the barely noticeable distance from the intended or expected goal of something we can measure. For example, we can say that the difference (“razlika”) between two candidates at an exam was negligible, that a cake contains only a negligible amount (“količina”) of sugar, or that the distance (“udaljenost”) between any two points on our horizon is negligible. However, a combination (“kombinacija”) cannot be described in terms of size, but in terms of quality - good, familiar, unexpected, etc. This is one of the limitations of word use that students have to recognize in order to successfully solve this task.

The verb “ugroziti” (Eng. to jeopardize) points to the fact that something is put in danger. This verb is usually used with words that carry a positive connotation. One can jeopardize someone’s life, health, safety, friendship, i.e., by jeopardizing something, we are negatively influencing our own or other people’s wellbeing, weakening our health or the strength of our friendship. For the same weakening effect on something negative such as illness, we will not use the verb “jeopardize”, but rather say that we have “defeated” or “eased” it. Students should recognize the limitations of use of the initial verb in terms of standing in combination with positively connoted nouns.

Each of the three subscales will be analyzed separately in the following part of the paper.

Mill Hill Vocabulary Scales

The Mill Hill Vocabulary Scales (MHV, Raven et al., 1994, in Križan & Matešić, 2001) are available in two forms. Both forms consist of two subscales, each containing one set of words. Each set of words appears on both subscales in parallel – in one test form, one set of words is in the word definition task, i.e., subscale, and in the other test form that set of words is on the subscale of identifying synonyms to items among six response options. One set of words consists of 14 verbs, 9 nouns, 7 adjectives and 2 adverbs, and the other set contains 12 verbs, 7 nouns, 12 adjectives and one adverb. In

other words, each subscale is composed of 32 items and it includes nouns, adjectives, verbs, and adverbs. A correct response to each item is scored one point. In this study, Cronbach's alpha coefficients for the first form of the MHV were .75 for the subscale of word definitions and .86 for the subscale of synonym identification, while in the second form of the MHV, Cronbach's alpha was .86 for the subscale of word definitions and .74 for the subscale of synonym identification. Cronbach's alpha coefficient for the total result on both forms of the MHV was .88.

Procedure

The study was approved by the ethics committees of the Psychology Department at the Faculty of Humanities and Social Sciences (University of Zagreb) and the Ministry of Science and Education of the Republic of Croatia. Prior to data collection, a signed parental consent was obtained for each student. On the day of data collection, in addition to signed parental consent, students also gave their assent to participate in the study. None refused to participate.

Participants first completed the researcher-designed GAVM, assessing general academic vocabulary breadth and depth, and then the aforementioned standardized measure of general vocabulary knowledge, the Mill Hill Vocabulary Scales. Students completed both instruments anonymously, in paper form. The order of completing them was the same for all students. No specific time limit was set for completing the measures, but most children completed the GAVM, as well as the MHV, in 15 to 20 minutes. After completing the two vocabulary measures, students participated in workshops we have prepared providing them and their teachers with examples of exercises for vocabulary development which can be used by the students independently or implemented in class.

Results

IBM SPSS v. 26 was used for data analysis. Descriptive statistics of students' results on the GAVM and on MHV are presented in Table 4. Results are separated by grade and subscale. The highest possible result on each of the three GAVM subscales (noun synonyms, adjective collocations, and verb collocations) is 20 points, and the maximum possible total GAVM result is 60 points. The highest possible result on the individual MHV subscales (word definitions and synonym identification) is 32 points, and the maximum possible total result on the MHV is 64 points.

Normality of distributions of students' results on the GAVM and MHV by grade was inspected visually by observing shapes of distributions and additionally by calculating the z-scores for skewness and kurtosis (Field, 2009; Petz et al., 2012). Visually, all shapes of distributions had a normal curve. The z-scores for skewness (Table 4) revealed negative skewness of all distributions, except for the distribution of 5th grade students' results on the subscale of word definitions on the MHV, which indicated positive skewness. On the GAVM, the negative skewness of the distribution of 5th grade students' results

is statistically significant (higher than +/- 1.96) only for the subscale of adjective collocations. In the 6th grade, negative skewness of the distribution is statistically significant for the subscale of noun synonyms and the total result. Negative skewness of the distribution for the 7th grade is statistically significant for all subscales and the total score. On the MHV, negative skewness is statistically significant for the distribution

Table 4

Descriptive statistics for the General academic vocabulary measure and the Mill Hill vocabulary scales by subscale and grade

			M	SD	Min	Max	z_s	z_k
GAVM	5 th grade	GAVM-N	12.65	2.44	7	17	-0.67	-0.89
		GAVM-ADJ	11.37	4.31	1	18	-3.43	0.54
		GAVM-V	9.09	4.39	1	18	-0.65	-1.03
		GAVM-TOT	33.11	9.43	12	48	-1.31	-0.73
	6 th grade	GAVM-N	14.06	3.85	1	19	-4.48	4.47
		GAVM-ADJ	13.61	2.74	6	19	-1.93	0.62
		GAVM-V	11.61	3.42	3	18	-1.22	0.09
		GAVM-TOT	39.27	8.81	16	54	-2.49	0.59
	7 th grade	GAVM-N	15.66	2.87	2	20	-7.24	16.25
		GAVM-ADJ	14.80	2.94	0	19	-8.30	20.49
		GAVM-V	14.09	2.57	6	18	-3.07	2.56
		GAVM-TOT	44.55	5.92	28	56	-2.65	1.75
MHV	5 th grade	MHV-D	17.09	4.37	8	28	0.79	-0.13
		MHV-S	20.53	5.59	6	32	-1.28	-0.29
		MHV-TOT	37.61	8.66	18	60	-0.49	-0.01
	6 th grade	MHV-D	21.82	5.14	10	32	-0.62	-0.67
		MHV-S	23.88	3.75	14	30	-2.69	1.43
		MHV-TOT	45.71	7.95	24	62	-1.36	0.23
	7 th grade	MHV-D	21.91	5.09	6	29	-2.58	1.31
		MHV-S	26.02	3.36	19	32	-0.92	-0.45
		MHV-TOT	47.93	7.49	25	61	-2.55	1.18

Legend. GAVM = General academic vocabulary measure; GAVM-N = subscale of noun synonyms on GAVM; GAVM-ADJ = subscale of adjective collocations on GAVM; GAVM-V = subscale of verb collocations on GAVM; GAVM-TOT = total result on the GAVM scale; MHV = Mill Hill vocabulary scales; MHV-D = subscale of word definitions on MHV; MHV-S = subscale of synonym identification on MHV; MHV-TOT = total result on MHV; M = mean result; SD = standard deviation; Min = the lowest score; Max = the highest score; z_s = z-score for skewness; z_k = z-score for kurtosis

of 6th grade students' results on the subscale of synonym identification and for the distribution of 7th grade students' results on the subscale of word definitions and the total result. The z-scores for kurtosis (Table 4) are not statistically significant (higher than +/- 1.96) for the results of 5th and 6th grade students, except for 6th grade students' results on the subscale of noun synonyms on the GAVM which have a leptokurtic distribution. 7th grade students' results have a leptokurtic distribution on all three

subscales of the GAVM, but the kurtosis of the distribution of the total result on the GAVM does not deviate from a normal curve.

Since the inspection of the normality of distributions has indicated that the distributions of some variables statistically significantly deviate from a normal curve, we have analyzed the data by using both parametric and nonparametric procedures and have subsequently compared their results. The results of parametric and nonparametric analyses did not differ, which we explain in more detail in further description of the results.

Testing the effects of gender

The effects of gender on the total result and the results on all individual subscales of the GAVM were tested separately from the effects of age, i.e., school grade, since the number of 6th and 7th grade girls and boys participating in this study was rather uneven. Levene's test was used to assess the equality of variance in the samples of girls and boys. The results indicated that equal variances of the results on the GAVM cannot be assumed, i.e., that there is a statistically significant difference in the variance in the samples of girls and boys on the subscales of adjective and verb collocations (the resulting values of Levene's test for noun synonyms, adjective collocations, verb collocations and the total result were, respectively, $F = 3.35$, $p = .069$; $F = 5.05$, $p = .026$; $F = 4.74$, $p = .031$, and $F = 3.78$, $p = .054$; in all tests $df_1 = 1$, $df_2 = 144$). Due to the inequality of variances, gender differences in the results on the GAVM were first inspected by the non-parametric Mann-Whitney test, and subsequently by independent samples t-test. The results of all conducted tests were statistically significant, indicating that girls achieved statistically significantly better results on all subscales of the GAVM and in total. Since the results of the Mann-Whitney test and the t-test did not differ, we will only report on the results of the t-test.

On the subscale of noun synonyms, girls achieved a mean result of 14.39 (SD = 3.13), and boys 13.02 points (SD = 3.69); $t(144) = 2.29$, $p = .023$; Cohen's $d = 0.40$. On the subscale of adjective collocations, girls had a mean result of 13.57 (SD = 3.39) while boys scored an average of 11.91 (SD = 4.33); $t(144) = 2.26$, $p = .027$; Cohen's $d = 0.42$. On the subscale of verb collocations, girls achieved a mean result of 12.27 (SD = 3.70) and boys 9.45 (SD = 4.62); $t(144) = 3.58$, $p = .001$; Cohen's $d = 0.60$. The total mean result of girls on the GAVM was 40.24 (SD = 8.66) and boys 34.39 (SD = 10.66); $t(144) = 3.49$, $p = .001$; Cohen's $d = 0.60$. Cohen's d values indicate that the effect of gender on GAVM results is of medium size.

Testing the effects of grade and word class

The effects of grade and word class on GAVM results were tested by the non-parametric Kruskal-Wallis (the effect of grade) and Friedman test (the effect of word class), as well as by the parametric mixed ANOVA with word class as within-subjects factor (because all students solved all subscales) and grade as between-subjects factor, whereby both factors had three levels (nouns, adjectives, and verbs and 5th, 6th, and 7th grade, respectively). Just as it was the case with testing the effects of gender, the results

of the parametric and nonparametric analyses were comparable, so we will report on the results of the mixed ANOVA which enabled us to test the interaction of grade and word class in addition to testing their main effects.

Since Mauchly's test of sphericity was statistically significant ($p = .045$), we will report the Greenhouse-Geisser corrected values for the main effect of word class, grade and the interaction of word class and grade. As already noted, since the number of girls and boys participating in this study was uneven, the interaction of gender, grade, and word class was not inspected.

The mixed ANOVA revealed a statistically significant main effect of word class ($F [1.921, 286.23] = 43.22, p < .001, \text{partial } \eta^2 = .225$). This result tells us that word class influences students' results on the GAVM and explains 22.5% of total variance. As can be seen from Table 4, students were most successful in solving the subscale with noun synonyms, then adjective collocations and finally verb collocations. The interaction of word class and grade was also statistically significant ($F [3.84, 286.23] = 2.71, p = .032, \text{partial } \eta^2 = .035$). Picture 1 and the means in Table 4 show that the word class effect was largest among 5th grade students, and smallest among 7th grade students, and that the grade effect was largest for verbs and smallest for nouns.

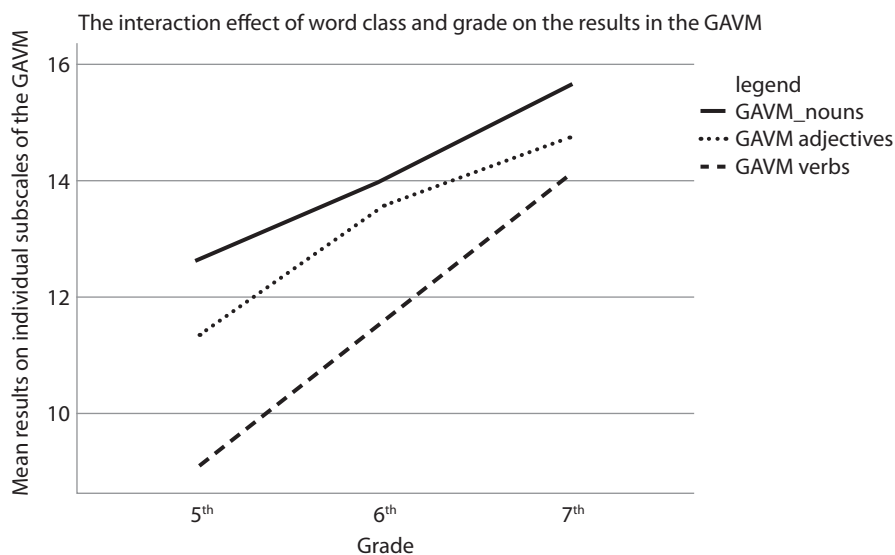


Figure 1. The interaction effect of word class and grade on the results on the GAVM

The analysis also revealed a significant main effect of grade ($F [2, 149] = 23.68, p < .001, \text{partial } \eta^2 = .241$), which indicates that older students achieved better results on the GAVM, i.e., that grade accounts for 24.1 % of total variance.

Post hoc Scheffe tests indicate that differences between grades are statistically significant on all levels (5th - 7th, $p < .001$; 5th - 6th, $p = .001$; 6th - 7th, $p = .010$).

Correlation patterns: Mill Hill Vocabulary scales and GAVM

The correlation of results on the GAVM and the results on the MHV in the total sample were tested by Spearman's non-parametric and Pearson's parametric correlation coefficients. Due to only slight differences between those two measures, only Pearson's coefficients will be presented in Table 5. There is a strong and statistically significant positive correlation between the total result on our researcher-designed GAVM, assessing general academic vocabulary knowledge, and the total result on the Mill Hill Vocabulary Scales, assessing general vocabulary knowledge. Moderate and statistically significant positive correlations are also present between the GAVM subscales and the multiple-choice task of identifying synonyms on the Mill Hill Vocabulary Scales. The correlations between the GAVM subscales and the word definition subscale on the Mill Hill Vocabulary Scales were somewhat weaker, but still of moderate size and statistically significant.

