# NEW INSIGHTS AND METHODS OF VOCABULARY ACQUISITION IN LATIN CLASSES [*]

*Andrea Beyer*[**]

0000-0002-8468-6411

Humboldt-Universität zu Berlin

[*] This paper extends a presentation at the *Congreso internacional sobre la enseñanza y el aprendizaje de lenguas clásicas* (2.9.2019, Bogotá). The underlying research project CALLIDUS (Computer-Aided Language Learning: Vocabulary Acquisition in Latin using Corpus-based Methods) is funded by the German Research Foundation (project number 316618374).

[**] beyeranz@hu-berlin.de

**Abstract**

Learning a historical language is in itself different from learning a modern language in view of emphasizing the work on texts instead of everyday communication. Therefore, not only the expectations and motivation differ, but also the teaching methodology. Whereas learners of modern languages focus on language production, learners of Latin read or translate their texts. Because of the overall low frequency of occurrence of a Latin word or a phrase in this kind of learning environment most students are often unfamiliar with a given word and therefore finally unable to translate the texts. To tackle this underlying problem of Latin classes an interdisciplinary research project conducted different studies using a data-driven learning (DDL) approach. So far, the findings are very multifaceted and sometimes even surprising, e.g. that the majority of students fail to lemmatise words correctly though they have learned Latin for four years or more.

**Keywords**: *Latin; vocabulary acquisition; data-driven learning (DDL); corpus-based exercises; historical language*.

## 1 Introduction

In contrast to other countries, it is common to learn Latin as a foreign language in German high schools. Following English and French, the historical language Latin is still the third most important foreign language with about 600000 students[1], esp. in grades 7 to 10. While the numbers of students have decreased rapidly for the last ten years, the future of Latin in secondary schools is as fiercely discussed (e.g. Behrendt & Korn, 2016; Kipf & Kuhlmann, 2015) as the conclusions drawn from this (e.g. Beyer, Kipf, Liebsch, & Zimmermann, 2019; Hensel, 2017; Korn & Kuhlmann, 2017). In fact, there are two main questions less solved than ever:

- First, what does a student gain by learning Latin in high school?
- Second, why do the students have so many problems in achieving a specific language level, esp. vocabulary related, that enables them to translate text passages of Latin literature?

For the first question, some answers are offered – knowledge about language(s), literature, ancient history and European culture, transferable skills (e.g. meta-learning) and strategies (to deal with texts) – that all are parts of the humanistic view on education, i.e. Humboldt's concept of *Bildung*, and lacking an economically motivated purpose. Obviously, students can achieve these results only, if they acquire a sufficient level of language competence and performance in Latin (and German). Therefore, finding answers to the second question is more important.

Studying the (German) literature referring to the Latin language proficiency of students, one topic is constantly recurring: the difficulties students have with learning Latin vocabulary (e.g. Hermes, 1988; Kuhlmann & Horstmann, 2018; Steinthal, 1971; Stirnemann, 2009; Utz, 2000). After a subsequent analysis of the circumstances of Latin as a subject in German high

---

[1] DESTATIS – Statistisches Bundesamt, *Schüler/-innen mit fremdsprachlichem Unterricht.* Available: https://www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Bildung-Forschung-Kultur/Schulen/Tabellen/allgemeinbildende-beruflicheschulen-fremdsprachl-unterricht.html. Accessed on: Apr. 20, 2020.

schools, one might conclude that there are probably three major reasons for this on-going problem of vocabulary acquisition in Latin classes:

1. A domain-specific reason: Latin as a historical language focuses on translation whereas reading and, more importantly, oral communication do not matter in German secondary schools (Einheitliche Prüfungsanforderungen in der Abiturprüfung - Latein: EPA, 2005). Thus, both teachers and students are not used to Latin language production though it is necessary to achieve a high level of language proficiency (cf. Output Hypothesis, Swain, 1995). Additionally, for the teaching objective "translation", they all tend to think cross-lingually (Latin – German) rather than intra-lingually (Latin – Latin) at the word level although a good translator needs intra-lingual understanding at word, sentence, and text level (cf. different translation theories, Siever, 2015).
2. A scientific reason: There is a lack of research, empirical studies, and data on learning Latin, esp. in the German-speaking community (Beyer, 2019, p. 1). Furthermore, findings from linguistics and its subordinate fields such as computational linguistics and psycholinguistics are hardly taken into account in the curriculum of Latin (or Ancient Greek) philology at German universities. Hence, it is not surprising that insights, e.g. from language acquisition research or the theory of the mental lexicon or a corpus-based approach of vocabulary learning, are known just to a small minority of professionals. Consequently, learning materials offered by publishing houses include these insights just as little so that esp. vocabulary acquisition does not receive new, research-based input.
3. A methodical reason: Due to the aforementioned reasons, there is a lack of a broad understanding of vocabulary or vocabulary competence and the importance of explicit vocabulary learning in Latin pedagogy. Thus, vocabulary acquisition does not follow a systematic approach and still receives almost no attention in classes (Kuhlmann, 2016, p. 50), because it is perceived as tedious and time-consuming. Overall, vocabulary learning is restricted to the homework, reduced to learning translational equivalents and finally tested in a completely context-free manner, despite suggestions to the contrary (Kuhlmann, 2016, pp. 50–55).

To tackle these difficulties at least partly, it seems appropriate to work in an interdisciplinary research project of corpus linguistics, Latin pedagogy and computer science, the so-called CALLIDUS project[2]. Within this project, the main research question is whether the data-driven (language) learning (DDL) approach (Braun, 2007; Talai & Fotovatnia, 2012) proposed in second language acquisition (SLA) research can be applied to a historical language like Latin. Because of the missing domain-specific theories and data, the research began by designing a model of Latin vocabulary acquisition according to the theory of the lexical representation in the mental lexicon (form level, lemma level, conceptual level) (Levelt, Roelofs, & Meyer, 1999, p. 4). To achieve a better understanding of what students know about Latin vocabulary and strategies to infer the meaning of words, two intervention studies with intermediate learners were conducted (see 2.1 and 2.2). On this basis, the promising methodology of DDL was further adapted to Latin classes by developing a study on a new concept of introducing and understanding vocabulary with beginners (see 2.3). In addition, the results of the first studies were combined with the DDL approach to develop a mobile-friendly software for creating corpus-based vocabulary exercises (Boulton, 2017; Gilquin & Granger, 2010). The so-called Machina Callida provided the possibility of a fourth study on the outcome of a context-based vocabulary approach compared to learning of simple word equivalents (see 2.4).

## 2 Methodology

In the four intervention studies conducted in three secondary schools in Berlin so far, 11 groups with overall 283 students participated. To preserve the authentic educational context, the samples were not be randomised (group = class) and exhibit a self-selection bias, i.e. they were influenced by the choice of the teachers willing to collaborate. For the reason of privacy policy, the students got an individual ID, but were not asked for their gender, other known languages, reading habits or anything else. In general, all studies had a DDL approach and a wide variety of vocabulary tasks (e.g. word formation or word meaning) in common.

[2] https://www.projekte.hu-berlin.de/en/callidus-en/index.html and https://eadh.org/projects/callidus-vocabulary-acquisition-latin-using-corpus-based-methods.

However, they differed in the degree of given context, the length of the intervention and the Latin text corpus depending on the language proficiency level and age of the participants.

### 2.1 Study with Intermediate Learners Based on a Text by Cicero

The first study was aimed at acquiring data about students' knowledge of Latin vocabulary and about their skills while dealing with Latin words. It was conducted in summer 2018 with two groups (Table 1) that were taught by the same teacher. Before starting the intervention, she assessed the students of the test group as significantly weaker esp. referring to their overall cognitive skills.

