universität wien

MASTERARBEIT / MASTER'S THESIS

Titel der Masterarbeit / Title of the Master's Thesis

„A multi-factor correlation study between student English vocabulary knowledge and English Medium Education in Multilingual University Setting related variables"

verfasst von / submitted by

Jennie Josephine Atwell

angestrebter akademischer Grad / in partial fulfillment of the requirements for the degree of

Master of Arts (MA)

Wien, 2023 / Vienna 2023

Studienkennzahl lt. Studienblatt / degree programme code as it appears on the student record sheet:	UA 066 812
Studienrichtung lt. Studienblatt / degree programme as it appears on the student record sheet:	Masterstudium English Language and Linguistics UG2002
Betreut von / Supervisor:	Univ.-Prof. Mag. Dr. Ute Smit

Acknowledgements

I would like to start out by thanking my supervisor, Prof. Dr. Ute Smit, for her endless support and feedback throughout my thesis. Her compassion, patience, and flexibility were paramount to my success and a large part of the reason why this thesis now exists in its entirety. Without her guidance and understanding nature I simply would have never been able to complete my project.

I would also like to thank my family, who supported me endlessly from across an ocean. To my mom, for only ever being a phone call away and for pushing me to finish my degree. She knows what it is like to be one paper away from a diploma and thus encouraged me to *just get it done*. To my dad, for giving me no other option but to love numbers from a small age, sitting at the kitchen table, working on extra math problems so I wasn't just "another girl bad at math". More recently, thank you dad for guiding me to combine my love for languages and knack for math. To my brother, Jake, for always being the example and encouraging me to study abroad and obtain a foreign degree. Sorry, unfortunately you cannot talk about me anymore in German. I know what you are saying. To my Aunt KJ, for all the love, support, and pop tarts over the years but most recently for editing my thesis and offering a critical eye. To Bri, my cousin, my best friend, and my shoulder to cry on through happy and sad, your support means the world to me.

I want to thank my boyfriend, Philip, for always being there for me as my sounding board, best critic, and programmer extraordinaire. In addition, thank you for taking on all the mundane domestic duties that have made all the difference in the last few months, from making dinner to cleaning dishes and taking out the trash. Thank you for pushing me to accomplish my goals and for being there for me through it all. When you shine, I shine. When I shine, you shine.

I would like to thank my close friends in Vienna, mostly Poppy, Maddi, Muhammed, and Fiona. Thank you, Poppy and Maddi, for always getting me out of the house and for distracting me when I really couldn't think about my thesis anymore. Thank you to my study friends, Fiona and Muhammed, for reminding me to finish my thesis and redirecting my distractions when I really couldn't think about my thesis anymore.

I would also like to thank Dr. Katharina Ghamarian-Krenn for letting me use her vocabulary test in my research and data collection as well as showing me one way to combine statistics and linguistics.

Abstract

The number of students who use English during their higher education in some shape or form is increasing worldwide. Additionally, English holds an undeniable role in the academic sphere as the main language for publishing and communicating scientific findings. Currently, a key debate centers around the potential impacts additional English exposure can have on the language proficiencies of students.

In the study that follows, Austrian master's students' vocabulary scores on two English vocabulary tests were compared to determine if certain factors suggested a statistically significant difference between groups of students. The first vocabulary test that was used was Schmitt, Schmitt, and Clapham's (2001) well-established Vocabulary Levels Test (VLT) that served a measure of general vocabulary knowledge. The second test was Ghamarian-Krenn's (2023) newly developed Form Recognition Test of Academic Verbs (FoRTAV) that focused on academic English. Several variables to group the participants and measure English-Medium exposure were developed for the present study: the academic disciplinary grouping of the students according to two separate sorting schemes; the amount of English exposure in the master's programs; students' perceived level in English; age; semester progress; study abroad experience; and commonly used languages inside and outside of the university setting.

The variables were developed with the aim of answering four main research questions. First, what vocabulary knowledge do master's students enrolled in Austrian universities possess at the different frequency levels and academic genre according to the VLT and FoRTAV results? Second, does vocabulary knowledge vary when students are grouped in academic discipline according to Biglan or Holland's schemes? Third, does English exposure in a student's current degree program or other biographical factors correlate to their vocabulary scores, and if so in what ways? Finally, how does student performance on the academic section (AWL) of the VLT compare to their performance on the FoRTAV?