Pearson's correlations between the individual subscales of the GAVM are positive and moderate and correlate highly with the total result on the GAVM.

Table 5

Pearson's correlation coefficients for the results on the GAVM and the results on the MHV in the total sample (N = 152)

	GAVM-ADJ	GAVM-V	GAVM-TOT	MHV-D	MHV-S	MHV-TOT
GAVM-N	.45	.63	.80	.45	.59	.58
GAVM-ADJ		.66	.83	.41	.54	.53
GAVM-V			.91	.51	.69	.66
GAVM-TOT				.54	.72	.70
MHV-D					.60	.90
MHV-S						.89

Legend. GAVM-N = subscale of noun synonyms on GAVM; GAVM-ADJ = subscale of adjective collocations on GAVM; GAVM-V = subscale of verb collocations on GAVM; GAVM-TOT = total result on the GAVM scale; MHV-D = subscale of word definition on MHV; MHV-S = subscale of synonym identification on MHV; MHV-TOT = total result on MHV.

Note. All correlations are significant at the $p < .001$ level.

Internal consistency of the GAVM

Internal consistency of the GAVM is high, as indicated by the standardized Cronbach's alpha coefficient ($\alpha = .90$). For the instrument subscales, the alpha coefficients were .78 (noun synonyms), .79 (adjective collocations), and .79 (verb collocations). The correlations between items ranged from low negative to high positive. Most correlations were in the range between .10 and .30. From the total of 190 correlations between items for every subscale, 54% of the correlations for the subscale of verb collocations fit to that category, as well as 53% of correlations for the subscale of noun synonyms and the subscale of adjective collocations. A substantial number of correlations exceeded .30 (21% of the correlations on the subscale of adjective collocations and 16% of the correlations on the subscales of verb collocations and noun synonyms). The highest correlation coefficient for the subscale of adjective collocations was .70, for the subscale of noun synonyms .56, and for the subscale of verb collocations .45. One part of the

correlations was low and positive (ranging from .00 to .10): 24% for the subscale of noun synonyms, 20% for the subscale of verb collocations, and 17% for the subscale of adjective collocations. The smallest number of correlations was low and negative (ranging from -.01 to -.15, with the exception of one correlation between two items on the subscale of noun synonyms with the value of -.29): 9% of the correlations for the subscales of adjective and verb collocations, and 7% for the subscale of noun synonyms.

Low and/or negative correlations between some items are mostly related to items which, if excluded from the subscales, would contribute to their internal consistency. For example, 79% of all negative correlations, as well as 30% of all low positive correlations on the subscale of noun synonyms is related to the items “oskudica” (Eng. scarcity) and “dobrobit” (Eng. wellbeing), including the abovementioned highest negative correlation within the instrument. Visual inspection of the correlation matrix for the subscale of noun synonyms reveals that low correlations are mostly related to several items (“ustanova [Eng. institution]”, “pothvat [Eng. endeavor]”, “odredba [Eng. decree]”, “oskudica [Eng. scarcity]”, “dobrobit [Eng. wellbeing]”). This indicates that those items should either be replaced, or the suggested answers should be adjusted. Scale reliability analysis confirms this conclusion for the item “dobrobit [Eng. wellbeing]”. Excluding it from the subscale of noun synonyms would result in a higher Cronbach’s alpha coefficient for that subscale (.79 with respect to the current .78). On the subscale of adjective collocations, 65% of all negative correlations and 30% of all low positive correlations are related to two items, “postojana [Eng. stable]” and “prisebna [Eng. self-possessed]”. Visual inspection of the correlation matrix for the subscale of adjective collocations shows that the most of low correlations are connected with the items “časna [Eng. honourable]”, “neznatna [Eng. negligible]”, “ravnomjeran [Eng. equally distributed]”, “postojana [Eng. stable]”, and “prisebna [Eng. self-possessed]”. Scale reliability analysis shows that excluding items “postojana [stable]” and “prisebna [Eng. self-possessed]” from the subscale would contribute the most to a better internal consistency of that subscale – in that case, Cronbach’s alpha coefficient for the subscale of adjective collocations would be .80, and .81, respectively, with respect to the current .79. Visual inspection of the correlation matrix for the subscale of verb collocations reveals that low correlations are most numerous with items “obilovati [Eng. abound]”, “modelirati [Eng. mould]”, “ovladati [Eng. master]”, “predvidjeti [Eng. foresee/predict]”, and “prerađivati [Eng. process]”. The highest number of low correlations is again related to two items, “predvidjeti [Eng. foresee/predict]” and “prerađivati [Eng. process]” (44% of the negative and 33% of the positive low correlations). If the items “modelirati [Eng. mould]” or “predvidjeti [Eng. foresee/predict]” were to be excluded from the subscale, Cronbach’s alpha coefficient for the subscale of verb collocations would rise equally for both items - to .80 with respect to the current .79.

Low and/or negative correlations are not usually present in instruments which are designed to measure a single construct and their presence indicates that some items of the GAVM should be replaced or slightly modified. However, they also reveal that a certain level of vocabulary breadth and/or depth of some words does not necessarily

presuppose equally developed semantic network for other words, especially when it comes to a specific register as it is the case with general academic vocabulary measured by the GAVM. It should be noted that the correlation matrix of the MHV for our sample also reveals a certain amount of low positive and negative correlations between items. The percentage of low positive correlations (ranging from .00 to .10) on the first form of the MHV was 32% on the subscale of word definitions and 17% on the subscale of synonym identification. The percentage of low negative correlations (from -.001 to -.10) on those subscales was 13% and 11%, respectively. Negative correlations between particles on the first form of the MHV ranging from -.10 to -.35 take up 9% of the total number of correlations on the subscale of word definitions and 3% on the subscale of synonym identification.

On the second form of the MHV, the percentage of low positive correlations (from .00 to .10) was 20% on the subscale of word definitions and 18% on the subscale of synonym identification. The percentage of low negative correlations (ranging from -.001 to -.10) on the second MHV form was 7% on the subscale of word definitions and 19% on the subscale of synonym identification, while the percentage of negative correlations in the range between -.10 to -.30 was 3% and 5%, respectively.

Discussion

The purpose of this study was to assess the sensitivity, reliability, and validity of a preliminary version of a researcher-designed measure of general academic vocabulary knowledge which was tested on a sample of upper-elementary school students. This general academic vocabulary measure, GAVM, consisted of three subscales, each containing 20 multiple-choice items—a subscale requiring identification of synonyms to nouns, a subscale requiring recognition of the least plausible collocational pairs for adjectives, and a subscale requiring recognition of the least plausible collocational pairs for verbs. With this type of questions, one must acknowledge the possibility of a false positive or a false negative result with individual items, since students might know a stimulus item but be insufficiently familiar with response items, which leads them to choose wrong answers, or they could choose the correct answer by random guessing. A more thorough inspection of the wrong answers given by the students in the GAVM revealed that the students are not familiar with very formal dimensions of meaning, or those too distant from their everyday experiences (e.g., *independent judgment*). The most frequent distractors students selected in identifying synonyms to nouns were contextually related to the initial noun. For example, when the initial word was “quote”, the students very frequently chose “book” as a synonym, while the correct answer was “citation”. Also, frequent choices included distractors which were incorporated in the tasks on the basis of their word formation or phonological similarity to the initial item. One such example is choosing the word “rukohvat” (Eng. handrail) instead of “akcija” (Eng. action) as a synonym for the word “pothvat” (Eng. undertaking).

The results of the independent samples t-tests, which were used to test the effects of gender on the results on the GAVM, suggest that girls achieve better results on all

subscales of the GAVM, i.e., that girls are better in recognizing synonyms to nouns belonging to general academic vocabulary, as well as in identifying implausible collocations to adjectives and verbs from that register. As already mentioned, numerous studies show that girls outperform boys in various vocabulary measures from an early age to adolescence (e.g., Marjanović-Umek & Fekonja-Pekljaj, 2017, p. 106; Simonsen et al., 2014, p. 18). In the Croatian context, Didović and Kolić-Vehovec (2009) report that, among the 8th graders participating in their study, girls had richer vocabularies than boys. Also, in the group of somewhat younger elementary school students (4th - 7th grade) who participated in a study conducted by Keresteš et al., (2020), girls also had richer vocabularies than their male peers. However, the results of some other studies indicate that the role of gender in vocabulary development changes as children grow older. In a study by Toivainen et al., (2017, p. 84) boys showed better results on vocabulary measures at the age of 10 and 12, and no statistically significant gender differences could be detected at the ages of 14 and 16. Similarly, Coxhead et al. (2015, p. 131) report finding no gender related differences in vocabulary breadth among 13- to 18-year-old boys and girls, native speakers of English. Since in our study it was not justified to test the interaction effects of gender and grade due to the uneven number of participating girls and boys, those effects remain to be tested in some future studies. However, the vocabulary advantage for girls found in our study indicates that a slightly gender-diversified approach should be applied to vocabulary learning, i.e., that boys might benefit more strongly from more extensive vocabulary practice opportunities during elementary education.

A significant main effect of grade revealed by mixed analysis of variance confirms that our instrument, GAVM, is sensitive to age-, i.e., grade-related vocabulary development. This is a basic requirement for any measure used for the purpose of measuring vocabulary, since older students know more words, have more knowledge about the world and have accumulated experience in using expository texts in the educational context. The fact that older students achieve higher results than the younger ones on every subscale of the GAVM confirms the validity of our instrument and shows that it can detect the well documented age differences which are a result of schooling and maturation (e.g., Biemiller & Slonim, 2001, p. 505; Keresteš et al., 2020, p. 55; Song et al., 2015, p. 10; Verhoeven et al., 2011, p. 15).

The results of the mixed ANOVA also show that word class influences students' results on the GAVM. The students were most successful in solving the subscale with nouns, then adjectives, and finally verbs. This indicates that, in vocabulary interventions, different word classes should be approached in a selective and differentiated way: based on the results of this study, we can assume that vocabulary interventions in which the approach would be diversified by word class would be more efficient than word class-nonselective interventions. Within such an approach, it remains important to specifically focus on vocabulary depth when it comes to verbs for two reasons. The first arises from the specifically rich derivational morphology of the Croatian

language which enables changing the meaning of a verb by adding prefixes, whereby the verb's valency and linking patterns are often changed as well. The second reason arises from psycholinguistic factors which influence the later formation of a holistic conceptualization of verb meanings in comparison to nouns and adjectives.

Even within a word class (e.g., verbs), there might be a need to approach the development of vocabulary depth differently with different age groups, since 3.5% of the variance in the results is explained by the interaction of word class and the grade the students were in, i.e., students' age. Nevertheless, it must be noted that the differences in students' results between word classes could also be due to the differences in task type, and not only word class. Even though it may seem that the differences in task type through which different word classes were approached in the preliminary version of GAVM (synonyms and limitations of collocational use) make an all-levels comparison among word classes impossible, it is important to keep in mind that the instrument was knowingly constructed in such a way because our guiding principles were precisely the characteristics of individual word classes and the typical context in which they occur. As already noted, verbs connect with other words based on their relational nature and valency characteristics, adjectives are used for describing nouns, and nouns can stand alone in the wholeness of their meaning.

Some authors, such as Vermeer (2001, p. 231), say that the two vocabulary dimensions, breadth and depth, are not two conceptually distinct entities, and therefore high correlations are to be expected between the measures of vocabulary breadth and measures of vocabulary depth. The correlations between individual subscales of the GAVM are moderate, which indicates the need to refine the composition of individual items included in each subscale. However, the strong positive correlation between our researcher-designed GAVM measuring a specific language register and the Mill Hill Vocabulary Scales measuring general vocabulary knowledge is in line with similar results provided by previous research comparing vocabulary knowledge in those two registers (e.g. Masrai & Milton, 2018, p. 50). At the same time, this strong positive correlation indicates good convergent validity of the new measure of general academic vocabulary knowledge, GAVM.

As noted earlier, the correlations between the GAVM subscales and the subscale of identifying synonyms on the Mill Hill Vocabulary Scales (both multiple-choice) are higher than the correlations between the GAVM subscales and the subscale with word definitions on the Mill Hill Vocabulary Scales. Since higher correlations between the two instruments are found between the subscales of the same task type, we can conclude that, when several instruments are used comparatively in the same study to assess a similar construct, it is important to consider which task types are used to form a specific instrument.

Finally, GAVM also demonstrated adequate internal consistency as a whole and at the level of individual subscales, as indexed by high Cronbach's alpha coefficients. This shows that this form of our instrument, i.e., the chosen tasks, can be applied for

the purpose of assessing the knowledge of general academic vocabulary breadth and depth. With additional testing of metric characteristics, new items (words) adjusted to other levels of the educational system could be included in the existing task types, in consistence with specific demands or goals of individual studies and interventions.

Conclusions

Since more and more studies report a high correlation between vocabulary breadth and depth and reading comprehension (e.g., Didović & Kolić-Vehovec, 2009, p. 109; Kieffer & Lesaux, 2007, p. 138; Keresteš et al., 2020, p. 61; Peti-Stantić 2019, p. 96), there is an increased awareness of the importance of a systematic analysis of vocabulary in textbooks and a comparison of that material with students' lexical resources at a given age. A mismatch between their lexical knowledge and the linguistic material used in their textbooks will result in poorer text comprehension. Therefore, more attention should be paid to the linguistic material used to present textbook content – it should at the same time be sufficiently understandable for the students and be an incentive for the development of their lexical abilities.