Table 1 Participants in the study (Cicero)

| Group | Group size | Number of participants | Grade | Age | Learning Latin |
|---|---|---|---|---|---|
| Test group | 30 | 27 | 9th | 13-15 | 5th year |
| Control group | 28 | 19 | 9th | 13-15 | 5th year |

After the pre-test (45 min) both groups worked for 12 lessons (about 4 weeks) with the same Cicero text prepared for intermediate learners (extracts of the letter *Ad Quintum fratrem* I, 1), but only the test group received an intervention[3] that focuses on different vocabulary tasks: basic form, polysemy, word formation rules, fixed word pairs, collocations, strategies for word linking. The 172 selected lemmata of the intervention and the tests pertain to a core vocabulary of the 500 most frequent words presented in all current textbooks (Utz, 2008). The intervention was concluded by a post-test (45 min) identical to the pre-test. Both tests[4] are restricted to the above-mentioned lemmata and contain tasks in five categories (basic morphological knowledge, word cross-linking strategies, strategies to decode word meaning, phrases, and application of vocabulary knowledge) including language production tasks, e.g. creating Latin phrases by using words given in pools. Additionally, they provide the students with the possibility to rate the difficulty of both the task type and the task assignment.

[3] DOI of the text and intervention: 10.5281/zenodo.3751196.
[4] DOI of the pre-test and the post-test: 10.5281/zenodo.3752556.

## 2.2 Study with Intermediate Learners Based on Texts by Ovid

The second study focused on context-based vocabulary tasks and was carried out in winter 2018/19. Different teachers taught the selected groups (Table 2) of which the test group appeared to be a group of learners who had a lower learning outcome than the control group.

Table 2 Participants in the study (Ovid)

| Group | Group size | Number of participants | Grade | Age | Learning Latin |
|---|---|---|---|---|---|
| Test group | 30 | 21 | 10th | 14-16 | 6th year |
| Control group | 30 | 27 | 10th | 14-16 | 6th year |

After the pre-test (45 min) both groups worked for about 8 weeks with the same Ovid text prepared for intermediate learners (the metamorphosis *Pyramus et Thisbe*, IV, vv. 55-166), but only the test group received the text with an intervention[5] consisting of a theoretical introduction to vocabulary acquisition and various vocabulary tasks. These tasks offer context-based learning activities that refer alternately to metacognitive strategies or to the form, lemma or conceptual level of the lexical representation in the mental lexicon. Like the first study, this one relies on a small number of selected lemmata (165) matching the Ovid texts and the previously mentioned core vocabulary. Finally, the intervention ended with a post-test (45 min) identical to the pre-test in structure, but not in content, for the context-based questions called for choosing another metamorphosis[6]. Both the tasks of the intervention and of the tests[7] are restricted to the afore-said lemmata. On the one hand, the tasks explicitly address aspects of vocabulary acquisition that have so far mostly been an implicit part of Latin classes (e.g. to give the basic form of an inflected word), and on the other hand they explicitly demand metacognitive action from the students (e.g. to explain how the basic form can be derived). Additionally, each test also covers an unabridged metamorphosis including ten test items and providing the students with the opportunity to rate the difficulty of both tasks and task assignments. In order to be able to master this amount

[5] DOI of the text and the intervention: 10.5281/zenodo.3751581.
[6] Pre-test: Salmacis (Ovid, Metamorphoses IV, vv. 271-388),
Post-test: Pygmalion (Ovid, Metamorphoses X, vv. 243-297).
[7] DOI of the pre-test and the post-test: 10.5281/zenodo.3752624.

of text and test items in one lesson (45 min), both metamorphoses were supplemented with subheadings, introductory and transitional German texts and were mainly offered in a bilingual or paraphrased form. At the end of the tests, the understanding of the context is checked by a bonus task. In this task, the students have to answer the question about the nature and end of the respective metamorphosis while providing text references. The aim of the entire test design is to give the learners as much context as possible when completing the vocabulary tasks.

### 2.3 Study with Beginners Based on Textbook Material Relying on a DDL Concept

The third study was concerned with implementing a DDL concept for vocabulary acquisition into Latin classes and was conducted in the school year 2018/19 (Table 3). The same teacher taught both test groups while different teachers instructed the control groups.

Table 3 Participants in the study (textbook)

| Group | Group size | Number of participants | Grade | Age | Learning Latin |
|---|---|---|---|---|---|
| Test group 1 | 28 | 27 | 5th | 10-11 | 1st year |
| Control group 1 | 30 | 25 | 5th | 10-11 | 1st year |
| Test group 2 | 29 | 27 | 5th | 10-11 | 1st year |
| Control group 2 | 29 | 24 | 5th | 10-11 | 1st year |

Because the students did not know any Latin at the start of the study, it was necessary to design a German-based test that evaluates skills students need for translating (i.e. decoding a text) and learning Latin vocabulary (e.g. technical terms, word formation rules). Thus, the identical pre-test and post-test[8] contain a text on the Punic wars (readability: 7th grade), a reading comprehension test (8 tasks, 10 points) and a test of vocabulary knowledge (5 tasks, 20 points). By relying additionally on a Design-based Research approach (Bakker & van Eerde, 2015) for constructing the intervention it was always possible to react to the learning requirements and keep the design agile. In general, the intervention[9] was designed as a complement to the used textbook, but it changed the practice and order of teaching

[8] DOI of the pre-test and the post-test: 10.5281/zenodo.3752943 .
[9] DOI of the intervention: 10.5281/zenodo.3751690.

absolutely, e.g. it was now necessary to start each unit with the context-based introduction of new words. After transferring the theoretical concepts of DDL and vocabulary competence into learning materials, the intervention consists of the following parts for each unit:

- Context-based introduction of vocabulary in which the meaning of a new Latin word has to be inferred
- Worksheets on vocabulary knowledge and strategies including tasks
- Exercises focusing on Roman culture and practicing vocabulary in a broad sense
- Vocabulary test (so-called DDL test) with the same task formats and order for each unit

When three units are done, a traditional vocabulary test (list of word pairs) follows. In this test, the vocabulary of the last three units is divided into four tests, clustered in word classes and distributed randomly in a test group. In this way, it was possible to collect a vast amount of data on the learning progress of each student (pseudonymously).

### 2.4 Study with Intermediate Learners Based on the So-called Vocabulary Unit

The fourth study aims at investigating the impact of context-dependent, corpus-based vocabulary learning on students' vocabulary competence. It is carried out in the so-called vocabulary unit of the project software[10] and is not yet finished. Therefore, the study is not confined to selected students or students at all, but to participants who know Latin. In autumn 2019, the data of 13 participants was evaluated – since all tests were performed almost simultaneously, it is likely to be a group of students. To provide the same conditions for all participants, the study has to work independently of the preceding learning environment. Therefore, the usual study setting is enhanced with a text basis simulating a typical Latin class with its text work (pre-test – Latin text: introduction and comprehension tasks – intervention: exercise – post-test) and takes just 40 min (about one lesson). Within this study, participants should learn vocabulary that is part of the processed Latin text and examined in the post-test (= pre-test). Both tests contain five exercise types on different aspects of

[10] cf. https://korpling.org/mc/test. Documentation: https://korpling.org/mc/doc-voc-unit.

vocabulary knowledge, e.g. word formation or inflection. Besides, the actual intervention consists either of cloze exercises built from the processed text (corpus-based learning, i.e. DDL) or of a word list with free text fields including all the words of the processed text (learning of word pair associations) given randomly to the users. This way, it is possible to measure different learning outcomes in the post-test depending on the assigned type of exercise.[11]

## 3 Results

In the following section, some of the numerous results are presented. The selection is driven by the intention to give an overview of the diverse insights in Latin vocabulary acquisition gained during the course of the research project. Additionally, the manifold findings might allow a deeper understanding of the main problems in Latin classes when the students try to engage but fail more often than they succeed.

### 3.1 Study with Intermediate Learners Based on a Text by Cicero

The results of the first study[12] were interesting, but also disappointing because of the overall performance of the test group. Although this group had been assessed as cognitively less successful learners, the intervention should have shown at least a small improvement in dealing with the tasks of the tests. Yet, as it is shown in Figure 1, the test group did not accomplish a better score in the second test. While the control group improved their performance between the two tests by 16 %,[13] the test group

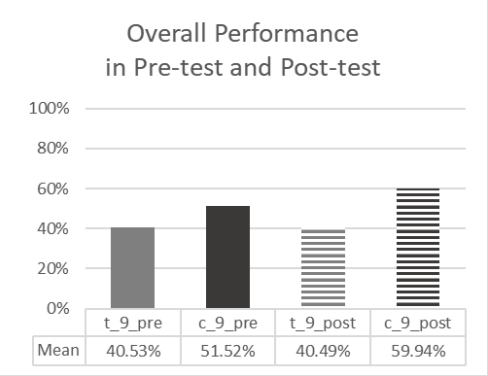



Figure 1 Relative averages (max.: 45 points)

Median: 38.89%; 51.11%; 41.11%; 60.00%
The control group in particular shows almost a normal distribution. Remarkable is also the big difference (approx. 20%) between the groups in the post-test.