With a sample size of 85 master's students attending Austrian universities, the students' reported levels of English showed a statistically significant positive medium strength correlation to the VLT results and the combined VLT & FoRTAV scores. Age had a statistically significant medium positive correlation to the 10k level score and the scores on the AWL section of the VLT and FoRTAV were found to have a statistically significant strong positive correlation. The academic disciplinary grouping schemes did not show any correlation to the vocabulary test scores. On average, the participants performed well, with median scores on the VLT sections as 2k = 30/30, 3k = 29/30, 5k = 29/30, 10k = 23/30, AWL = 30/30, and the median score on the FoRTAV = 61/72. With a passing minimum requirement set at 86% correctly answered items, 95.3% of participants passed the 2k section, 90.6% of participants passed the 3k section, 76.5% of participants passed the 5k section, 41.2% of participants passed the 10k section, and 96.5% of participants passed the AWL section. The student results from the AWL section of the VLT correlated strongly and positively with the FoRTAV results. Further discussion of the potential implications of these results and caveats to these results are also discussed.

Table of Contents

List of abbreviations i
List of Figures i
List of Tables ii

1. Introduction 1
2. Literature Review 2
2.1. EMEMUS 2
2.2. ROADMAPPING Framework 4
2.2.1. RO 5
2.2.2. AD 6
2.2.3. M 7
2.2.4. A 8
2.2.5. PP 9
2.2.6. ING 10
2.3. Academic Disciplines 10
2.3.1. Smart et al. and Holland's Approach 14
2.3.2. Biglan and Kolb 20
2.3.3. Biglan & Holland's approaches compared 25
2.4. Vocabulary Acquisition 26
2.4.1. Academic Disciplines and Academic Discourse 26
2.4.2. Word Knowledge 30
2.5. Academic English and the Academic Word List 36
2.5.1. Academic English 38
2.5.2. Academic Word Lists (AWL and others) 38
2.5.3. Criticisms of the AWL 39
2.5.4. Other Written Academic Word Lists 42
2.5.5. Spoken Academic Word Lists – HSWL and SSWL 43
2.6. Vocabulary Testing 46
2.6.1. Vocabulary Levels Test (VLT) 47
2.6.2. Form-Recognition Test of Academic Verbs (FoRTAV) 50
3. Methodology and Research Design 51
3.1. VLT and FoRTAV 53
3.2. Questionnaire 55
3.3. Pre-testing 57
3.4. Data Processing and Variable Development 59
3.5. Participants 61
4. Results & Statistical Analysis 66
4.1. VLT & FoRTAV Test Results 67
4.2. Biglan Groups and Vocabulary Scores 72
4.3. Holland Groups and Vocabulary Scores 77
4.4. English Medium Education and Vocabulary Scores 81
4.5. Other Factors and Vocabulary Scores 86
4.6. AWL vs. FoRTAV 90
4.7. Analysis and Interpretation 93
5. Conclusion 102
6. References 104
7. Appendix I – Questionnaire 112

8. Appendix II – Vocabulary Tests ……...…… 122
9. Appendix III – Data and R-code ……..…… 122
10. Appendix IV – Extra Figures ……...…… 123
11. Appendix V – German Abstract …….. 126

List of abbreviations

AVL ……. Academic Vocabulary List
ASWL …….. Academic Spoken Word List
AWL ……. Academic Word List
BNC ……..… British National Corpus
CEFR……..…….. Common European Framework of Reference for Languages
CLIL…… Content and Language Integrated Learning
COCA …… Corpus of Contemporary American English
EAP ……...…… English for Academic Purposes
ELF ……...…… English as a Lingua Franca
EME ……..…….. English Medium Education
EMEMUS ……..…. English-Medium Education in Multilingual University Settings
EMI …… English Medium Instruction
EML …….…… English Medium Learning
ESP …….…….. English for Specific Purposes
FHs ……..… Fachhochschulen
FoRTAV …… Form Recognition Test of Academic Verbs
GSL ……...…. General Service List
HEIs …….…… Higher Education Institutions
HSWL …….…….. Hard science Spoken Word List
L1…… First language or languages
L2 ……...…… Any language learned after an L1
NS ……..……. Native Speaker
NNS …….. Nonnative Speaker
SSWL …… Soft science Spoken Word List
VLT …….…… Vocabulary Levels Test
VST …….. Vocabulary Size Test