Sensitive, reliable, and valid vocabulary measures are a prerequisite for researching vocabulary development and for a reliable assessment of the effects of different vocabulary interventions. Since the instrument tested in this study, the general academic vocabulary measure for upper-elementary school students, is sensitive to age and gender differences, strongly positively correlated with a standardized general measure of vocabulary knowledge and has good internal consistency (reliability), the most important conclusion of this analysis is that GAVM can be used in basic (fundamental) scientific studies which explore (the factors influencing) vocabulary development and in evaluating vocabulary interventions aimed at developing general academic vocabulary knowledge of upper-elementary school students in Croatia. Except for the explicit focus of our new instrument on the academic register, which is currently not covered by other available vocabulary assessment instruments in Croatian language, an additional valuable characteristic of GAVM is its focus on three different word classes, supported by the significant main effect of word class in the results. Finally, compared to other vocabulary measures, GAVM is firmly grounded in theory because of the attentive approach to the choice of words, as well as on account of the careful choice of distractors in multiple-choice tasks which contribute to task quality. However, the version of the instrument presented in this paper is only preliminary and should be verified on new, larger and more representative samples, as well as additionally subsequently refined based on the results of future analyses.

While keeping in mind the recommendations of the European Parliament and Council of Europe on developing communicative competence in one's mother tongue as one of the key competences of lifelong learning (2006), analyses such as this one should also serve as guidelines for educational policies. Many educational institutions and policy-makers actively encourage projects which focus on developing students'

communicative and reading competences, especially academic literacy (e.g., FörMig, BISS [Germany], IAL [USA], SALAD [West Sussex]). One such example is a scientific project with a holistic approach to academic literacy led by Paola Uccelli and associates (e.g. Uccelli et al., 2015). Her research group focuses on defining the set of language skills which go beyond academic vocabulary and are salient and relevant in the educational context. They have also developed an instrument for the assessment of academic language skills of upper-elementary school students (CALSI) which assesses several dimensions of text-level abilities. Those skills include understanding argument structure, inferencing from local and global context, understanding of syntactical structures, semantics, and morphology. Along with specific studies complementing the one we have conducted, such comprehensive instruments could also serve as models for further development of measures of academic vocabulary and academic language abilities in Croatian language.

Data availability statement:

The data that support the findings of this study are openly available in figshare at <https://figshare.com/s/f6b21e70f4b59dbb8733>

References

- Anderson, R. C., & Freebody, P. (1979). *Vocabulary knowledge. Technical report*. No. 136. <https://eric.ed.gov/?id=ED177480>
- Biemiller, A., & Slonim, N. (2001). Estimating root word vocabulary growth in normative and advantaged populations: Evidence for a common sequence of vocabulary acquisition. *Journal of Educational Psychology*, 93(3), 498–520. <https://doi.org/10.1037/0022-0663.93.3.498>
- Bornstein, M. H., Hahn, C.-S., & Haynes, O. M. (2004). Specific and general language performance across early childhood: Stability and gender considerations. *First Language*, 24(3), 267–304. <https://doi.org/10.1177/0142723704045681>
- Clark, E. V. (1995). Later lexical development and word formation. In P. Fletcher & B. MacWhinney (Eds.), *The Handbook of Child Language* (pp. 393–412). Blackwell Publishing Ltd. <https://doi.org/10.1111/b.9780631203124.1996.00016.x>
- Cordier, F., Croizet, J.-C., & Rigalleau, F. (2013). Comparing nouns and verbs in a lexical task. *Journal of Psycholinguistic Research*, 42(1), 21–35. <https://doi.org/10.1007/s10936-012-9202-x>
- Corson, D. (1997). The learning and use of academic English words. *Language Learning*, 47(4), 671–718. <https://doi.org/10.1111/0023-8333.00025>
- Coxhead, A., Nation, P., & Sim, D. (2015). Measuring the vocabulary size of native speakers of English in New Zealand secondary schools. *New Zealand Journal of Educational Studies*, 50(1), 121–135. <https://doi.org/10.1007/s40841-015-0002-3>

- Cronbach, L. J. (1942). An analysis of techniques for diagnostic vocabulary testing. *The Journal of Educational Research*, 36(3), 206–217. <https://doi.org/10.1080/00220671.1942.10881160>
- Didović, M., & Kolić-Vehovec, S. (2009). Različiti aspekti poznavanja rječnika i razumijevanje teksta kod učenika osnovne škole [Different aspects of vocabulary knowledge and text comprehension in elementary school students]. *Psihologijske teme* 18(1), 99–117. Retrieved from https://hrcak.srce.hr/index.php?show=clanak&id_clanak_jezik=62574
- Dunn, L. M., Dunn, L. M., Kovačević, M., Padovan, N., Hržica, G., Kuvač Kraljević, J., Mustapić, M., Dobravac, G., & Palmović, M. (2009). *Priručnik za PPVT-III-HR*. [PPVT-III-HR Manual.] Naklada Slap.
- Eriksson, M., Marschik, P. B., Tulviste, T., Almgren, M., Pérez Pereira, M., Wehberg, S., Marjanović-Umek, L., Gayraud, F., Kovacevic, M., & Gallego, C. (2012). Differences between girls and boys in emerging language skills: Evidence from 10 language communities: Differences between girls and boys in early language. *British Journal of Developmental Psychology*, 30(2), 326–343. <https://doi.org/10.1111/j.2044-835X.2011.02042.x>
- European Commission. (2006). Recommendation of the European parliament and of the council of 18 December 2006 on key competences for lifelong learning. *Official Journal of European Union* (L 394), 10–18. Retrieved from <https://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2006:394:0010:0018:en:PDF>
- Field, A. (2009). *Discovering statistics using SPSS for Windows (3rd edition)*. SAGE.
- Feilke, H. (2012). Bildungssprachliche Kompetenzen—fördern und entwickeln [Fostering and developing academic language skills]. *Praxis Deutsch*. 39 (233), 4–13.
- Hawkins, R. O., Hale, A., Sheeley, W., & Ling, S. (2011). Repeated reading and vocabulary-previewing interventions to improve fluency and comprehension for struggling high-school readers. *Psychology in the Schools*, 48(1), 59–77. <https://doi.org/10.1002/pits.20545>
- Hiebert, E., & Lubliner, S. (2010). The nature, learning, and instruction of general academic vocabulary. In A. E. Farstrup & S. J. Samuels (Eds.), *What Research Has to Say about Vocabulary Instruction* (pp. 106–129) International Reading Association.
- IBM Corp. (2019). IBM SPSS Statistics for Windows, Version 26.0. IBM Corp.
- Indrisano, R., & Chall, J. S. (1995). Literacy development. *Journal of Education*, 177(1), 63–83.
- Kieffer, M. J., & Lesaux, N. K. (2007). Breaking down words to build meaning: Morphology, vocabulary, and reading comprehension in the urban classroom. *The Reading Teacher*, 61(2), 134–144. <https://doi.org/10.1598/RT.61.2.3>
- Killian, J. (2016). Wortschatzerweiterung und Wortschatzvertiefung [Expanding vocabulary]. In I. Pohl & W. Ulrich (Eds.), *Deutschunterricht in Theorie und Praxis (DTP): Wortschatzarbeit* (2. Auflage, pp. 133–142). Schneider Verlag Hohengehren.
- Križan, L., & Matešić, K. (2001). *Priručnik za Ravenove progresivne matrice i ljestvice rječnika*. Mill Hill ljestvica rječnika [Manual for Raven's progressive matrices and vocabulary scales]. Naklada Slap.
- Kaushanskaya, M., Gross, M., & Buac, M. (2013). Gender differences in child word learning. *Learning and Individual Differences*, 27, 82–89. <https://doi.org/10.1016/j.lindif.2013.07.002>
- Keresteš, G., Lugomer Armano, G., Čičko, V., Goldstein, T., Jerković, K., Juras, L., Lindić, B., Matuško, M., Pera, K., Pešorda, D., Polak, P., Šoštarić, M., Šutić, L., & Vrdoljak, A. (2020).

- Čitanje u ranoj adolescenciji - dobne i spolne razlike te individualni i obiteljski prediktori [Reading in early adolescence – age and gender differences, and personal and familial predictors]. FF Press.
- Lesaux, N. K., Kieffer, M. J., Faller, S. E., & Kelley, J. G. (2010). The effectiveness and ease of implementation of an academic vocabulary intervention for linguistically diverse students in urban middle schools. *Reading Research Quarterly*, 45(2), 196–228. <https://doi.org/10.1598/RRQ.45.2.3>
- Li, M., & Kirby, J. R. (2014). The effects of vocabulary breadth and depth on English reading. *Applied Linguistics*. <https://doi.org/10.1093/applin/amu007>
- Logan, J. K., & Kieffer, M. J. (2017). Evaluating the role of polysemous word knowledge in reading comprehension among bilingual adolescents. *Reading and Writing*, 30(8), 1687–1704. <https://doi.org/10.1007/s11145-017-9745-1>
- Manuel, H. T. (1966). *Development of Inter-American Test materials*. <https://eric.ed.gov/?q=inter-american+series&id=ED010670>
- Marjanovič-Umek, Lj., & Fekonja-Peklaj, U. (2017). Gender differences in children's language: a meta-analysis of Slovenian studies. *Center for Educational Policy Studies Journal* 7(2), 97–111. <https://files.eric.ed.gov/fulltext/EJ1145869.pdf>
- Masrai, A., & Milton, J. (2018). Measuring the contribution of academic and general vocabulary knowledge to learners' academic achievement. *Journal of English for Academic Purposes*, 31, 44–57. <https://doi.org/10.1016/j.jeap.2017.12.006>
- McGeown, S. P., Duncan, L. G., Griffiths, Y. M., & Stothard, S. E. (2015). Exploring the relationship between adolescent's reading skills, reading motivation and reading habits. *Reading and Writing*, 28(4), 545–569. <https://doi.org/10.1007/s11145-014-9537-9>
- Nation, I. S. P. (2001). *Learning vocabulary in another language* (1st ed.). Cambridge University Press. <https://doi.org/10.1017/CBO9781139524759>
- Nation, P., & Beglar, D. (2007). A vocabulary size test. *The Language Teacher* 31(7), 9–13.
- National Reading Panel (U.S.). (2000). *Report of the National Reading Panel: teaching children to read: an evidence-based assessment of the scientific research literature on reading and its implications for reading instruction: reports of the subgroups*. <https://www.nichd.nih.gov/sites/default/files/publications/pubs/nrp/Documents/report.pdf>
- OECD. (2016). *Pisa 2015 results (Volume I): Excellence and equity in education*. <https://doi.org/10.1787/9789264266490-en>
- OECD. (2019). *Pisa 2018 results (Volume I): What students know and can do*. <https://doi.org/10.1787/5f07c754-en>
- Pearson, P. D., Hiebert, E. H., & Kamil, M. L. (2007). Vocabulary assessment: What we know and what we need to learn. *Reading Research Quarterly*, 42(2), 282–296. <https://doi.org/10.1598/RRQ.42.2.4>
- Pečjak, S., Kolić-Vehovec, S., Rončević Zubković, B., & Ajdišek, N. (2009). (Meta)kognitivni i motivacijski prediktori razumijevanja teksta adolescenata u Hrvatskoj i Sloveniji [(Meta) cognitive and motivational predictors of text comprehension of adolescents in Croatia and Slovenia]. *Suvremena psihologija* 12(2), 257–269. <https://hrcak.srce.hr/82952>
- Peti-Stantić, A., Anđel, M., Keresteš, G., Ljubešić, N., Stanojević, M.-M., & Tonković, M. (2018). Psiholingvističke mjere ispitivanja 3.000 riječi hrvatskoga jezika: Konkretnost i predočivost