[11] The constraints of this intervention are specified in Beyer and Schulz (2020, p. 1752).
[12] DOI of the test data: 10.5281/zenodo.3783475.
[13] To a degree, this might be attributed to the unusual tasks of the tests: After the pre-test, the students had to some extent experience with the tasks. Furthermore, the students themselves who assessed even the tasks in the post-test as easy that had been rated before as difficult, e.g. task 1B, support this assumption, too.

achieved almost the same score in pre-test (18.24 points) and post-test (18.22 points). Some of the circumstances might explain this, e.g. the pre-test in the control group took place after a strenuous class test (i.e. a weaker performance than usual), the study took place just before the summer break (i.e. a grading with marks was no longer possible) and the teacher used the material of the intervention only in every second lesson. Nevertheless, this does not fully account for the results. Besides the necessity that an intervention has to focus more on just one or two aspects of vocabulary competence if it is running out of time, it is obvious that a long-standing cognitive disparity between two classes cannot be overcome in a short intervention. This disparity is best shown in Figure 2 and Figure 3. Whereas task 1B refers particularly to the language proficiency level in Latin, because students need to recognise the lexical form of the given words, the other task (3B) relies more on general, higher-order thinking skills like inferring a word meaning and explaining one's own procedure. Obviously, the control group performs better in cognitively demanding tasks, but this ability does not help this group to surpass the other group in the Latin-specific task just as distinctly. This is interesting since it might suggest that at an early stage of second (foreign) language acquisition even well-developed strategies cannot compensate for a reduced representation of a word in the mental lexicon. That being said the results show above all that the Latin language level of both groups is very low even after learning Latin almost for five years.

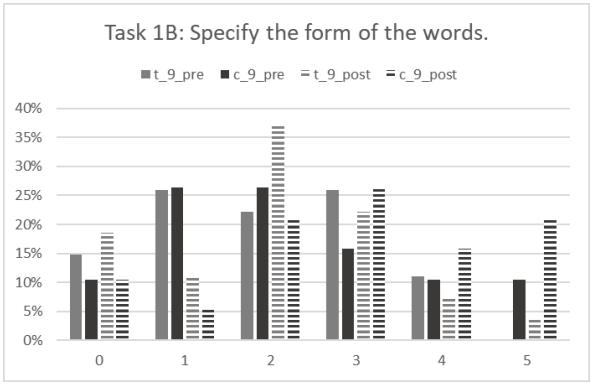



Figure 2 Frequency distribution of achieved scores (max. 5)

The students' majority of both groups think that this Latin-specific teaching task is difficult. In both groups the middle 50 % achieve 1 to 3 points out of 5 in the pre-test (post-test: t = 1-3 points, c = 2-4 points). The language proficiency does not seem to differ much.

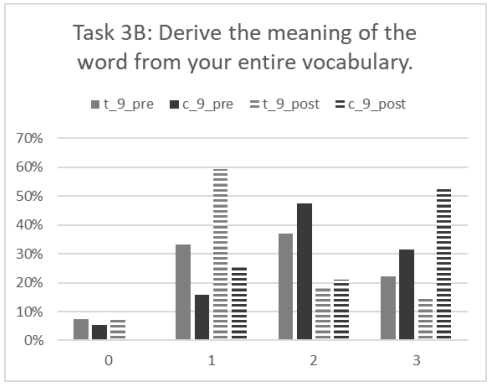



Figure 3 Frequency distribution of achieved scores (max. 3)

This task focuses on strategic language skills and does not require a certain Latin proficiency level. The control group performs much better, e.g. the median in the post-test is c = 3 (t = 1).

### 3.1.1 **Interesting Linguistic Errors**

Apart from the more or less disappointing quantitative results, a qualitative evaluation of the students' answers is productive in the sense of understanding the learning of a historical language like Latin better. For example, there is evidence that students are affected by interferences and overgeneralise when using language rules, as they would do when learning a modern language. Furthermore, they show many uncertainties concerning all levels of the lexical representation of a word in the mental lexicon. Some of the most interesting findings are given in Table 4.

Table 4 Examples of the qualitative analysis of students' answers in pre-test and post-test.

| **Task (3A, Pool): Create compounds and translate them.** | | | |
|---|---|---|---|
| Student's answer | Target hypothesis | English meaning | Type of error |
| *ab-scribere – abschreiben* | *abscribere* is non-existent in Latin; *ab-ducere* | to lead away | Word formation rule of L1 (*ab-schreiben* – to copy) causes wrong association. |
| *vide-tor – Seher* | *videtor* is non-existent in Latin; *scri[b/]p-tor* | writer | Word formation rule of L1 (*Seh-er* – the seer) and a weak understanding of the German concept of *Seher* cause wrong association. |
| *facil-iter – leichter* | *faciliter – leicht* (adverb) | easily | Neglecting the part of speech; phonological form of L1 (*leichter;* comparative) causes wrong association. |

| **Task (4B, 3 Pools): Construct a Latin phrase and translate it.** | | | |
|---|---|---|---|
| Student's answer | Target hypothesis | English meaning | Type of error |
| *litteras scribere habeo – Ich habe Briefe geschrieben.* | *litteras scribere possum – Ich kann einen Brief schreiben.* | I can write a letter. | Lexical form of L1 (*ich habe geschrieben* – I have written) causes wrong association, but L2 lexical form is built correctly (*habeo*) disregarding the participle. |
| *litteras scribere debet – Briefe müssen geschrieben werden.* | *litteras scribere debet – Er/Sie muss einen Brief schreiben.* | S/He has to write a letter. | Semantic-based understanding disregarding L2 form and a lack of awareness of active or passive voice in L1 cause wrong association. |
| *nos epistulam videt – Er sieht unseren Brief.* | *nos epistulam scribimus – Wir schreiben einen Brief.* | We write a letter. | Phonological resp. orthographic similarity to *noster* (*unser*) and a missing link to the part of speech of *nos* cause wrong association. |
| *amicitia magnam est – Die Freundschaft ist groß.* | *amicitia dignus est* – Er ist der Freundschaft würdig. | He is worthy of friendship. | Semantic-based understanding disregarding L2 form causes wrong association. |
| **Task (5A, Cloze): Give the appropriate German meaning.** | | | |
| Student's answer | Target hypothesis | English meaning | Type of error |
| *frater – Vater* | Bruder | brother | Phonological resp. orthographic form of L1 (*Vater*) causes wrong association. |

Though the intervention did not improve the performance of the test group quantitatively, the teacher's feedback was very positive. After working with the intervention material, she understood the necessity to spend time on vocabulary learning in class, but also suggested to integrate the tasks into the prepared text and shorten them as well to avoid boring students with long-lasting exercises.

### 3.2 Study with Intermediate Learners Based on Texts by Ovid

For avoiding the so-called ceiling effect (Sparrow, Newman, & Pfeiffer, 2005, p. 290), the tests were designed in a way that between 65% and 80% of the total score should have been achievable. On average, the students should reach 75-80% for the less complex tasks, e.g. 1 and 8, and only 65-70% for the linguistically and cognitively more demanding tasks, e.g. 2 and 9. Surprisingly, both groups did not exceed 60% on average and very few achieved more

than 70% (5 students/pre-test, 6 students/post-test).[14] In general, the test group achieved on average lower scores (Figure 4), but shows a growth in performance (mean: 10.83%; median: 12.33%). Contrarily, the control group performs worse in the second test (mean: -3.62%; median: -7.15%). Since the median shows slightly higher growth rates than the mean value, it can be concluded that the respective changes were actually triggered by the majority and less by individual outliers of both groups. Thus, it seems that the intervention helped the weaker test group to perform better in the second test so that the difference in outcomes decreases by more than 6% and the output of the groups is converging. Despite these encouraging results with regard to the possible effect of the intervention, the question remains why both groups did not reach the expected 75-80% on average. However, apart from unknown confounding factors, the less common test formats focusing on metacognition and implicit knowledge, the scope of the tests or a lack of motivation on the part of the students could be among the potential causes for these results. Nevertheless, the analysis of the tasks seems to be helpful to understand better the difficulties students have with the Latin vocabulary. Thus, four particularly interesting outcomes are presented in this context.