- [Psycholinguistic estimates of 3000 words of Croatian: concreteness and imageability]. *Suvremena lingvistika*, 44(85), 91–112. <https://doi.org/10.22210/suvlin.2018.085.05>
- Peti-Stantić, A. (2019). *Čitanjem do (spo)razumijevanja. Od čitalačke pismenosti do čitateljske sposobnosti [The road to comprehension—developing reading literacy]*. Ljevak.
- Petz, B., Kolesarić, V., & Ivanec, D. (2012). *Petzova statistika - Osnovne statističke metode za nematematičare [Petz's statistics - basic statistical methods for nonmathematicians]*. Naklada Slap.
- Pignot-Shahov, V. (2012). Measuring L2 receptive and productive vocabulary knowledge. *Language Studies Working Papers*, 4 (1), 37-45. Retrieved from <https://api.semanticscholar.org/CorpusID:17785514>
- Proctor, C. P., Dalton, B., & Grisham, D. L. (2007). Scaffolding English language learners and struggling readers in a universal literacy environment with embedded strategy instruction and vocabulary support. *Journal of Literacy Research*, 39(1), 71–93. <https://doi.org/10.1080/10862960709336758>
- Ouellette, G. P. (2006). What's meaning got to do with it: The role of vocabulary in word reading and reading comprehension. *Journal of Educational Psychology*, 98(3), 554–566. <https://doi.org/10.1037/0022-0663.98.3.554>
- Raven, J. (2000). The Raven's progressive matrices: Change and stability over culture and time. *Cognitive Psychology*, 41(1), 1–48. <https://doi.org/10.1006/cogp.1999.0735>
- Read, J. (1993). The development of a new measure of L2 vocabulary knowledge. *Language Testing*, 10(3), 355–371. <https://doi.org/10.1177/026553229301000308>
- Read, J. (2000). *Assessing vocabulary*. Cambridge Core. <https://doi.org/10.1017/CBO9780511732942>
- Richards, J. C. (1976). The role of vocabulary teaching. *TESOL Quarterly*, 10(1), 77. <https://doi.org/10.2307/3585941>
- Schmitt, N. (2014). Size and depth of vocabulary knowledge: What the research shows. *Language Learning*, 64(4), 913–951. <https://doi.org/10.1111/lang.12077>
- Simonsen, H. G., Kristoffersen, K. E., Bleses, D., Wehberg, S., & Jørgensen, R. N. (2014). The Norwegian Communicative Development Inventories: Reliability, main developmental trends and gender differences. *First Language*, 34(1), 3–23. <https://doi.org/10.1177/0142723713510997>
- Swanborn, M. S. L., & de Glopper, K. (1999). Incidental word learning while reading: A meta-analysis. *Review of Educational Research*, 69(3), 261–285. <https://doi.org/10.3102/00346543069003261>
- Song, S., Su, M., Kang, C., Liu, H., Zhang, Y., McBride-Chang, C., Tardif, T., Li, H., Liang, W., Zhang, Z., & Shu, H. (2015). Tracing children's vocabulary development from preschool through the school-age years: An 8-year longitudinal study. *Developmental Science*, 18(1), 119–131. <https://doi.org/10.1111/desc.12190>
- Toivainen, T., Papageorgiou, K. A., Tosto, M. G., & Kovas, Y. (2017). Sex differences in non-verbal and verbal abilities in childhood and adolescence. *Intelligence*, 64, 81–88. <https://doi.org/10.1016/j.intell.2017.07.007>
- Townsend, D., Filippini, A., Collins, P., & Biancarosa, G. (2012). Evidence for the importance of academic word knowledge for the academic achievement of diverse middle school students. *The Elementary School Journal*, 112(3), 497–518. <https://doi.org/10.1086/663301>
- Uccelli, P., Barr, C. D., Dobbs, C. L., Galloway, E. P., Meneses, A., & Sánchez, E. (2015). Core

- academic language skills: An expanded operational construct and a novel instrument to chart school-relevant language proficiency in preadolescent and adolescent learners. *Applied Psycholinguistics*, 36(5), 1077–1109. <https://doi.org/10.1017/S014271641400006X>
- Uusen, A., & Määrsepp, M. (2012). Gender differences in reading habits among boys and girls of basic school in Estonia. *Procedia - Social and Behavioral Sciences*, 69, 1795–1804. <https://doi.org/10.1016/j.sbspro.2012.12.129>
- Verhoeven, L., van Leeuwe, J., & Vermeer, A. (2011). Vocabulary growth and reading development across the elementary school years. *Scientific Studies of Reading*, 15(1), 8–25. <https://doi.org/10.1080/10888438.2011.536125>
- Vermeer, A. (2001). Breadth and depth of vocabulary in relation to L1/L2 acquisition and frequency of input. *Applied Psycholinguistics*, 22(2), 217–234. <https://doi.org/10.1017/S0142716401002041>
- Wiemer-Hastings, K. K., & Xu, X. (2005). Content differences for abstract and concrete concepts. *Cognitive Science*, 29(5), 719–736. https://doi.org/10.1207/s15516709cog0000_33
- Williams, J. P. (2018). Text structure instruction: The research is moving forward. *Reading and Writing*, 31(9), 1923–1935. <https://doi.org/10.1007/s11145-018-9909-7>

Vedrana Gnjiđić

Faculty of Humanities and Social Sciences
University of Zagreb
Ivana Lučića 3, 10000 Zagreb, Croatia
vgnjidic@ffzg.hr

Gordana Keresteš

Department of Psychology
Faculty of Humanities and Social Sciences
University of Zagreb
Ivana Lučića 3, 10000 Zagreb, Croatia
gordana.kerestes@ffzg.hr

Anita Peti-Stantić

Department of South Slavic languages and literatures
Faculty of Humanities and Social Sciences
University of Zagreb
Ivana Lučića 3, 10000 Zagreb, Croatia
anita.peti-stantic@ffzg.hr

Mjerenje širine i dubine akademskoga vokabulara učenika viših razreda osnovne škole

Sažetak

U ovom se radu analizira novorazvijena mjera procjene poznavanja općega akademskog vokabulara učenika viših razreda osnovne škole u Hrvatskoj. Njome se procjenjuje poznavanje sinonima (kao mjera širine vokabulara) i ograničenja kolokacijske upotrebe (kao mjera dubine vokabulara) unutar općega akademskog registra. U istraživanju su sudjelovali učenici petog, šestog i sedmog razreda osnovne škole (N = 152, 69,7 % djevojčica). Osim mjere općega akademskog vokabulara, učenici su ispunili i Mill Hill ljestvice rječnika, standardiziranu mjeru općega vokabulara. Rezultati pokazuju da je novi instrument za mjerenje poznavanja općega akademskog vokabulara pouzdan i valjan—ima dobru unutarnju konzistentnost, osjetljiv je na očekivane dobne i spolne razlike u dubini i širini vokabulara te visoko korelira s rezultatima dobivenim na Mill Hill ljestvicama rječnika. Ova bi nova mjera mogla popuniti prazninu u istraživanju akademskoga vokabulara jer nekoliko trenutačno dostupnih mjera procjene vokabulara na hrvatskom jeziku ne testira specifično taj segment, te bi se mogla koristiti za evaluaciju različitih intervencija usmjerenih na razvoj akademskoga vokabulara u školama.

Ključne riječi: kolokacije; opći akademski vokabular; sinonimi

Uvod

Djeca se tijekom osnovnoškolskoga obrazovanja susreću s brojnim novim riječima. Prema nekim procjenama za engleski jezik, broj riječi s kojima se djeca susreću u svakoj godini školovanja od desete godine života nadalje može doseći i do deset tisuća novih riječi (Clark, 1995, str. 393).

Tekstovi su osnovni oblik poučavanja i prijenosa znanja u obrazovnom kontekstu, bilo u obliku učeničkih bilježaka, bilježaka na ploči ili udžbeničkih tekstova. Od učenika viših razreda osnovne škole očekuje se da samostalnije i češće koriste udžbeničke tekstove nego što je to slučaj s učenicima u nižim razredima. Osim toga, učenici viših razreda osnovne škole najčešće uče iz udžbeničkih tekstova objasnidbenoga tipa koji

su nerijetko vrlo dugački i relativno složeni (Williams, 2018, str. 1924). Polazeći od pretpostavke da bi udžbenički tekstovi učenicima trebali služiti kao temelj razvoja znanja o svijetu, nema dileme o tome kako učenici mogu razumjeti cjelinu teksta koji čitaju i iz kojega uče samo ako razumiju značenja velike većine riječi (Peti-Stantić, 2019, str. 169, 198). Prema rezultatima PISA testiranja iz 2015. i 2018. godine, razina čitalačke pismenosti hrvatskih 15-godišnjaka statistički je značajno ispod prosjeka rezultata zemalja OECD-a (pri čemu su djevojčice uspješnije od dječaka). U ciklusu iz 2015. godine hrvatski su učenici bili na 31. mjestu među zemljama OECD-a, a 2018. zauzeli su 29. mjesto (OECD 2016, str. 149; OECD 2019, str. 97). Takvi rezultati pokazuju da bi, do uvođenja sustavne promjene načina poučavanja, hrvatski učenici mogli imati koristi od različitih vrsta intervencija usmjerenih na razvoj čitalačke pismenosti. Intervencije usmjerene na razvoj vokabulara prepoznate su kao jedna od ključnih sastavnica intervencija usmjerenih na razvoj čitalačke pismenosti (Lesaux, Kieffer, Faller i Kelley, 2010, str. 198; Hawkins, Hale, Sheeley i Ling, 2011, str. 61; Proctor, Dalton i Grisham, 2007, str. 75).

Većina udžbeničkih sadržaja napisana je formalnim jezičnim registrom akademskoga tipa za koji se očekuje da ga većina učenika svlada tijekom školovanja. Taj je registar, uz specifičnost vokabulara, obilježen sintaktički i semantički zahtjevnim strukturama (Feilke, 2012, str. 8). Hiebert i Lubliner (2008, str. 110) vokabular koji se koristi u školskim tekstovima i zadacima dijele na četiri vrste: vokabular specifičan za pojedino predmetno područje, opći akademski vokabular, vokabular koji se koristi u školskim zadacima i književni vokabular. Riječi koje pripadaju vokabularu specifičnom za pojedino predmetno područje su one riječi koje se, primjerice, koriste u nastavi kemije, matematike ili jezika (npr., *redoks*, *višekratnik* ili *glagolski*), dok je vokabular koji se koristi u školskim zadacima karakterističan za upute u testovima i zadacima za vježbu (*podcrtati*, *kategorizirati*). Književni je vokabular tipičan za književne tekstove i uključuje riječi poput *zgranut* ili *bremenit*, dok riječi poput *djelomičan*, *sugerirati* i *svrha* koje se koristi u mnogim nastavnim predmetima, ali i u sporazumijevanju o apstraktnim uvjetima života, pripadaju općem akademskom vokabularu.

Međupredmetna prisutnost riječi koje pripadaju općem akademskom vokabularu ukazuje na važnost njihova svladavanja — ako učenici ne razumiju značenja tih riječi i potencijal njihove upotrebe, ne mogu razumjeti ni tekstove u svojim udžbenicima iz kojih bi trebali učiti (Peti-Stantić, 2019, str. 169, 198). Riječi koje pripadaju akademskom vokabularu u literaturi se karakteriziraju kao apstraktne, nisko predočive i niske relativne čestote (Corson, 1997, str. 696). Budući da se procesu razvoja vokabulara u višim razredima osnovne škole pristupa kao da se on odvija automatizirano, učenike se u hrvatskim školama akademskom vokabularu ne poučava eksplicitno, što proturječi saznanjima o tome da se proces usvajanja riječi ne odvija uvijek automatizirano. Istraživanja, naime, pokazuju da učenici u dobi od četvrtog razreda osnovne škole do trećeg razreda srednje škole u prosjeku spontano usvoje tek 15 % nepoznatih riječi koje susretnu samostalno čitajući tekstove (Swanborn i de Gloppe, 1999, str. 272).

Zato smo odlučile analizirati poznavanje općega akademskog vokabulara učenika viših razreda osnovne škole i razviti instrument koji bi se mogao koristiti u evaluaciji budućih intervencija u svrhu unaprjeđivanja razumijevanja i sposobnosti korištenja općega akademskog vokabulara. Jedan od zaključaka Nacionalnog odbora za čitanje SAD-a jest da mjere vokabulara skrojene prema težištima specifičnih intervencija bolje detektiraju učinke rada na vokabularu nego standardizirane mjere vokabulara (National Reading Panel, 2000, str. 4-24). U skladu s tim zaključkom nastojale smo razviti instrument koji će pouzdano i osjetljivo mjeriti opći akademski vokabular, odnosno širinu i dubinu poznavanja akademskoga vokabulara s ciljem mjerenjenja učinkovitosti intervencija usmjerenih na opći akademski vokabular. U ovom se radu služimo pojmovima širine i dubine vokabulara, premda u literaturi nema potpunoga slaganja kad je riječ o njihovu određenju. Anderson i Freebody (1979, str. 22) određuju koncept širine vokabulara kao broj riječi koje poznajemo (to znanje uključuje samo osnovne dimenzije značenja), dok se dubina vokabulara odnosi na kvalitetu znanja o riječima. Prihvaćajući te definicije, mi ih preciziramo u skladu s Killianovim opisom (2016, str. 133) tako da širinom vokabulara smatramo poznavanje prototipnoga konteksta i struktura riječi koje se tvore od istoga korijena, a dubinom vokabulara poznavanje denotativnih i konotativnih značenja, asocijativnih semantičkih potencijala te metonimijskih i metaforičkih prijenosa.

Naslanjajući se na Readov (2000, str. 185) pristup dubini poznavanja vokabulara njezinim razdvajanjem na komponente, odnosno dimenzije, Schmitt (2014, str. 915) dubinu vokabulara opisuje kao poznavanje više pojedinačnih leksičkih dimenzija riječi, navodeći kao primjer poznavanje višestrukih značenja, i kao holističko poznavanje značenja riječi u koje je, primjerice, uključeno postojanje bogatih asocijativnih mreža među riječima, dok širinu vokabulara operacionalizira kao zbrajanje svih leksičkih jedinica koje su nam poznate, odnosno kojima smo do neke mjere ovladali (Schmitt, 2014, str. 942). Već rečeno ukazuje na rasprave o detaljima razumijevanja širine i dubine vokabulara među znanstvenicima, kao i na to da je, kao i u drugim psiholingvističkim mjerama kao što su npr. apstraktnost i konkretnost riječi u odnosu prema predočivosti, i ovdje riječ o dimenzijama koje se do neke mjere preklapaju. Schmitt (2014, str. 942) tako kaže da širina do neke mjere uvijek zahvaća i dubinu značenja riječi te da se dubina poznavanja vokabulara može definirati jedino kao međusobna povezanost i isprepletenost različitih aspekata značenja riječi.