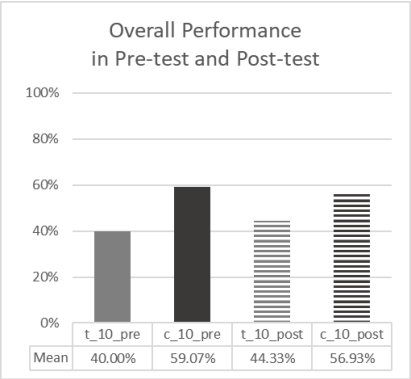



Figure 4 Relative averages of a maximum of 47 resp. 48 points

Median:
36.17%; 61.70%; 40.63%; 57.29%
After the intervention the test group performs in the post-test relatively better than the control group.

**Task 1 and Task 2: Metrics and Metacognitive Reflection**

After practicing metrics (hexameter) in the regular Latin class before this study started, both tasks should have been an easy introduction to the tests. Astonishingly, the results contradicted this assumption. Although the students rated the difficulty of both tasks as rather easy, they failed to achieve a high score on average. Despite the visible improvement between pre-test and post-test in task 1 (Figure 5), the students of both groups had problems with the second task that focuses on metacognitive understanding and German technical language to

[14] DOI of the test data: 10.5281/zenodo.3783480.

provide an answer. What is more, the generally high-performing control group obtained a remarkable growth rate (33%) in task 1, whereas the outcome of the test group changed by just 16%. That is notable, because it might suggest that both groups learned the metrics just by practicing in class, not by the tasks of the intervention. Consequently, it is recommendable to provide students with more explicit exercises for metrics repeatedly. Furthermore, the results in task 2 – in particular the overall failure of the control group (Figure 6) – call for more attention on metacognitive tasks within class so that the students have many opportunities to improve their metacognitive skills and, consequently, their overall language performance.

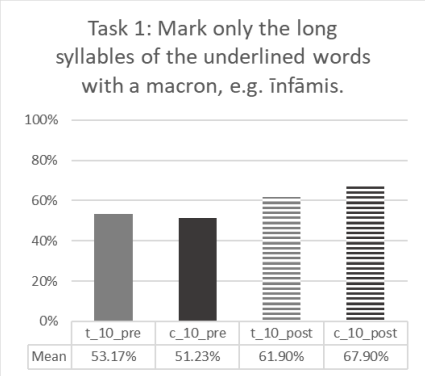



Figure 5 Relative averages of a maximum of 3 points

Median: 50.00%; 50.00%; 66.67%; 83.33%
Esp. the development of the median indicates that the majority of all students achieve higher scores in the post-test than in the pre-test.

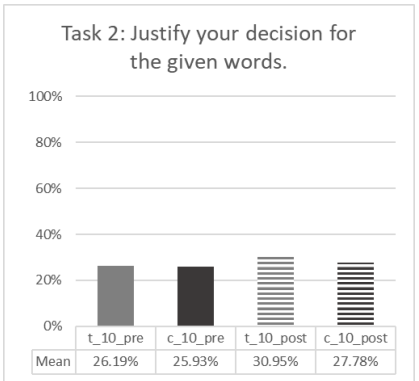



Figure 6 Relative averages of a maximum of 2 points

Median: 16.67%; 16.67%; 33.33%; 0.00%
The median is very striking. Whereas the majority of the test group performs better in the second test, more than half of the control achieve 0 points in the post-test.

### 3.2.1 **Task 8: Basic Form**

This task asks for a basic skill that is assumed to be present without being tested by the teacher. That this skill is often not sufficiently acquired, however, is also shown by the study of Florian (2015), where students do not recognize an inflected word as the word already learned in the basic form, i.e. the appropriate dictionary entry is not found. Therefore, the results (Figure 7) are not as shocking and unexpected as they could have been, even if on

average the students could give just one basic form out of three that were asked for. Thereby, the findings including the assessment of the task as rather difficult by the students indicate that most students fail to lemmatise a word correctly or to retrieve for a given word relevant information from their mental lexicon. This conjecture is supported by the replies in the second part of the task. Even though some students could give smart answers like "*cupidine* kommt von *cupidus*" (i.e. *cupidine* is derived from *cupidus*), most students who gave a basic form correctly gave no explanation, but the expression "gelernt" (i.e. learned). This can only lead to the conclusion that they were not able to use the context of the word or its morphology to deduce the basic form. Thus, the test results strengthen the perception that learners even after more than five years of learning Latin fail inevitably at a very low threshold level in comprehending Latin words, sentences and texts, because most have only reached the lowest stage of language acquisition (cf. the SLA model of Jiang, 2000).

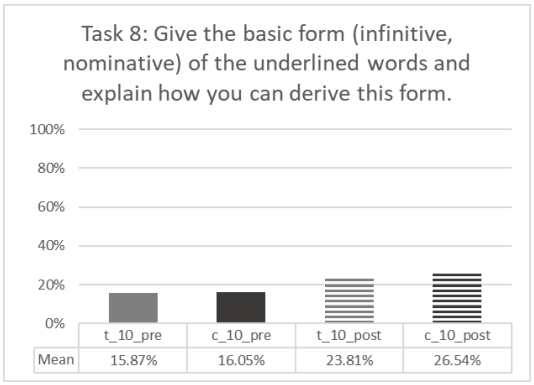



Figure 7 Relative averages of a maximum of 6 points

Median: 16.67%; 16.67%;16.67%;16.67%
The constant median shows that in general the small improvements in the second test are due to outliers who perform much better than at least half of all students.

### 3.2.2 **Task 9: Derivation of the Word Meaning**

To solve this task students have to rely on (metacognitive) strategies when retrieving lexical information from their mental lexicon. The underlined words (pre-test: *mixta*, *iunguntur*, *facies*; post-test: *admovet*, *flectitur*, *facies*) are neither part of the above-mentioned 165 lemmata of the learning vocabulary nor totally unknown. Although the students declare this task rather difficult, even the usually low-performing test group achieves 42% in the pre-test (c = 73%) and improves its results in the post-test a little (48%, growth rate: 15%) (Figure 8). This might indicate that the intervention helped the test group to perform better. However, what happened to the control group? The average score drops by almost 20% and is now only 54%. There might have been some confounding factors like a lack of working time (task 10 shows a similar decline). Nevertheless, another explanation is also possible, if one evaluates

the answers given for the Latin word *facies* that is asked for in both tests. In the pre-test, 24 students of the control group know the word meaning and derivation without almost any erroneous associations. However, in the post-test only 10 students of this group are able to recall the meaning and the derivation. This might indicate that the word *facies* was "known" by chance in the pre-test, because it had been used previously. Therefore, almost all students knew it (24 out of 27). Contrarily, in the post-test only these students could answer correctly who really had learned the word and could recall it from the mental lexicon.

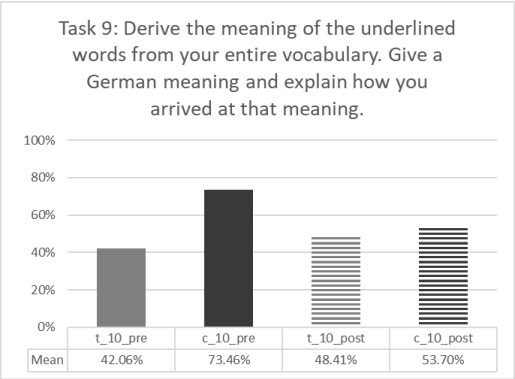



Figure 8 Relative averages of a maximum of 6 points

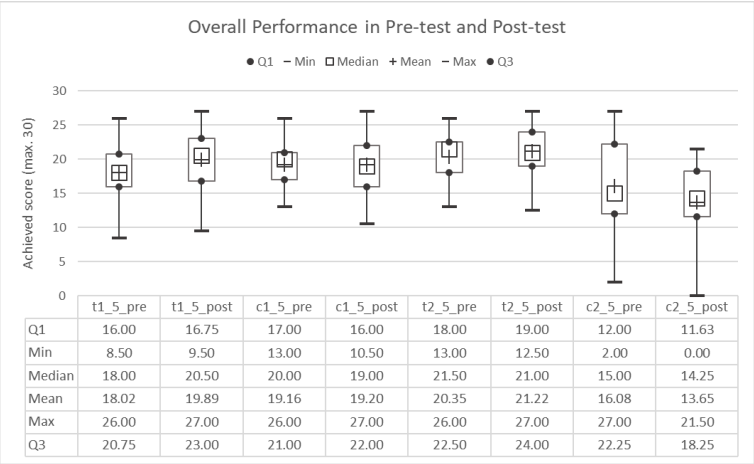
Median: 33.33%; 66.67%; 50.00%; 50.00%
The contrasting development brings both groups close together so that finally at least 50% of each group accomplish 50% (3 points) in the post-test.

### 3.3 Study with Beginners Based on Textbook Material Relying on a DDL Concept

In general, the study can just vaguely point the way ahead for vocabulary work and learning in Latin classes. Because of the unmanageable confounding factors in an authentic educational context and the long period it took, the correlation of test results and intervention is at most a weak one.[15] The reading comprehension test especially refers to an ability that is particularly strongly demanded and promoted during the first year at high school, since learning is increasingly carried out via texts. Thus, it is not entirely surprising that the overall performance of the studied groups reached a higher level in the post-test – with one exception (Figure 9). The results of the second control group (c2) cannot be included in the evaluation since the group obviously did not cooperate in the post-test.[16]

[15] DOI of the test data: 10.5281/zenodo.3816709.

[16] Especially in the vocabulary knowledge test, the students commented the test itself with rather annoyed remarks like *kein Bock* (I'm not up for it), *voll blöd* (bloody stupid), *weil Banane krumm* (German nonsensical expression for bloody stupid), *das is voll sinnlos* (that's utterly pointless). Many students also decorated the materials with scribbles.

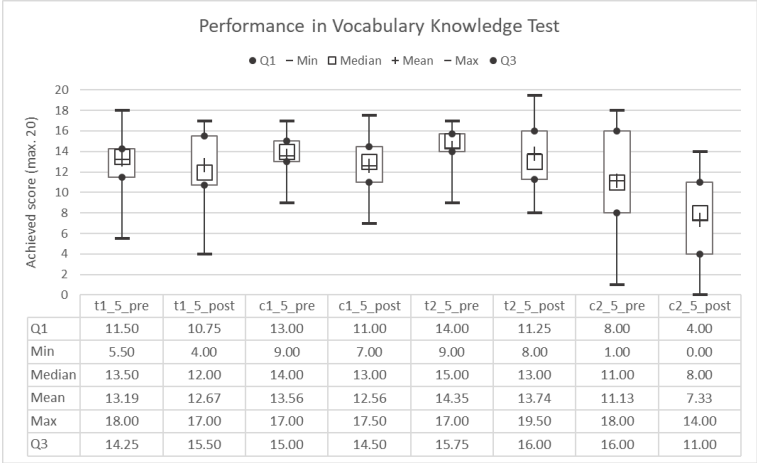



| | t1_5_pre | t1_5_post | c1_5_pre | c1_5_post | t2_5_pre | t2_5_post | c2_5_pre | c2_5_post |
|---|---|---|---|---|---|---|---|---|
| Q1 | 16.00 | 16.75 | 17.00 | 16.00 | 18.00 | 19.00 | 12.00 | 11.63 |
| Min | 8.50 | 9.50 | 13.00 | 10.50 | 13.00 | 12.50 | 2.00 | 0.00 |
| Median | 18.00 | 20.50 | 20.00 | 19.00 | 21.50 | 21.00 | 15.00 | 14.25 |
| Mean | 18.02 | 19.89 | 19.16 | 19.20 | 20.35 | 21.22 | 16.08 | 13.65 |
| Max | 26.00 | 27.00 | 26.00 | 27.00 | 26.00 | 27.00 | 27.00 | 21.50 |
| Q3 | 20.75 | 23.00 | 21.00 | 22.00 | 22.50 | 24.00 | 22.25 | 18.25 |

Figure 9 Distribution of achieved scores in general

Mean growth rate: t1= 10.38%; c1 = 0.21%; t2 = 4.28%; c2 = -15.16%
Median growth rate: t1 = 13.89%; c1 = -5.00%; t2 = -2.33%; c2 = -5.00%
For the groups t1, c1 and t2, the interquartile range increases. That might indicate, that the students enhanced their skills in the first high school year answering the questions now more consciously by relying on learned facts and methods.

However, as Figure 10 shows, the general improvement is not due to the knowledge about vocabulary, but to the increased scores in the reading comprehension test (Figure 11). In a way, the higher scores of the reading comprehension were anticipated due to the increased practice of reading, but there is one striking feature: On average, both test groups achieved a higher score with a median of 75% (c1 = 65%, c2 = 63%). Furthermore, there are no longer extreme outliers in both test groups. Considering the explicit training on context awareness the groups received

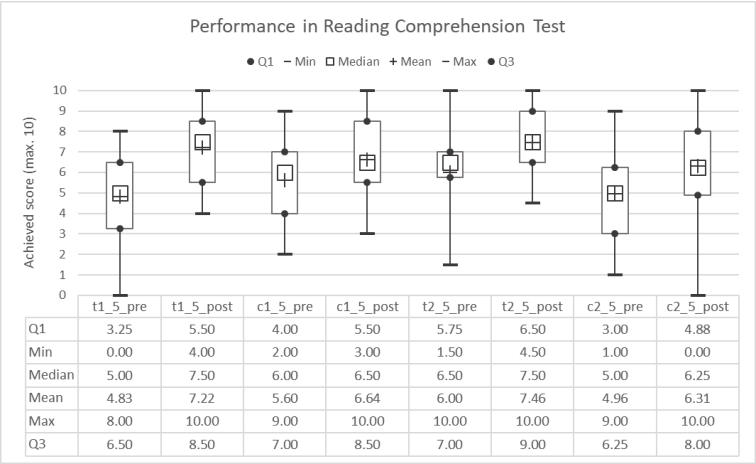



| | t1_5_pre | t1_5_post | c1_5_pre | c1_5_post | t2_5_pre | t2_5_post | c2_5_pre | c2_5_post |
|---|---|---|---|---|---|---|---|---|
| Q1 | 11.50 | 10.75 | 13.00 | 11.00 | 14.00 | 11.25 | 8.00 | 4.00 |
| Min | 5.50 | 4.00 | 9.00 | 7.00 | 9.00 | 8.00 | 1.00 | 0.00 |
| Median | 13.50 | 12.00 | 14.00 | 13.00 | 15.00 | 13.00 | 11.00 | 8.00 |
| Mean | 13.19 | 12.67 | 13.56 | 12.56 | 14.35 | 13.74 | 11.13 | 7.33 |
| Max | 18.00 | 17.00 | 17.00 | 17.50 | 17.00 | 19.50 | 18.00 | 14.00 |
| Q3 | 14.25 | 15.50 | 15.00 | 14.50 | 15.75 | 16.00 | 16.00 | 11.00 |

Figure 10 Distribution of achieved scores in the vocabulary knowledge test

Every group achieves a lower score on average, as the growth rates show:
Mean growth rate: t1 = -3.93%; c1 = -7.37%; t2 = -4.26%; c2 = -34.08%
Median growth rate: t1 = -11.11%; c1 = -7.14%; t2 = -13.33%; c2 = -27.27%
Similarly, to the overall performance, the interquartile range increases considerably for t1, c1, and t2.

in the intervention, this might suggest that more students in these groups were able to decode information by noticing the context.