Te nejasne granice odražavaju se i u korištenju i interpretaciji različitih mjera poznavanja vokabulara koje se najčešće dijele na osnovi toga procjenjuju li receptivno ili produktivno poznavanje vokabulara (Pearson, Hiebert i Kamil, 2007, str. 284). Mjere produktivnoga poznavanja vokabulara procjenjuju se kroz vještine govorenja ili pisanja, odnosno kroz zadatke koji uključuju spontanu aktivnu upotrebu riječi (Pignot-Shahov, 2012, str. 39). Mjere receptivnoga poznavanja vokabulara obično se oslanjaju na zadatke definiranja riječi (npr., Mill Hill ljestvice rječnika, Raven, 2000), odabir točne definicije riječi (npr., The Vocabulary Size Test, Nation i Beglar, 2007)

ili povezivanje riječi s njezinom definicijom (npr., Inter-American Series, Manuel, 1966). Osim definicija riječi, Cronbach (1942, str. 207) i Richards (1976, str. 79, 81, 82) predlažu nekoliko sastavnica znanja o riječima koje treba obuhvatiti prilikom testiranja vokabulara u dijagnostičke svrhe. Te sastavnice uključuju prepoznavanje višestrukih značenja riječi i ograničenja njihove upotrebe. Za Cronbacha je znanje o nekoj riječi potpuno onda kad riječ možemo razumjeti i pravilno upotrijebiti, dok Richards naglašava važnost razvijanja bogatih asocijativnih veza među riječima u našim mentalnim leksikonima. Te asocijativne veze uključuju odnose sinonimije, antonimije, hipero-, kohipo- ili hiponimije te kolokacije. Nation (2001, pp. 27) također sažima komponente znanja o riječima i dijeli ih na poznavanje triju dimenzija riječi: oblika (govorenoga, pisanoga, morfološkoga), značenja (povezanost oblika i značenja, koncepata i referenata, asocijacije) i upotrebe (semantička, sintaktička i pragmatička ograničenja upotrebe).

Primjena određenih vrsta zadataka u istraživanjima vokabulara razlikuje se ovisno o načinu konceptualiziranja znanja o riječima, odnosno o definicijama dubine i širine vokabulara koje autori prihvaćaju. Primjerice, Quелlette (2006, str. 558) širinu vokabulara definira kao ukupan broj riječi koje postoje u našem rječniku i operacionalizira je kroz zadatak imenovanja slika (koristeći dvije podskale testa The Test of Word Knowledge), kao što je to najčešće i slučaj (primjerice, test The Peabody Picture Vocabulary Test korišten u Logan i Kieffer, 2017). Koliko god takav tip zadatka bio primjeren za mnoge pojmove i riječi, njime nije moguće ispitati akademski vokabular koji se uglavnom sastoji od apstraktnih riječi, zbog čega imenovanje slika nije primjeren za njegovo istraživanje.

Quелlette (2006, str. 558) je za procjenu dubine poznavanja vokabulara upotrijebio zadatak prepoznavanja sinonima i zadatak definiranja riječi jer smatra da ti zadatci omogućavaju mjerenje razvijenosti kvalitete poznavanja vokabulara. Zadatak definiranja riječi kao mjeru dubine poznavanja vokabulara koriste i Li i Kirby (2014), a Read (1993) kao dio procjene dubine vokabulara u testu Word Associates Test uz tipove zadataka koji uključuju biranje valjanih kolokacija (sintagmatski odnosi) i prepoznavanje semantičkih asocijacija (analitički odnosi) koristi i prepoznavanje sinonima (paradigmatske odnose).

Iako Quелlette i Read poznavanje sinonima svrstavaju u mjere dubine poznavanja vokabulara, zadatak prepoznavanja sinonima u nekim se recentnim istraživanjima koristi kao mjera širine vokabulara, odnosno mjera osnovnoga razumijevanja značenja riječi (npr., The Vocabulary Subtest, dio testa Gates-MacGinitie Reading Test u Townsend, Filippini, Collins i Biancarosa, 2012; Li i Kirby, 2014).

Schmitt (2014, str. 941) zaključuje da je operacionalizacija konstrukta dubine vokabulara, neovisno o zadatcima koji se koriste za njezino mjerenje, uvijek nepotpuna jer će uvijek moći mjeriti samo ograničeni broj segmenata koji čine cjelovitost, odnosno kvalitetu poznavanja neke riječi, a nikada ukupnost te kvalitete. U ovom smo istraživanju stoga odabrale one zadatke za koje smo smatrale da će najbolje obuhvatiti

bitne dimenzije širine i dubine značenja. Prihvaćamo tezu da svi zadatci kojima se mjeri širina vokabulara djelomično obuhvaćaju i neku razinu dubine poznavanja vokabulara, ali smatramo da pojedinačni zadatci ipak većim dijelom zahvaćaju jednu ili drugu dimenziju vokabulara, odnosno širinu ili dubinu. Tako držimo da zadatak prepoznavanja sinonima pretežito mjeri širinu poznavanja vokabulara jer se u njemu riječi uspoređuju na razini cjeline pojedinačnih značenja, a zadatak prepoznavanja kolokacija, u kojem se riječi promatraju iz perspektive ograničenja međusobnih interakcija, kako semantičkih, tako i sintaktičkih, pretežito mjeri dubinu poznavanja vokabulara.

Dimenzija koja se često zanemaruje kod ispitivanja vokabulara jest važnost razlikovanja poznavanja značenja različitih vrsta riječi koju potvrđuju nalazi psiholingvističkih istraživanja. Poznato je, naime, da se u zadatcima leksičke odluke imenice prepoznaju brže od glagola (npr., Cordier, Croizet i Rigalleau, 2013, str. 28). Zbog toga smo u istraživanju koje smo provele upravo tu dimenziju izdvojile kao relevantnu. Isto tako, imenice se u općem vokabularu procjenjuju konkretnijima od glagola, a glagoli konkretnijima od pridjeva (Peti-Stantić, Anđel, Keresteš, Ljubešić, Stanojević i Tonković, 2018, str. 104).

U tom je kontekstu važno upozoriti na razlike u načinu konceptualizacije apstraktnih i konkretnih riječi. Wiemer-Hastings i Xu (2005, str. 724) apstraktne koncepte opisuju kao relacijske jer ih većinom promatramo u odnosu i usporedbi s drugim konceptima, dok konkretne koncepte obično promatramo kroz njihova intrinzična svojstva ili obilježja po kojima su karakteristični. Isto tako, apstraktni su koncepti relativno nespecificirani u usporedbi s konkretnim konceptima jer im se značenje često prilagođava situaciji, odnosno kontekstu u kojem se koriste (Wiemer-Hastings i Xu, 2005, str. 721).

Akademski registar općenito karakterizira veća apstraktnost vokabulara u odnosu na vokabular općega jezika, a možemo pretpostaviti da se unutar akademskoga registra na vokabular preslikava i razlika u konkretnosti među vrstama riječi iz općega jezika, iako ne nužno na jednak način. Odnosno, ako su prema procjenama općega jezika imenice konkretnije od glagola, a glagoli od pridjeva, moguće je da su upravo radi više apstraktnosti akademskoga registra i relacijske prirode glagola kao vrste riječi glagoli u akademskom registru apstraktniji od pridjeva, a pridjevi od imenica.

Osim apstraktnosti registra koja dodatno naglašava razlike između vrsta riječi, u obzir treba uzeti i različite potencijale dubinskoga vezanja različitih vrsta riječi – imenice i pridjevi vežu se na linearnoj razini, a glagoli prema valenciji, čime postaje primjetnom razlika u kompleksnosti tih vrsta riječi. Time su, naime, glagoli svojom relacijskom prirodom dodatno odvojeni od imenica i pridjeva. Stoga su upravo tipovi odnosa koje imenice, pridjevi i glagoli u skladu sa svojim semantičkim, sintaktičkim i pragmatičkim ograničenjima mogu oblikovati ono što uvjetuje nemogućnost korištenja istih tipova zadataka pri procjeni širine i dubine poznavanja značenja svake vrste riječi.

Iako su zadatci u instrumentima poznavanja općega vokabulara, poput Mill Hill ljestvica rječnika, sastavljeni od više vrsta riječi (imenica, glagola, pridjeva i priloga),

razlikama između vrsta riječi unutar instrumenta ne pridaje se dovoljno pozornosti. Zbog toga smo se prilikom konstrukcije instrumenta za mjerenje širine i dubine akademskoga vokabulara usredotočile na pojedinačne vrste riječi unatoč ograničenjima njihove upotrebe koje proizlaze iz njihova unutrašnjega semantičkoga i sintaktičkoga ustrojstva. To je razlog zbog kojeg su imenice ispitane pomoću sinonimnih odnosa, a pridjevi i glagoli kolokacija, odnosno čvrstih i slobodnih veza.

Nacionalni odbor za čitanje SAD-a (2000, str. 4-3) grupirao je najčešće metode poučavanja vokabulara u nekoliko skupina. Te metode uključuju eksplicitno poučavanje vokabulara (usmjereno na definicije), ali i implicitno poučavanje vokabulara koje se oslanja na ekstenzivno čitanje, multimedijalne metode (usmjerene na multisenzoričko učenje), metode povećavanja kapaciteta obrade namijenjene mlađim učenicima i usmjerene na automatizaciju procesa čitanja te metode asocijacije koje učenike potiču na uspostavljanje asocijacija između novoga i postojećega znanja. Jedan od zaključaka izvješća Nacionalnog odbora za čitanje SAD-a (2000, str. 4-27) jest da dob učenika i njihove sposobnosti utječu na učinkovitost svih navedenih metoda poučavanja vokabulara. Iz toga proizlazi da bi sve intervencije koje koriste bilo koju od tih metoda trebale koristiti instrumente koji omogućavaju procjenu učinaka poučavanja na različite dobne skupine i na učenike s različitim sposobnostima.

Na hrvatskom jeziku postoji nekoliko dijagnostičkih instrumenata za mjerenje općega poznavanja vokabulara—hrvatska verzija testa Peabody Picture Vocabulary Test (PPVT-III-HR, Dunn, Dunn, Kovačević, Padovan, Hržica, Kuvač Kraljević, Mustapić, Dobravac i Palmović, 2009), hrvatska verzija Mill Hill ljestvica rječnika (MHV, Raven, Court i Raven, 1994, u Križan i Matešić 2001) te prijevod Manuelova testa iz niza *Inter-American Series* koji koriste neki istraživači (npr., Didović i Kolić-Vehovec, 2009; Pečjak, Kolić-Vehovec, Rončević Zubković i Ajdišek, 2009). Iako su konstrukti općega poznavanja vokabulara i poznavanja općega akademskog vokabulara povezani te je utvrđena visoka korelacija instrumenata koji mjere ta dva konstrukta (npr., Masrai i Milton, 2018., $r = .78$, str. 50), istraživanja pokazuju da je ipak riječ o dva odvojena konstrukta. Primjerice, Townsend i sur. (2012, str. 513) pokazale su da rezultati postignuti na zadacima poznavanja općega akademskog vokabulara objašnjavaju dodatni dio varijance u rezultatima na standardiziranim mjerama akademskoga postignuća, čak i uz kontrolu širine općega vokabulara sudionika istraživanja. Budući da se čestice za mjerenje općega akademskog vokabulara biraju iz udžbeničkih tekstova i akademskih članaka, instrumenti sastavljeni od takvih čestica mogu se koristiti za evaluaciju specifičnih učinaka intervencija usmjerenih na razvoj razumijevanja i korištenja općega akademskog vokabulara. S obzirom na to da su mjere znanja vokabulara trenutačno dostupne na hrvatskom jeziku mjere općega vokabulara, prepoznale smo potrebu za razvojem i provjerom ovoga novog instrumenta koji bi se u Hrvatskoj mogao koristiti za mjerenje poznavanja općega akademskog vokabulara u višim razredima osnovnoškolskoga obrazovanja. Na temelju teorija razvoja čitanja (Indrisano i Chall, 1995) zaključujemo da su viši razredi osnovne škole razdoblje u kojima je ključno raditi

na razvoju akademskoga vokabulara, dok se u nižim razredima ipak treba koncentrirati na automatizaciju same vještine čitanja. Tijekom srednjoškolskoga obrazovanja naglasak može biti na nadograđivanju akademskoga vokabulara na materinskom jeziku uspostavljenoga u temeljnom obliku tijekom dotadašnjega obrazovanja te se učenike može usmjeriti prema složenijim strategijama samostalnoga rada na obogaćivanju vokabulara i strategijama razumijevanja teksta.

Istraživanje

Rezultati hrvatskih učenika na međunarodnim procjenama čitalačke pismenosti, kao što je već rečeno, ukazuju na potrebu provođenja različitih tipova intervencija usmjerenih na razvoj čitalačkih sposobnosti učenika. Prema našim saznanjima ne postoje sustavne intervencije koje se provode u hrvatskim školama, a koje su usmjerene na razvoj učeničkoga poznavanja akademskoga vokabulara ili na čitalačku pismenost, kao što ne postoje ni instrumenti na hrvatskom jeziku kojima se može mjeriti poznavanje općega akademskog vokabulara. S obzirom na to da je valjan instrument za mjerenje širine i dubine poznavanja općega akademskog vokabulara preduvjet za procjenu učinaka potencijalnih školskih intervencija usmjerenih na razvoj akademskoga vokabulara, u okviru rada na projektu MEGahr odlučile smo osmisliti mjerni instrument za procjenu poznavanja općega akademskog vokabulara. Preliminarnu smo verziju oblikovale kako bismo omogućile ostvarivanje specifične svrhe instrumenta – mjerenja širine i dubine poznavanja imenica, glagola i pridjeva u okviru općeg akademskog vokabulara. U ovom radu prikazujemo rezultate ispitivanja te preliminarne verzije instrumenta. Ako se pokaže da je ta verzija instrumenta valjana i pouzdana, instrument se može razviti i koristiti u budućim istraživanjima kao mjera poznavanja općega akademskog vokabulara, kao i za mjerenje učinaka različitih pristupa u poučavanju akademskoga vokabulara.