| | t1_5_pre | t1_5_post | c1_5_pre | c1_5_post | t2_5_pre | t2_5_post | c2_5_pre | c2_5_post |
|---|---|---|---|---|---|---|---|---|
| Q1 | 3.25 | 5.50 | 4.00 | 5.50 | 5.75 | 6.50 | 3.00 | 4.88 |
| Min | 0.00 | 4.00 | 2.00 | 3.00 | 1.50 | 4.50 | 1.00 | 0.00 |
| Median | 5.00 | 7.50 | 6.00 | 6.50 | 6.50 | 7.50 | 5.00 | 6.25 |
| Mean | 4.83 | 7.22 | 5.60 | 6.64 | 6.00 | 7.46 | 4.96 | 6.31 |
| Max | 8.00 | 10.00 | 9.00 | 10.00 | 10.00 | 10.00 | 9.00 | 10.00 |
| Q3 | 6.50 | 8.50 | 7.00 | 8.50 | 7.00 | 9.00 | 6.25 | 8.00 |

Figure 11 Distribution of achieved scores in the reading comprehension test

Every group achieves a higher score on average, as the growth rates show:
Mean growth rate: t1 = 49.43%; c1 = 18.57%; t2 = 24.38%; c2 = 27.31%
Median growth rate: t1 = 50.00%; c1 = 8.33%; t2 = 15.38%; c2 = 25.00%

In contrast, there is no good explanation for the failure in the vocabulary knowledge test, in particular referring to the test groups. After focusing in the intervention on word fields, word formation, and technical language (among other things), at least the students of the test groups should have increased their scores. Instead, measured by the median, they performed as groups even worse than control group 1. In general, the variation grows in all three considered groups. That might indicate that the students became more heterogeneous with regard to their metacognitive abilities during the first high school year.

### 3.3.1 **The Overall Performance of the Test Groups in Latin Classes**

Comparing all results related to the given grades[17] during the school year three facts are obvious (Figure 12, Figure 13). First, the second test group gets better grades on average in the newly introduced DDL tests, the traditionally structured vocabulary (list) tests and the exams. Second, the performance of this group as measured by the received marks declines significantly in the second half of the year. Third, the students of the second test group appear to have more difficulties with the DDL tests compared to their success in learning word equivalents (vocabulary tests). In contrast, test group 1 performs in a more constant way. In particular, it is interesting that the variance measured for both vocabulary test types is almost

[17] In Germany, the grades are ranked from 1 (highest) to 6 (lowest). Sometimes it is possible to get a $1^{+}$, i.e. a student has achieved more than 100% (in the figures below 1.0).

the same (DDL: 0.56; voc: 0.57). Thus, it is likely that this group was able to adapt better to the new vocabulary test type over time as is also suggested by the comparison of the overall achievement in the DDL tests of both groups (Figure 14, Figure 15).

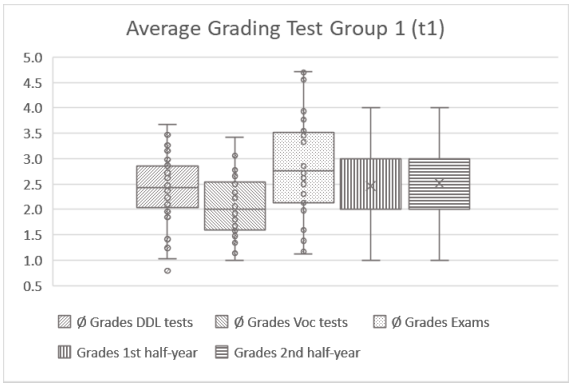



Figure 12 Distribution of achieved grades (t1)

Mean: DDL = 2.4; Voc = 2.1; Exams = 2.8;
1st half = 2.5; 2nd half = 2.5
Median: DDL = 2.4; Voc = 2.0; Exams = 2.8;
1st half = 3.0; 2nd half = 3.0

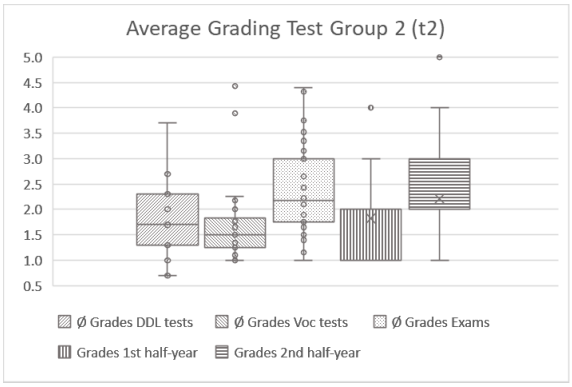



Figure 13 Distribution of achieved grades (t2)

Mean: DDL = 1.7; Voc = 1.7; Exams = 2.4;
1st half = 1.8; 2nd half = 2.2
Median: DDL = 1.7; Voc = 1.5; Exams = 2.2;
1st half = 2.0; 2nd half = 2.0

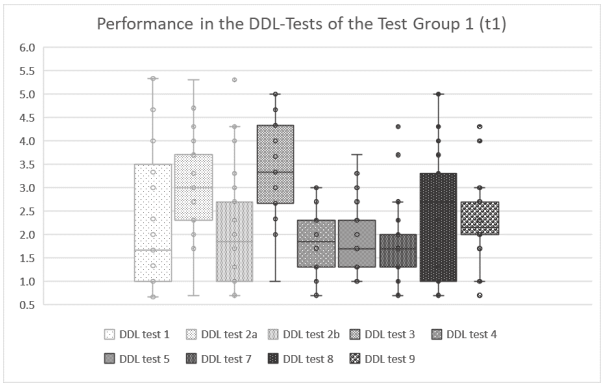



Figure 14 Distribution of grades per each DDL test

Mean: 1 = 2.3; 2a = 3.1; 2b= 2.1; 3 = 3.4; 4 = 1.8;
5 = 1.9; 7 = 1.8; 8 = 2.5; 9 = 2.4
Median: 1 = 1.7; 2a = 3.0; 2b = 1.9; 3 = 3.3; 4 = 1.9;
5 = 1.7; 7 = 1.7; 8 = 2.7; 9 = 2.4
Test 8 contains an usual form of task 1 that the students could not solve: Give a synonym to … e.g. *dat*.
(Test 6 is missing because there is no comparative value.)

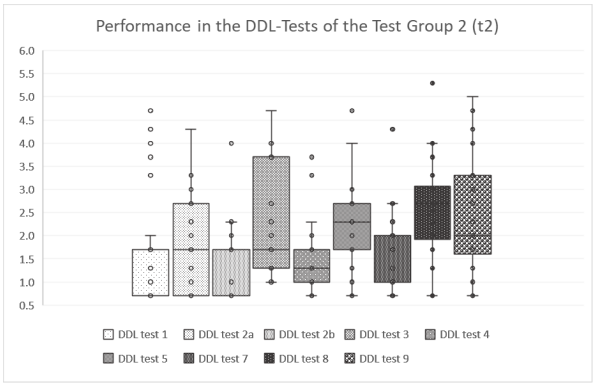



Figure 15 Distribution of grades per each DDL test

Mean: 1 = 1.5; 2a = 1.8; 2b= 1.2; 3 = 2.3; 4 = 1.5;
5 = 2.2; 7 = 1.5; 8 = 2.5; 9 = 2.3
Median: 1 = 0.7; 2a = 1.7; 2b = 0.7; 3 = 1.7; 4 = 1.3;
5 = 2.3; 7 = 1.0; 8 = 2.7; 9 = 2.0
Test 8 contains an usual form of task 1 that the students could not solve: Give a synonym to … e.g. *dat*.
(Test 6 was cancelled due to reasons at school.)

Starting with test 4, the results of the first group (t1) are approaching those of the other group in the second half-year, which might be attributed to a learning effect or a better motivation. However, although the reason for the improvement is uncertain, it is obvious that this new kind of vocabulary test is manageable even though it possibly takes time for the students to adjust to the new tasks.

### 3.3.2 The Latin Cloze Task in the DDL Tests

Usually, in a historic language like Latin there is no need for language production though it would be advisable for better learning outcomes (Izumi, 2003). Therefore, students do not have to fill in Latin cloze tests, as a rule. Thus, the results in this test format are very interesting, since they might at least show whether the majority of (young) learners can handle a cognitively less demanding version of context-based language production: a Latin cloze with answers in a pool. Figure 16 and Figure 17 display for the cloze tasks of three tests[18] the relative results for each task. In general, students of the first test group achieve a higher percentage and almost 80% of these students reach at least two third of the possible scores. Thus, the group (t1) that performs at the beginning of the intervention worse adapts successfully to the cloze task. Comparing these results to the average grades in the traditional vocabulary tests, it is informative that the second test group stays ahead of the other except for the first test[19]. This might lead to the conclusion that both tasks (cloze vs. word list) are not correlated in the way that it would be possible to predict the outcome of a context-sensitive reading or translation task by evaluating the results of a test on translational equations. Although this does not seem surprising because of the involved procedures in retrieving answers from the mental lexicon, it is astonishing that this traditional vocabulary test type prevails. It does not seem to be very helpful for preparing complex tasks like

[18] These tests are the only ones that can be compared due to a lack of data, since other individual results are missing.
[19] Median: t1/test 1 = 1.0; t2/test 1 = 1.0; t1/test 2 = 1.7; t2/test 2 = 1.0; t1/test 3 = 2.7; t2/test 3 = 1.9; t1/test 4 = 1.7; t2/test 4 = 1.3

decoding a Latin text nor does it enhance the understanding of the lexical form in the second language Latin.

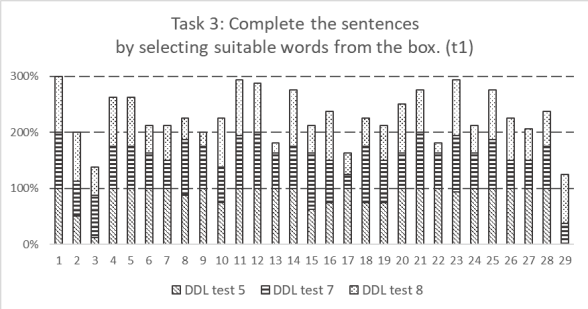



Figure 16 Relative results per DDL test, task 3

(100% each as maximum, numbers = students) (t1)
Mean: test 5 = 89.51%; test 7 = 73.49%; test 8 = 67.89%
Median: test 5 = 100.00%; test 7 = 75.00%; test 8 = 75.00%

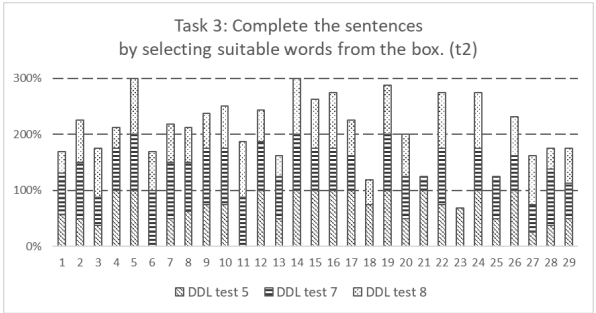



Figure 17 Relative results per DDL test, task 3

(100% each as maximum, numbers = students) (t2)
Mean: test 5 = 68.53%; test 7 = 75.00%; test 8 = 64.87%
Median: test 5 = 75.00%; test 7 = 75.00%; test 8 = 68.75%

### 3.4 Study with Intermediate Learners Based on the So-called Vocabulary Unit

The first results[20] of the on-going study indicate that participants who completed their tests very fast seem to reveal a higher level of vocabulary competence. Furthermore, these learners generally also made less mistakes, which implies that advanced learners do not just produce better test results, but also need less time to do so (Beyer & Schulz, 2020, pp. 1753–1754). Apart from these less research-specific findings, there are signs of support for the DDL approach. Figure 18 indicates that most participants of the intervention show no development at all, which suggests that they either a) already achieved the highest score or b) failed completely in both the pre-test and the post-test. Typically, the former is the case, esp. in a highly advanced learner group as the study group. While this case may be of little interest since there is hardly any opportunity for improvement except for presenting harder exercises, the cases that deviated in the post-test are much more remarkable: If there was a deviation, participants who learned with a vocabulary list generally changed for the worse, while those with a cloze usually improved their score. Even though it cannot be disputed that there are a

[20] DOI of the test data: 10.5281/zenodo.3601182.

few exceptions where cloze learners' performance decreased slightly, there is on the other hand only one example of a student learning with a vocabulary list who scores a higher result for a specific exercise in the post-test. Thus, in general, learning vocabulary in context seems to be correlated with better performance on the given vocabulary test.

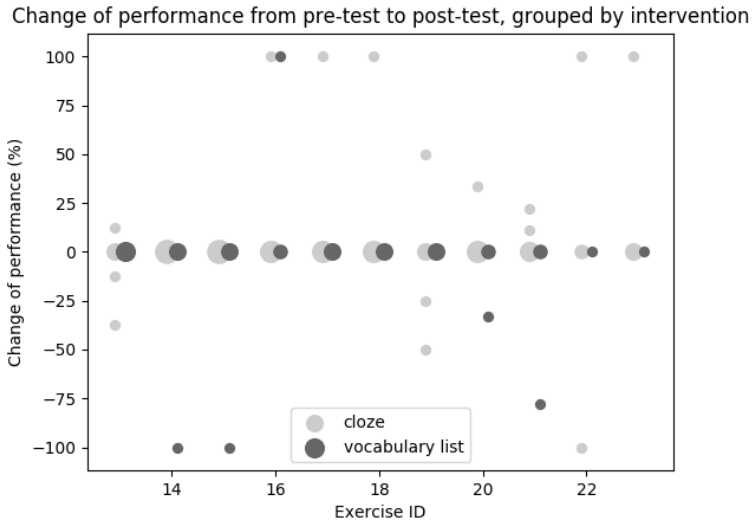



Figure 18 Comparison of students' performance before and after certain interventions.

Larger markers indicate multiple students with the same percental change.
Exercise IDs correspond to those in the learner dataset

## 4 Conclusions

Overall, the studies allow some general remarks on testing Latin vocabulary in high school classes. First, the students' assessment of the task statements and difficulty proved to be helpful for the evaluation because it mostly matched the individual performance. Therefore, this kind of assessment should be included in tests and perhaps even in exams. Second, for a more precise evaluation it is imperative to split each task into the amount of operators and to assign the scores correspondingly, e.g. "Give the basic form.(1)" and "Explain how you derive it.(1)". Third, the metacognition skills also required for language acquisition are for most untrained students only available to a limited extent and are obviously further limited by a low level of language competence in German (technical) language, e.g. to reason why a

syllable is long or to understand a task statement. Fourth, the feedback of the teachers suggests that students generally appreciate a vocabulary-focused classroom design, if it is diverse and helpful for solving the tasks in Latin classes, since they enjoy working with words. For this reason alone, vocabulary learning should take place in Latin classes regularly and variedly.

## 4.1 Language Proficiency

As it was shown above, the causes of errors are similar to the ones when learning a modern language: Students tend to overgeneralise and to make inferences from L1 in sound, form and syntax (cf. 3.1.1). Furthermore, the automation of retrieval processes from the mental lexicon is rarely successful, e.g. the recall of the basic form for a given word (cf. 3.2.1). What is more, even though they have learned just a small amount of vocabulary (approx. 240 words), they have major difficulties to connect a Latin word with another Latin one, although both words were recently learned and discussed as synonyms, e.g. in the task "Give a synonym to … e.g. *dat*" (cf. 3.3.1, test 8). All these examples illustrate a low outcome of Latin language acquisition, in particular, if students have learned Latin for more than four years. Accordingly, their overall proficiency level reaches just a formal stage of lexical development (Jiang, 2000, p. 51) that is not sufficient for handling words in contexts or understanding texts. Consequently, this seems to lead to two major conclusions: Firstly, there has to be more diverse and regular vocabulary practice in class that enables to achieve the second and third stages of language acquisition, i.e. L1 lemma mediation and L2 integration. However, since there is no need for language production and the occurrence frequencies of each word are low due to the different literary genres, the second and third stage might be difficult to reach in Latin classes. Secondly, if this assumption is correct, then it might be time to rethink the goals of learning Latin in high school, e.g. instead of focusing on Latin-specific skills needed for translation it might be wise to focus more on overall language skills.