Očekujemo da će naš mjerni instrument biti osjetljiv na dob, odnosno školski razred, budući da se iskustvo učenika sa složenijim konceptima i riječima kojima se ti koncepti izražavaju povećava sa stupnjem obrazovanja. Prema prethodno iznesenoj pretpostavci da su glagoli koji pripadaju akademskom registru apstraktniji od pridjeva i imenica u tom registru, očekujemo i učinak vrste riječi, odnosno očekujemo najveću uspješnost učenika pri rješavanju podskale s imenicama, a najmanju pri rješavanju podskale s glagolima.

U skladu s razvojem kognitivnih sposobnosti učenika s dobi, moguće je očekivati i interakciju ta dva faktora: učinak vrste riječi može biti različit u različitoj dobi, odnosno školskom razredu, a učinak razreda može biti različit za različite vrste riječi.

Nadalje, očekujemo da će naš instrument biti osjetljiv na spolne razlike koje se često pronalaze u istraživanjima vokabulara i ukazuju na to kako djevojčice obično postižu bolje rezultate na različitim mjerama vokabulara od dječaka (npr., Bornstein, Hahn i Haynes, 2004, str. 296; Eriksson, Marschik, Tulviste, Almgren, Pérez Pereira, Wehberg, Marjanovič-Umek, Gayraud, Kovacevic i Gallego, 2012, str. 337-338; Kaushanskaya, Gross i Buac, 2013, str. 86).

Rezultate prikupljene mjerenjem općega akademskog vokabulara predloženom metodom usporedit ćemo s učeničkim rezultatima na Mill Hill ljestvicama rječnika koje predstavljaju standardiziranu mjeru poznavanja općega vokabulara. Na taj ćemo način provjeriti konvergentnu valjanost mjere poznavanja općega akademskog vokabulara (MOAV) koju predlažemo na temelju ovdje opisanoga istraživanja. Očekujemo da će rezultati na MOAV-u pozitivno te umjereno do visoko korelirati s rezultatima na Mill Hill ljestvicama rječnika.

Naposljetku, očekujemo da će MOAV pokazati adekvatnu unutarnju konzistentnost mjerenu Cronbachovim alfa koeficijentom.

Metoda

Sudionici

U istraživanju su sudjelovali učenici viših razreda (od petog do sedmog) četiriju osnovnih škola iz Varaždinske, Međimurske i Koprivničko-križevačke županije (N = 152). U šestom i sedmom razredu raspodjela djevojčica i dječaka nije bila uravnotežena (Tablica 1).

Tablica 1.

Jedan od mogućih razloga slabijega odaziva dječaka na sudjelovanje u ovom istraživanju leži u činjenici da se čitanje i aktivnosti povezane s čitanjem tradicionalno smatraju prikladnijima za djevojčice nego za dječake. Budući da je sudjelovanje u istraživanju bilo dobrovoljno, tema istraživanja potaknula je manje dječaka na sudjelovanje. Osim toga, djevojčice obično imaju veću motivaciju za čitanje (McGeown, Duncan, Griffiths i Stothard, 2015, str. 561) dok dječaci svoje sposobnosti čitanja procjenjuju prosječnima ili ispodprosječnima i provode manje vremena čitajući (Usen i Mürsepp, 2012, str. 1799). Na osnovi takvih stavova o čitanju očekivana je i smanjena motivacija dječaka za sudjelovanjem u istraživanju u kojem se procjenjuje njihovo poznavanje vokabulara. Budući da su instrumenti za mjerenje poznavanja vokabulara slični školskim ispitima, moguće je da dječaci unaprijed procjenjuju da neće biti uspješni u njihovom rješavanju što dodatno umanjuje želju za sudjelovanjem iako im je rečeno da rezultati neće biti povezani s njihovim identitetom i neće imati utjecaja na njihov školski uspjeh.

Instrumenti

Mjera općega akademskog vokabulara, MOAV

Mjeru općega akademskog vokabulara (MOAV) koja je predmet analize u ovom radu osmislila je posljednja autorica ovoga rada. U prvoj fazi oblikovanja te mjere odabrani su leksemi iz udžbenika za hrvatski jezik, matematiku, povijest, zemljopis te prirodu i društvo/biologiju koji se najčešće koriste u hrvatskim školama u četvrtom, petom i šestom razredu. Nekoliko sveučilišnih profesora te studenata lingvistike i psihologije iz tih je udžbenika odabralo oko 3000 riječi općega akademskog vokabulara visoke frekvencije za koje je procijenjeno da bi mogle biti teže razumljive ili manje poznate

učenicima za koje su udžbenici pisani. U drugoj je fazi taj popis sveden na 1500 riječi. U trećoj je fazi uz pomoć osnovnoškolskih učitelja za svaki razred izdvojeno po 100 najčešće upotrebljivanih riječi s tog popisa. Tim probirom ukupno je izdvojeno 300 riječi. U četvrtoj je fazi od tih 300 riječi odabrano po 20 imenica, pridjeva i glagola. Svi uključeni u probir složili su se s tim da bi svi učenici nakon završene osnovne škole trebali znati značenje i upotrebu tih riječi, zato što se radi o riječima koje se široko upotrebljavaju i u javnom, a ne samo u akademskom diskursu. Od tih su riječi oblikovani zadatci za mjerenje širine i dubine vokabulara koji čine osnovicu MOAV-a.

MOAV se, dakle, u svojem preliminarnom obliku sastoji od 60 čestica koje su organizirane kao mjera širine i dubine vokabulara. Točan odgovor za svaku česticu nosi jedan bod. Prvi dio MOAV-a usmjeren je na procjenu širine vokabulara, tj. pronalaženje sinonima imenica (20 čestica) među pet ponuđenih odgovora. Posebnu pažnju posvetile smo odabiru distraktora po načelu tvorbene, semantičke i fonološke sličnosti sa zadanom česticom. U Tablici 2 prikazan je primjer čestice iz te podskale.

Tablica 2.

Riječi „rukohvat” i „prihvat” odabrane su prema načelu tvorbene i fonološke sličnosti sa zadanom česticom. Riječ „neuspjeh” može se značenjski vezati za zadanu česticu jer pothvat može biti neuspješan ili uspješan. Isto tako, naš „pothvat” može izazvati određeni „učinak”, ali „učinak” i „pothvat” nisu sinonimi. Točan odgovor u ovom zadatku je riječ „akcija”.

Drugi dio MOAV-a sastoji se od dvije podskale istoga tipa gdje je trebalo prepoznati najmanje prikladne (ili sasvim neprikladne) kolokacije pridjeva i glagola. Svaka podskala sadrži po 20 čestica. Uz svaku je česticu ponuđeno pet mogućih odgovora od kojih samo jedan može biti prepoznat kao neodgovarajući, odnosno neprikladan. Taj je tip zadatka odabran kao mjera dubine vokabulara jer su sudionici istraživanja trebali prepoznati ograničenja upotrebe riječi. U Tablici 3 prikazani su primjeri čestica korišteni u tom tipu zadatka.

Tablica 3.

U dva smo navedena zadatka distraktore birale tako da među riječima koje su ovjerene kolokacije bude zahvaćena što već raznolikost značenjskih nijansi i upotreba pojedine riječi, pri čemu se vodilo računa i o denotativnoj i o konotativnoj razini značenja.

Pridjev „neznatan” označava jedva vidljiv odmak od željenoga ili očekivanoga cilja nečega što možemo mjeriti. Tako možemo reći da je „razlika” između dva kandidata na ispitu bila neznatna, da je „pogreška” pri mjerenju ili lijepljenju bila neznatna, da u nešto trebamo staviti tek neznatnu „količinu” šećera ili da je neznatna „udaljenost” između neke dvije točke na obzoru koje vidimo pred sobom. Međutim, „kombinaciju” ne opisujemo veličinom, već kvalitetom—dobra, poznata, neočekivana i slično. To je jedno od ograničenja značenja riječi koje učenici trebaju prepoznati kako bi uspješno riješili ovaj zadatak.

Glagol „ugroziti” upućuje na to da se nešto dovodi u opasnost i obično se koristi s riječima koje imaju pozitivnu vrijednosnu konotaciju. Može se ugroziti nečiji život, zdravlje, sigurnost, prijateljstvo, odnosno, ugrožavanjem negativno utječemo na svoju ili tuđu dobrobit, slabimo svoje zdravlje ili postojanost prijateljstva. Za isti oslabljujući učinak na nešto negativno poput bolesti nećemo reći da smo je „ugrozili”, već da smo bolest pobijedili ili ublažili. Učenici bi trebali prepoznati ograničenja upotrebe zadanoga glagola na pojavljivanje s imenicama pozitivne vrijednosne konotacije.

U nastavku rada svaku od tri podskale analiziramo zasebno.

Mill Hill ljestvice rječnika

Mill Hill ljestvice rječnika (MHV, Raven i sur., 1994, u Križan i Matešić 2001) dolaze u dva oblika s dvije skupine riječi i dvije podskale. Svaka je skupina riječi u jednom obliku instrumenta u podskali koju čini zadatak definiranja značenja riječi, a u drugom obliku instrumenta u podskali prepoznavanja sinonima odabirom među šest mogućih odgovora. Jednu skupinu riječi čini 14 glagola, 9 imenica, 7 pridjeva i 2 priloga, a drugu 12 glagola, 7 imenica, 12 pridjeva i jedan prilog. Dakle, svaka je podskala sastavljena od ukupno 32 čestice i uključuje imenice, pridjeve, glagole i priloge. Svaki točan odgovor nosi jedan bod. U ovom istraživanju, Cronbachovi alfa koeficijenti za prvi oblik Mill Hill ljestvice rječnika bili su .75 za podskalu definiranja riječi i .86 za podskalu prepoznavanja sinonima, a za drugi oblik .86 za podskalu definiranja riječi i .74 za podskalu prepoznavanja sinonima. Cronbachov alfa koeficijent za ukupan rezultat na oba oblika MHV-a bio je .88.

Postupak

Istraživanje je odobrilo etičko povjerenstvo Odsjeka za psihologiju Filozofskog fakulteta Sveučilišta u Zagrebu i Ministarstvo znanosti i obrazovanja Republike Hrvatske. Prije početka istraživanja prikupile smo potpisane roditeljske pristanke od svih učenika koji su sudjelovali u istraživanju. Na dan prikupljanja podataka učenici su s potpisanim roditeljskim pristancima dodatno dali usmeni pristanak za sudjelovanje u istraživanju.

Sudionici su prvo ispunili standardiziranu mjeru znanja općega vokabulara, Mill Hill ljestvice rječnika, a potom i MOAV test, mjeru širine i dubine općega akademskog vokabulara. Učenici su oba testa ispunjavali anonimno, na papiru. Redoslijed ispunjavanja bio je isti za sve učenike. Vrijeme rješavanja nije bilo ograničeno, no većina učenika ispunila je Mill Hill ljestvice rječnika i MOAV test za 15 do 20 minuta. Nakon ispunjavanja zadataka poznavanja vokabulara, organizirale smo radionice za učenike i učitelje s primjerima vježbi za poticanje razvoja vokabulara kojima se učenici mogu služiti samostalno, a mogu se upotrebljavati i u nastavi.

Rezultati

Za analizu podataka korišten je IBM SPSS v. 26. Deskriptivni statistički pokazatelji rezultata učenika na MOAV-u i MHV-u prikazani su u Tablici 4. Rezultati su odvojeni po razredu i podskali. Najviši mogući rezultat na svakoj od tri podskale MOAV-a

(sinonimi imenica, kolokacije pridjeva i glagola) iznosi 20 bodova, a maksimalni ukupni rezultat na MOAV-u 60 bodova. Najviši mogući rezultati na svakoj od dviju skala MHV-a (definiranje značenja riječi i prepoznavanje sinonima) iznosi 32 boda, a maksimalni ukupni rezultat na MHV-u 64 boda.

Normalitet distribucija rezultata učenika različite dobi (razreda) na MOAV-u i MHV-u provjerili smo vizualnom inspekcijom oblika distribucija te izračunavanjem z-vrijednosti za indekse asimetričnosti (*skewness*) i spljoštenosti (*kurtosis*) (Field, 2009; Petz, Kolesarić i Ivanec, 2012). Vizualno, sve su distribucije imale oblik normalne krivulje. Indeksi asimetričnosti (Tablica 4) pokazali su da su sve distribucije negativno asimetrične, osim distribucije rezultata učenika petog razreda na podskali definiranja značenja riječi MHV-a, koja je pozitivno asimetrična. Na MOAV-u u petom razredu negativna asimetričnost distribucije statistički je značajna (veća od $+/-1,96$) samo za podskalu pridjevskih kolokacija, dok je u šestom razredu negativna asimetričnost distribucije statistički značajna za podskalu sinonima imenica i ukupni rezultat, a u sedmom razredu za sve podskale i ukupni rezultat. Na MHV-u je negativna asimetričnost statistički značajna za distribuciju rezultata učenika šestog razreda na podskali prepoznavanja sinonima te za distribuciju ukupnoga rezultata i rezultata na podskali definiranja značenja riječi učenika sedmog razreda. Indeksi spljoštenosti (Tablica 4) u petom i šestom razredu nisu statistički značajni (z-vrijednosti manje od $+/-1,96$), osim za podskalu sinonima imenica MOAV-a u šestom razredu na kojoj je distribucija rezultata leptokurtična. U sedmom razredu distribucije rezultata na sve tri podskale MOAV-a su leptokurtične, ali spljoštenost distribucije ukupnoga rezultata na MOAV-u ne odstupa od normalne krivulje.