## 4.2 Methods for Learning and Evaluating Vocabulary

Due to the varied Latin literary texts the DDL approach has its limits, because students know less words in each text than is necessary for incidental learning (95-98% should be known,

Nation, 2013, p. 352). Thus, the DDL approach in these studies refers to intentional learning by using context-based vocabulary exercises and introducing a broad understanding of vocabulary knowledge. However, this (corpus-based) language acquisition theory cannot be implemented by a mere contextualization of vocabulary tasks, but must be made explicit; otherwise, students do not learn to use the context for finding answers (cf. the bonus task in the Ovid study, 2.2). Additionally, the results of the last two studies indicate that even young learners have no long-lasting problems with context-based exercises (cf. 3.3.2), adapt quite fast to these more complex tasks (cf. 3.3.1), and perform better in context-based classroom activities like decoding a text after learning vocabulary in a context (cf. 3.4). Thus, the DDL approach seems to be a promising way to design Latin vocabulary learning more attractively and to prepare the students more successfully for decoding Latin texts.

### 4.3 Consequences

In terms of the mental lexicon, Latin language acquisition is not different to learning a modern language (e.g. storage of information), but the words are less used and scarcely useful. Thus, students must feel the need to know a word (e.g. by comprehension tasks), otherwise they learn the words just for tests and forget them instantaneously. In addition, they need a lot of diverse experience in dealing with vocabulary, e.g. learning rules of word formation, explaining their own actions or inferring a word meaning, to reach a higher level of language proficiency. For that reason, it is mandatory to work on vocabulary in class. According to this intensified and diverse vocabulary work, the test formats should resemble the underlying broad understanding of vocabulary knowledge and competence. Therefore, some kind of the so-called DDL tests should supplement the traditional vocabulary (list) tests. Consequently, if vocabulary is emphasised in class, the exams should equally shift their focus to more vocabulary tasks. Finally, to further the improvement of teaching Latin and to increase the outcome of learning Latin vocabulary, it is generally recommendable to conduct more finely granulated studies for modelling language acquisition in a historical language like Latin.

## 5 References


Bakker, A., & van Eerde, D. (2015). An Introduction to Design-based Research with an Example from Statistics Education. In A. Bikner-Ahsbahs, C. Knipping, & N. C. Presmeg (Eds.), *Advances in mathematics education. Approaches to qualitative research in mathematics education: Examples of methodology and methods* (pp. 429–466).

Behrendt, A., & Korn, M. (2016). Schülerzahlen im Fach Latein und Entwicklungsperspektiven der Fachdidaktik. *Forum Classicum*, *59*(3), 156–157. https://doi.org/10.11588/fc.2016.3.36322

Beyer, A. (2019). Im Lateinunterricht: cupidine kommt von cupidi und ne ist Fragepartikel - Wortschatzprobleme und ihre Ursachen. *Pegasus-Onlinezeitschrift*, *19*, 1–19.

Beyer, A., Kipf, S., Liebsch, A.-C., & Zimmermann, S. (2019). Zwischen Aktualität und historischer Forschung: Entwicklungstendenzen in der Fachdidaktik Latein. *Forum Classicum*, *62*(2), 85–96.

Beyer, A., & Schulz, K. (2020). Using NLP to Create Corpus-based Vocabulary Exercises in Latin Classes. In L. Gómez Chova, A. López Martínez, & I. Candel Torres (Eds.), *INTED Proceedings, INTED2020 Proceedings* (pp. 1750–1757). IATED. https://doi.org/10.21125/inted.2020.0562

Boulton, A. (2017). Data-Driven Learning and Language Pedagogy. In S. L. Thorne & S. May (Eds.), *Encyclopedia of Language and Education, 3rd ed. Language, Education and Technology* (3rd ed., pp. 181–192). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-02237-6

Braun, S. (2007). Integrating Corpus Work into Secondary Education: From Data-driven Learning to Needs-driven Corpora. *ReCALL*, *19*(3), 307–328.

Florian, L. (2015). *Heimliche Strategien: Wie übersetzen Schülerinnen und Schüler? Dissertation*. Göttingen: V&R unipress.

Gilquin, G., & Granger, S. (2010). How Can Data-driven Learning Be Used in Language Teaching. In A. O'Keeffe & M. McCarthy (Eds.), *The Routledge handbook of corpus linguistics* (Vol. 359370, pp. 359–370).

Hensel, A. (2017). Überprüfen von Textverständnis - neue Wege der schriftlichen Leistungsfeststellung im altsprachlichen Unterricht. *Der Altsprachliche Unterricht*, *60*(4+5), 2–11.

Hermes, E. (1988). Wortschatz und Grammatikwiederholung bei der Texterschließung. *Der Altsprachliche Unterricht*, *31*(6), 55–78.

Izumi, S. (2003). Comprehension and Production Processes in Second Language Learning: In Search of the Psycholinguistic Rationale of the Output Hypothesis. *Applied Linguistics*, *24*(2), 168–196. https://doi.org/10.1093/applin/24.2.168

Jiang, N. (2000). Lexical Representation and Development in a Second Language. *Applied Linguistics*, *21*(1), 47–77.

Kipf, S., & Kuhlmann, P. (Eds.) (2015). *Perspektiven für den Lateinunterricht*. Bamberg: Buchner.

Korn, M., & Kuhlmann, P. (2017). Latein - nur gymnasial?: Herausforderungen und Lösungsansätze. In P. Kuhlmann (Ed.), *Studienbücher Latein. Perspektiven für den Lateinunterricht II* (pp. 70–79). Bamberg: Buchner.

Kuhlmann, P. (2016). Wortschatzlernen im Lateinunterricht - Mythen und Fakten. *Forum Schule*, *63*, 40–56.

Kuhlmann, P., & Horstmann, H. (2018). *Wortschatz und Grammatik üben: Didaktische Kriterien und Praxisbeispiele für den Lateinunterricht*. Göttingen: Vandenhoeck & Ruprecht.

Einheitliche Prüfungsanforderungen in der Abiturprüfung - Latein: EPA (2005).

Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, *22*, 1–75.

Nation, P. (2013). *Learning vocabulary in another language* (2nd ed.). Cambridge: Cambridge University Press.

Siever, H. (2015). *Übersetzungswissenschaft*. Tübingen: Narr Francke Attempo.

Sparrow, S. S., Newman, T. M., & Pfeiffer, S. I. (2005). Assessment of Children Who Are Gifted with the WISC-IV. In A. Prifitera, D. H. Saklofske, & L. G. Weiss (Eds.), *Wisc-IV clinical use and interpretation: Scientist-practitioner perspectives* (pp. 281–298). Amsterdam, Boston: Elsevier Academic Press. https://doi.org/10.1016/B978-012564931-5/50009-8

Steinthal, H. (1971). Zum Aufbau des Wortschatzes im Lateinunterricht. *Der Altsprachliche Unterricht*, *14*(2), 20–50.

Stirnemann, S. (2009). Zu scharfe Ränder. Vom Umgang mit Wortschatz und Wörterbuch. *Der Altsprachliche Unterricht*, *52*(6), 38–44.

Swain, M. (1995). Three Functions of Output in Second Language Learning. In G. Cook & B. Seidlhofer (Eds.), *Principle & Practice in Applied Linguistics: Studies in honour of H. G. Widdowson* (pp. 125–144). Oxford: Oxford University Press.

Talai, T., & Fotovatnia, Z. (2012). Data-driven Learning: A Student-centered Technique for Language Learning. *Theory and Practice in Language Studies*, *2*(7), 1526–1531.

Utz, C. (2000). Mutter Latein und unsere Schüler - Überlegungen zu Umfang und Aufbau des Wortschatzes. In P. Neukam (Ed.), *Dialog - Klassische Sprachen und Literaturen. Antike Literatur - Mensch, Sprache, Welt* (Vol. 34, pp. 146–172). München.

Utz, C. (2008). *Adeo: Das lateinische Basisvokabular* (1st ed.). *Bamberger Wortschatz*. Bamberg: Buchner.