Budući da je provjera normaliteta distribucija pokazala da distribucije nekih varijabli statistički značajno odstupaju od normalne krivulje, podatke smo analizirale i parametrijskim i neparametrijskim postupcima te smo usporedile njihove rezultate. Rezultati dobiveni parametrijskim i neparametrijskim analizama nisu se razlikovali, o čemu detaljnije izvještavamo u nastavku prikaza rezultata.

Tablica 4.

Testiranje učinka spola

Učinke spola na ukupni rezultat i rezultate na pojedinačnim podskalama MOAV-a testirali smo odvojeno od učinaka dobi, odnosno razreda, budući da zbog neujednačenoga broja djevojčica i dječaka u šestim i sedmim razredima nije bilo opravdano testirati interakciju spola i razreda. Kako bi se testirali učinci spola, prvo su Leveneovim testom provjerene jednakosti varijanci rezultata u uzorcima djevojčica i dječaka. Rezultati su pokazali da se varijance rezultata na MOAV-u u ta dva uzorka statistički značajno razlikuju na podskalama pridjevskih i glagolskih kolokacija (vrijednosti Leveneova testa za rezultate na podskalama sinonima imenica, pridjevskih kolokacija i glagolskih kolokacija te ukupan rezultat na MOAV-u bile su, redom: $F = 3,35$, $p = .069$; $F = 5,05$, $p = .026$; $F = 4,74$, $p = .031$ i $F = 3,78$, $p = .054$; u svim analizama $df_1 = 1$, $df_2 = 144$).

Zbog nejednakosti varijanci, spolne razlike u rezultatima na MOAV-u prvo su provjerene neparametrijskim Mann-Whitney testom, a zatim i parametrijskim t-testom za nezavisne uzorke. Rezultati svih provedenih testova bili su statistički značajni, pokazujući da djevojčice postižu statistički značajno više rezultate na svim podskalama MOAV-a i ukupno u cijelom instrumentu. Kako se rezultati Mann-Whitney testa i t-testa nisu razlikovali, prikazat ćemo samo rezultate t-testa.

Na podskali prepoznavanja sinonima imenica djevojčice su ostvarile prosječan rezultat od 14,39 bodova (SD = 3,13), a dječaci 13,02 (SD = 3,69); $t(144) = 2,29, p = .023$; Cohenov $d = 0,40$. Na podskali pridjevskih kolokacija prosječan rezultat djevojčica bio je 13,57 (SD = 3,39), a dječaka 11,91 (SD = 4,33); $t(144) = 2,26, p = .027$; Cohenov $d = 0,42$. Prosječan rezultat na podskali glagolskih kolokacija za djevojčice je bio 12,27 (SD = 3,70), a za dječake 9,45 (SD = 4,62); $t(144) = 3,58, p = .001$; Cohenov $d = 0,67$. Prosječan ukupan rezultat djevojčica na MOAV-u bio je 40,24 (SD = 8,66), a dječaka 34,39 (SD = 10,66); $t(144) = 3,49, p = .001$; Cohenov $d = 0,60$. Vrijednosti Cohenova d indeksa pokazuju da je učinak spola na rezultate na MOAV-u srednje veličine.

Testiranje učinka razreda i vrste riječi

Učinke razreda i vrste riječi na rezultate na MOAV-u testirale smo neparametrijskim Kruskal-Wallisovim (učinak razreda) i Friedmanovim (učinak vrste riječi) testom te parametrijskom mješovitom analizom varijance u kojoj je vrsta riječi bila unutargrupna varijabla (jer su svi učenici rješavali sve podskale), a razred međugrupna varijabla. Obje nezavisne varijable imale su tri razine (imenice, pridjevi i glagoli, te peti, šesti i sedmi razred). Kao i kod testiranja učinaka spola, rezultati neparametrijskih i parametrijskih analiza bili su sukladni te ćemo prikazati samo rezultate mješovite analize varijance, koja je, uz testiranje glavnih učinaka razreda i vrste riječi, omogućila i testiranje njihove interakcije.

Budući da je Mauchlyjev test sferičnosti bio statistički značajan ($p = .045$), navest ćemo p vrijednosti s Greenhouse-Geisser korekcijom za glavni efekt vrste riječi, razreda i interakciju vrste riječi i razreda. Kao što smo već napomenule, budući da je broj djevojčica i dječaka koji su sudjelovali u istraživanju u pojedinačnim razredima bio nejednak, nismo testirale interakciju spola, razreda i vrste riječi.

Mješovita analiza varijance pokazala je statistički značajan glavni učinak vrste riječi ($F[1,92, 286,23] = 43,22, p < .001$, parcijalna $\eta^2 = .225$). Taj rezultat pokazuje da vrsta riječi ima učinak na učeničke rezultate na MOAV-u i objašnjava 22,5 % ukupne varijance. Kao što je vidljivo iz Tablice 4, učenici su bili najuspješniji u rješavanju podskale sa sinonimima imenica, zatim u rješavanju pridjevskih pa onda glagolskih kolokacija. Interakcija vrste riječi i razreda također je bila statistički značajna ($F[3,84, 286,23] = 2,71, p = .032$, parcijalna $\eta^2 = .035$), iako je objasnila razmjerno mali postotak varijance rezultata. Slika 1 i prosječne vrijednosti u Tablici 4 pokazuju da je učinak vrste riječi bio najveći među učenicima petog razreda, a najmanji među učenicima sedmog razreda te da je učinak razreda bio najveći za glagole, a najmanji za imenice.

Slika 1.

Analiza je također pokazala i značajan glavni učinak razreda ($F [2, 149] = 23,68$, $p < .001$, parcijalna $\eta^2 = .241$), što pokazuje da su stariji učenici postigli bolje rezultate na MOAV-u, odnosno da razred objašnjava 24,1 % ukupne varijance.

Scheffeovi *post hoc* testovi pokazuju da su razlike među razredima statistički značajne na svim razinama (5.—7., $p < .001$; 5.—6., $p = .001$; 6.—7., $p = .010$).

Korelacije između Mill Hill ljestvica rječnika i MOAV-a

Povezanosti rezultata na MOAV-u i rezultata na MHV-u u ukupnom uzorku ispitali smo Spearmanovim neparametrijskim i Pearsonovim parametrijskim koeficijentima korelacije. Budući da su se Spearmanovi i Pearsonovi koeficijenti korelacije tek neznatno razlikovali, prikazat ćemo samo Pearsonove koeficijente (Tablica 5). Utvrđena je visoka i statistički značajna pozitivna korelacija između ukupnoga rezultata na MOAV-u, kojim se procjenjuje poznavanje općega akademskog vokabulara, i ukupnoga rezultata na Mill Hill ljestvicama rječnika, kojima se procjenjuje poznavanje općega vokabulara. Umjerene i statistički značajne pozitivne korelacije bile su prisutne i između podskala MOAV-a i podskale prepoznavanja sinonima na Mill Hill ljestvicama rječnika. Korelacije između podskala MOAV-a i podskale definiranja riječi na Mill Hill ljestvicama rječnika bile su nešto slabije, no i dalje umjerene i statistički značajne.

Sve podskale MOAV-a u međusobnim su pozitivnim i umjerenim korelacijama te u visokim korelacijama s ukupnim rezultatom na MOAV-u.

Tablica 5.

Unutarnja konzistentnost MOAV-a

Unutarnja konzistentnost MOAV-a mjerena Cronbachovim alfa koeficijentom iznosi .90. Alfa koeficijenti za pojedinačne podskale bili su .78 (sinonimi imenica), .79 (pridjevske kolokacije) i .79 (glagolske kolokacije). Korelacije među česticama kretale su se od niskih negativnih do visokih pozitivnih. Većina korelacija bila je između .10 i .30 (od ukupno 190 korelacija među česticama za svaku podskalu, u tom je rasponu bilo 54 % korelacija za podskalu glagolskih kolokacija te 53 % korelacija za podskalu sinonima imenica i podskalu pridjevskih kolokacija). Znatno dio korelacija bio je veći od .30 (21 % korelacija za podskalu pridjevskih kolokacija te po 16 % korelacija za podskale glagolskih kolokacija i sinonima imenica), pri čemu je maksimalna korelacija za podskalu pridjevskih kolokacija bila .70, za podskalu sinonima imenica .56, a za podskalu glagolskih kolokacija .45. Dio korelacija bio je nizak i pozitivan (u rasponu od .00 do .10): 24 % za podskalu sinonima imenica, 20 % za podskalu glagolskih kolokacija i 17 % za podskalu pridjevskih kolokacija. Najmanji udio korelacija bio je nizak negativan (u rasponu od -.01 do -.15, uz iznimku jedne korelacije između dvaju zadataka podskale sinonima imenica koja je iznosila -.29): 9 % korelacija za podskale pridjevskih i glagolskih kolokacija te 7 % za podskalu sinonima imenica.

Niske i/ili negativne korelacije među nekim česticama najvećim se dijelom nalaze uz čestice čije bi isključivanje iz MOAV-a prema analizama doprinijelo boljoj unutarnjoj konzistentnosti. Primjerice, uz čestice „oskudica” i „dobrobit” na podskali sinonima imenica nalazi se 79 % svih negativnih korelacija na toj podskali, uključujući gore spomenutu najveću negativnu korelaciju u instrumentu, te 30 % svih niskih pozitivnih korelacija na podskali sinonima imenica. Vizualnom inspekcijom korelacijske matrice za podskalu imenica primjetno je da niske korelacije prevladavaju u nekoliko čestica („ustanova”, „pothvat”, „odredba”, „oskudica”, „dobrobit”) što upućuje na to da bi te zadane čestice trebalo zamijeniti ili preoblikovati ponuđene odgovore. Analiza pouzdanosti skale potvrđuje taj zaključak za česticu „dobrobit”, čije bi isključivanje iz MOAV-a rezultiralo nešto većim koeficijentom unutarnje konzistentnosti – u tom bi slučaju Cronbachov alfa koeficijent podskale sinonima imenica bio .79 u odnosu na sadašnjih .78. Na podskali pridjevskih kolokacija 65 % svih negativnih korelacija te 30 % svih niskih pozitivnih nalazi se uz dvije čestice („postojana” i „prisebna”). Vizualna inspekcija korelacijske matrice za podskalu pridjevskih kolokacija pokazuje da je uz čestice „časna”, „nezatna”, „ravnomjeran”, „postojana” i „prisebna” vezano najviše niskih korelacija, a analiza pouzdanosti skale pokazuje da bi isključivanje čestica „postojana” i „prisebna” najviše doprinijelo boljoj unutarnjoj konzistentnosti te podskale. Cronbachov alfa koeficijent podskale pridjevskih kolokacija bio bi .80, odnosno .81 u odnosu na sadašnjih .79. Vizualna inspekcija korelacijske matrice za podskalu glagolskih kolokacija otkriva da niske korelacije prevladavaju uz čestice „obilovati”, „modelirati”, „ovladati”, „predvidjeti” i „prerađivati”. Ponovno je najveći dio niskih korelacija vezan uz dvije čestice, „predvidjeti” i „prerađivati”, 44 % negativnih i 33 % pozitivnih niskih korelacija. Analiza pouzdanosti skale pokazuje da bi isključivanje čestica „modelirati” i „predvidjeti” najviše doprinijelo unutarnjoj konzistentnosti podskale glagolskih kolokacija. Cronbachov alfa koeficijent podskale glagolskih kolokacija jednako bi porastao za obje čestice - na .80 u odnosu na sadašnjih .79.

Iako niske i/ili negativne korelacije nisu uobičajene u instrumentima koji bi trebali mjeriti isti konstrukt i pokazuju da bismo neke čestice MOAV-a trebali zamijeniti ili prilagoditi, one istovremeno govore o tome da određena razina poznavanja širine i/ili dubine značenja određenih riječi ne podrazumijeva nužno poznavanje širine i/ili dubine značenja drugih riječi, pogotovo kada se radi o specifičnom registru kao što je slučaj s općim akademskim vokabularom koji MOAV mjeri. Treba naglasiti i da u našem uzorku korelacijska matrica MHV-a također sadrži određeni dio niskih pozitivnih i negativnih korelacija među česticama. Postotak niskih pozitivnih korelacija (u rasponu od .00 do .10) na prvom obliku MHV-a iznosio je 32 % na podskali definiranja riječi i 17 % na podskali prepoznavanja sinonima, dok je postotak niskih negativnih korelacija (od -.001 do -.10) na prvoj podskali bio 13 %, a na drugoj 11 %. Negativne korelacije među česticama na prvom obliku MHV-a u rasponu od -.10 do -.35 zauzimaju 9 % ukupnih korelacija na podskali definiranja riječi i 3 % na podskali prepoznavanja sinonima. Postotak niskih pozitivnih korelacija (u rasponu od .00 do

.10) na drugom obliku MHV-a iznosio je 20 % na podskali definiranja riječi i 18 % na podskali prepoznavanja sinonima, postotak niskih negativnih korelacija u rasponu od (-.001 do -.10) bio je 7 % na prvoj, a 19 % na drugoj podskali, dok je postotak negativnih korelacija u rasponu od -.10 do -.30 bio 3 % na podskali definiranja riječi i 5 % na podskali prepoznavanja sinonima.

Rasprava

Svrha ovoga istraživanja bila je procijeniti osjetljivost, pouzdanost i valjanost preliminarne verzije mjere poznavanja općega akademskog vokabulara koja je testirana u uzorku učenika viših razreda osnovnih škola. Ta mjera općega akademskog vokabulara, MOAV, sastojala se od tri podskale od kojih je svaka bila sastavljena od 20 čestica tipa višestrukoga izbora—podskala s prepoznavanjem sinonima imenica, podskala s prepoznavanjem najmanje prikladnih pridjevskih kolokacija te podskala s prepoznavanjem najmanje prikladnih glagolskih kolokacija. Pri korištenju takvih tipova zadataka višestrukoga izbora mora se uzeti u obzir i mogućnost lažno pozitivnoga ili lažno negativnoga rezultata u pojedinačnim česticama budući da učenici mogu znati značenje čestice, ali ne znati značenje svih ponuđenih odgovora zbog čega mogu izabrati pogrešan odgovor ili izabrati točan odgovor nasumičnim pogađanjem. Detaljnijim uvidom u netočne odgovore pokazalo se da su učenicima nepoznate vrlo formalne dimenzije značenja ili one koje su daleko od njihove svakodnevice (npr., *neovisna procjena*). Najčešći distraktori koje su učenici birali pri prepoznavanju sinonima imenica bili su kontekstualno povezani s početnom česticom. Primjerice, učenici su kao sinonim zadanoj čestici „citat” vrlo često birali riječ „knjiga”, a točan je odgovor bila riječ „navod”. Isto tako, birali su i distraktore koji su uvršteni u zadatak radi tvorbene i fonološke sličnosti s početnom česticom. Primjerice, kao sinonim riječi „pohvat” birali su „rukohvat” umjesto „akcija.”

Rezultati t-testa za nezavisne uzorke korištenoga za testiranje učinka spola na rezultate na MOAV-u pokazuju da djevojčice postižu bolje rezultate na svim podskalama MOAV-a, odnosno da su uspješnije od dječaka u prepoznavanju sinonima imenica, kao i pri prepoznavanju neprikladnih kolokacija pridjeva i glagola općega akademskog vokabulara. Kao što je već rečeno, brojna istraživanja ukazuju na to da djevojčice postižu bolje rezultate od dječaka na raznovrsnim mjerama vokabulara od rane dobi do adolescencije (npr., Marjanović-Umek i Fekonja-Peklaj, 2017, str. 106; Simonsen, Kristoffersen, Bleses, Wehberg i Jørgensen, 2014, str. 18). U kontekstu hrvatskoga jezika, rezultati Didović i Kolić-Vehovec (2009, str. 106) pokazuju da je u završnom razredu osnovne škole rječnik djevojčica bogatiji od rječnika dječaka. Isto tako, rezultati istraživanja Keresteš, Lugomer Armano, Čičko, Goldstein, Jerković, Juras, Lindić, Matuško, Pere, Pešorde, Polak, Šoštarić, Šutić i Vrdoljak (2020) s nešto mlađim učenicima osnovne škole (4.—7. razred) pokazuju da djevojčice i u tom razdoblju školovanja imaju bogatiji vokabular od dječaka. Ipak, na temelju nekih istraživanja može se zaključiti da se uloga spola u razvoju vokabulara mijenja s dobi, odnosno

tijekom školovanja. Istraživanje Toivainena, Papageorgioua, Tosto i Kovas (2017, str. 84) pokazalo je da dječaci postižu bolje rezultate na mjerama vokabulara u dobi od 10 i 12 godina, a nisu pronađene statistički značajne razlike između djevojčica i dječaka u dobi od 14 i 16 godina. Slično tome Coxhead, Nation i Sim (2015, str. 131) nisu pronašli nikakve razlike u širini vokabulara djevojčica i dječaka između 13. i 18. godine. Zbog neujednačenoga broja djevojčica i dječaka različite dobi (razreda) u našem istraživanju nije bilo opravdano testirati interakcijske učinke spola i razreda. Te je učinke potrebno ispitati u budućim istraživanjima. Ipak, bolji rezultati djevojčica u našem istraživanju mogu se shvatiti kao indikacija za to da pristup poučavanju vokabulara tijekom školovanja treba djelomično diferencirati prema spolu, odnosno, da bi dječaci u osnovnoj školi mogli imati veće dobrobiti od većega broja prilika za rad na razvoju vokabulara.

Značajan glavni učinak razreda koji je pokazala mješovita analiza varijance potvrđuje to da je MOAV osjetljiv na napretke u rječniku do kojih dolazi s porastom dobi i izloženosti formalnom obrazovanju. To je osnovni preduvjet za korištenje bilo kojeg instrumenta u svrhu mjerenja vokabulara budući da stariji učenici znaju više riječi, imaju više znanja o svijetu i duže se služe objasnidbenim tekstovima u sklopu obrazovanja. Činjenica da stariji učenici postižu bolje rezultate od mlađih na svakoj podskali MOAV-a potvrđuje valjanost instrumenta i pokazuje kako on može detektirati dobro dokumentirane dobne razlike koje su rezultat školovanja i sazrijevanja (npr., Biemiller i Slonim, 2001, str. 505; Keresteš i sur., 2020, str. 55; Song, Su, Kang, Liu, Zhang, McBride-Chang, Tardif, Li, Liang, Zhang i Shu, 2015, str. 10; Verhoeven, van Leeuwe i Vermeer, 2011, str. 15).

Rezultati mješovite ANOVA-e također pokazuju da vrsta riječi utječe na rezultate prikupljene ispitivanjem pomoću MOAV-a. Učenici su bili najuspješniji u rješavanju podskale s imenicama, zatim pridjevima i naposljetku s glagolima. Takav rezultat ukazuje na to da u okviru intervencija usmjerenih na vokabular različitim vrstama riječi treba pristupati selektivno i diferencirano: na temelju istraživanja može se pretpostaviti da bi intervencije koje različitim vrstama riječi pristupaju različito bile korisnije od neselektivnih intervencija. Pri tome, kad je riječ o glagolima, posebnu pozornost vrijedi usmjeriti na dubinu vokabulara. Dva su razloga za to: (1) specifičnost hrvatskoga jezika koji je tvorbeno bogat i omogućava mijenjanje značenja glagola dodavanjem prefiksa, čime se utječe i na njihovu valenciju i sposobnosti vezanja te (2) psiholingvistički faktori koji utječu na kasnije formiranje cjelovitoga značenja glagola u usporedbi s imenicama i pridjevima. Čak bi i unutar jedne vrste riječi (primjerice, glagola) mogla postojati potreba za drugačijim pristupom s različitim dobnim skupinama, budući da 3,5 % varijance objašnjava interakcija vrste riječi i razreda koji učenici pohađaju, odnosno njihove dobi. Ipak, treba reći da su na razlike u rezultatima po vrstama riječi možda utjecali tipovi zadataka, a ne samo vrsta riječi. Premda se može činiti da razlike u tipu zadataka u okviru preliminarne verzije MOAV-a (sinonimi i ograničenje kolokacijske upotrebe) na određenu vrstu riječi onemogućuju usporedivost na svim razinama, treba naglasiti kako je instrument svjesno konstruiran na taj način jer smo

se vodile karakteristikama pojedinačnih vrsta riječi i tipičnoga konteksta u kojem se pojavljuju. Kao što je već istaknuto, glagoli se vežu s drugim riječima prema svojoj relacijskoj prirodi i valencijskim svojstvima, pridjevi su u službi opisivanja imenica, a imenice mogu stajati samostalno u punini svojega značenja.

Neki autori, kao na primjer Vermeer (2001, str. 231), smatraju da se dimenzije širine i dubine vokabulara konceptualno ne razlikuju te da bi stoga između mjera širine i dubine vokabulara trebale postojati visoke korelacije. Međusobne korelacije između podskala MOAV-a su umjerene, što ukazuje na potrebu za dodatnim promišljanjem i provjerom izbora čestica uključenih u svaku podskalu. Ipak, visoka pozitivna korelacija između MOAV-a, kojim se mjeri poznavanje specifičnoga jezičnog registra, i Mill Hill ljestvice rječnika kojom se mjeri opće poznavanje vokabulara, u skladu je s rezultatima drugih istraživanja toga tipa (npr., Masrai i Milton, 2018, str. 50). Ta visoka korelacija istovremeno pokazuje dobru konvergentnu valjanost nove mjere općega akademskog vokabulara, MOAV-a.

Kao što je već rečeno, korelacije između sve tri podskale MOAV-a i podskale identifikacije sinonima u sklopu Mill Hill ljestvica rječnika (sve zadatci višestrukoga izbora) više su od korelacija sve tri podskale MOAV-a i podskale definiranja riječi na Mill Hill ljestvicama rječnika. Budući da su više korelacije između dva instrumenta prisutne između zadataka istoga tipa, možemo zaključiti da se pri usporednom korištenju više dijagnostičkih mjera kojima je cilj procjena sličnoga konstrukta trebaju uzeti u obzir i tipovi zadataka od kojih su te mjere sastavljene.

Uz sve navedeno, u provedenom je istraživanju za MOAV dobivena i visoka unutarnja konzistentnost mjerena Cronbachovim alfa koeficijentom, kako za instrument u cjelini, tako i za pojedinačne podskale. Time smo pokazale kako je ovaj oblik instrumenta, odnosno oblik uvrštenih zadataka, primjenjiv u mjerenju dubine i širine općega akademskog vokabulara. Te bi se vrste zadataka mogle testirati s drugim česticama (riječima) na drugim razinama obrazovne vertikale sukladno posebnim zahtjevima i ciljevima pojedinih istraživanja i intervencija uz metrijsku provjeru novih čestica.

Zaključak

Budući da se sve više govori o visokoj korelaciji između širine i dubine vokabulara i razumijevanja teksta (npr., Didović i Kolić-Vehovec, 2009, str. 109; Kieffer i Lesaux, 2007, str. 138; Keresteš i sur., 2020, str. 61; Peti-Stantić 2019, str. 96), raste i svijest o važnosti sustavne analize vokabulara u udžbenicima i usporedbe toga materijala s leksičkim resursima učenika u određenoj dobi. Nesrazmjer između leksičkoga znanja učenika i jezičnoga materijala korištenoga u njihovim udžbenicima rezultirat će lošijim razumijevanjem pročitana teksta. Stoga valja naglasiti da bi jezični materijal u udžbenicima trebao biti oblikovan tako da je istovremeno dovoljno razumljiv učenicima, ali i da dodatno potiče njihov jezični razvoj.

Osjetljive, pouzdane i valjane mjere vokabulara preduvjet su za istraživanje razvoja rječnika i pouzdanu procjenu učinaka različitih intervencija usmjerenih na razvoj

vokabulara. Budući da je instrument koji smo ispitali u ovom istraživanju, mjera procjene poznavanja općega akademskog vokabulara učenika viših razreda osnovne škole, osjetljiv na razlike u dobi i spolu te pokazuje visoku pozitivnu korelaciju sa standardiziranom mjerom općega znanja vokabulara i ima dobru unutarnju konzistentnost (pouzdanost), najvažniji zaključak ove analize jest da se MOAV može koristiti u temeljnim znanstvenim istraživanjima u kojima se ispituje razvoj vokabulara i čimbenici koji na njega utječu te u procjeni intervencija usmjerenih na razvoj znanja općega akademskog vokabulara učenika viših razreda osnovne škole u Hrvatskoj. Uz to, osim izričite usmjerenosti na akademski registar koja ne postoji među trenutno dostupnim instrumentima na hrvatskom jeziku, treba naglasiti vrijednost usmjerenosti instrumenta na različite vrste riječi. U tom je smislu njegova dodatna prednost to što rezultati pokazuju učinak vrste riječi. Na kraju, kako zbog pažljivoga izbora riječi, tako i zbog pažljivoga izbora distraktora u zadacima višestrukoga izbora koji doprinosi kvaliteti zadataka, MOAV je čvrsto teorijski utemeljen instrument u usporedbi s drugim mjerama vokabulara. Ipak, verzija instrumenta prikazana u ovom radu tek je preliminarna i treba ju provjeriti na novim, većim i reprezentativnijim uzorcima te na temelju dobivenih rezultata dalje usavršavati.

Na kraju ističemo da bi, uzevši u obzir preporuke Europskog parlamenta i Vijeća Europe (2006) o razvoju komunikativne kompetencije na materinskom jeziku, analize poput ove trebale služiti i kao smjernice za obrazovne politike. Mnoge obrazovne institucije i tvorcima obrazovnih politika aktivno potiču projekte koji su usmjereni na razvoj učeničke komunikativne i čitalačke kompetencije, posebno akademske pismenosti (npr., FörMig, BiSS [Njemačka], IAL [USA], SALAD [Zapadni Sussex]). Kao primjer navodimo tek jedan znanstveni projekt toga tipa u kojem se cjelovito pristupa akademskoj pismenosti. Riječ je o projektu koji razvija skupina istraživača predvođena Paolom Uccelli i njezinim kolegama (npr., Uccelli, Barr, Dobbs, Galloway, Meneses i Sánchez, 2015). Oni su usmjerili pažnju na definiranje skupa jezičnih vještina koje nadilaze akademski vokabular, a nužne su i relevantne u obrazovnom kontekstu te su u okviru projekta razvili instrument za procjenu akademskih jezičnih vještina učenika viših razreda osnovnih škola (CALS-I) kojim se procjenjuje nekoliko dimenzija sposobnosti vezanih za razumijevanje teksta. Među ostalim, procjenjuje se razumijevanje strukture argumenata, zaključivanje na osnovi lokalnoga i globalnoga konteksta, razumijevanje sintaktičkih struktura, semantike i morfologije. Takvi sveobuhvatni instrumenti, uz specifična istraživanja koja smo započeli, mogli bi poslužiti kao model razvoja mjera akademskoga vokabulara i akademskih jezičnih sposobnosti na hrvatskom jeziku.

Izjava o dostupnosti podataka:

Podatci prikupljeni u ovom istraživanju javno su dostupni na servisu *figshare* na adresi <https://figshare.com/s/f6b21e70f4b59dbb8733>