List of Figures

Figure 1. ROADMAPPING from Dafouz & Smit 2023 ……..……..… 5
Figure 2. Holland's Congruency Hexagon …….……..… 17
Figure 3. Example question from Schmitt, Schmitt and Clapham's 2001 VLT ……...… 48
Figure 4. Example question from Ghamarian-Krenn's FoRTAV ……..….. 51
Figure 5. Example question from VLT in grid-format ……..…….. 53
Figure 6. Example question from FoRTAV in grid-format …….…… 54
Figure 7. Histogram of participant genders ……..……..… 62
Figure 8. Histogram of participant ages …….…… 62
Figure 9. Histogram of participant-ranked English level …… 63
Figure 10. Histogram of Biglan's groups and number of participants in each group ……..… 64
Figure 11. Histogram of Holland's groups and number of participants in each group ……... 64
Figure 12. Boxplots of participant scores (in percentages) for each test type …… 68
Figure 13. Quantile-Quantile Plot of 10k residuals …… 70
Figure 14. Bar plot of pass/fail rates for all participants by test section …….. 70
Figure 15. Boxplots of Biglan's groups vs. 2k results …… 73
Figure 16. Boxplots of Biglan's groups vs. 3k results …… 73
Figure 17. Boxplots of Biglan's groups vs. 5k results ……...…73
Figure 18. Boxplots of Biglan's groups vs. 10k results ……73
Figure 19. Boxplots of Biglan's groups vs. FoRTAV results ……..74

Figure 20. Boxplots of Biglan's groups vs. AWL section of the VLT results ……74
Figure 21. Boxplots of Biglan's groups vs. VLT & FoRTAV results ……74
Figure 22. Boxplots of Biglan's groups vs. VLT results ……..74
Figure 23. Bar plot of passing rates of each test section split into Biglan's groups ……75
Figure 24. Boxplots of Holland's groups vs. 2k results ……..77
Figure 25. Boxplots of Holland's groups vs. 3k results ……..77
Figure 26. Boxplots of Holland's groups vs. 5k results ……..78
Figure 27. Boxplots of Holland's groups vs. 10k results ……78
Figure 28. Boxplots of Holland's groups vs. FoRTAV results ……78
Figure 29. Boxplots of Holland's groups vs. AWL section of the VLT results ……..78
Figure 30. Boxplots of Holland's groups vs. combined VLT & FoRTAV results ……..79
Figure 31. Boxplots of Holland's groups vs. VLT results ……79
Figure 32. Bar plot of passing rates of each test section split into Holland's groups …….. 80
Figure 33. Scatter plot of writing EME scores vs. FoRTAV scores ……... 83
Figure 34. Scatter plot of presenting EME scores vs. FoRTAV scores …….. 83
Figure 35. Scatter plot of reading EME scores vs. combined VLT & FoRTAV scores ……. 84
Figure 36. Scatter plot of reading EME scores vs. FoRTAV scores ……... 84
Figure 37. Scatter plot of summed EME scores vs. combined VLT & FoRTAV scores …... 85
Figure 38. Scatter plot of summed EME scores vs. FoRTAV scores …… 85
Figure 39. Scatter plot of number of semesters vs. 10k score …… 86
Figure 40. Scatter plot of number of semesters vs. VLT score ……... 86
Figure 41. Scatter plot of participant-ranked English level vs. VLT score …… 88
Figure 42. Scatter plot of participant-ranked English level vs. VLT & FoRTAV score …..… 88
Figure 43. Scatter plot of participant-ranked English level vs. FoRTAV score ……..… 88
Figure 44. Scatter plot of participant-ranked English level vs. 10k score ……... 88
Figure 45. Scatter plot of age vs. 10k score …… 89
Figure 46. Scatter plot of age vs. VLT score …….. 90
Figure 47. Scatter plot of AWL score vs. FoRTAV score ……..…… 91
Figure 48. Scale of AWL and FoRTAV items based on correctness frequency …… 92

List of Tables

Table 1. Holland types ……..… 14
Table 2. Kolb & Biglan academic disciplines ……..…… 21
Table 3. Bauer & Nation (1993) word family scheme …… 31
Table 4. Nation's 18 aspects of word knowledge (2016) …… 33
Table 5. Findings from spoken and written hard and soft science vocabulary studies …… 44
Table 6. Pilot testing scores …… 58
Table 7. Participant nationalities …… 65
Table 8. Participant universities …… 66
Table 9. Test results and big 5 statistics …… 69
Table 10. Top 15 frequently incorrectly answered words ……71
Table 11. Most frequently incorrectly answered items for all participants and Biglan's groups …… 76
Table 12. Most frequently incorrectly answered items for all participants and Holland's groups …… 81
Table 13. Spearman's ρ correlation coefficient for EME values and test results …… 82
Table 14. Spearman's ρ or Kruskal-Wallis p-value for other variables ……..... 87

1. Introduction

With English-Medium Education in Multilingual University Settings (EMEMUS) already existing as a prevalent occurrence all the while continuing to grow in popularity and diversity, the need for adequate reports on the present situation of the students enrolled in these programs is abundantly clear (Richards & Pun 2021: 12). The potential for large vocabulary acquisition in English due to these types of programs is large with many students enrolling in these programs with some hopes of maintaining or improving their English skills. Universities advertise these programs with the claim that students will be more prepare for an English-speaking workplace in the future by enrolling in these English Medium programs.

The ROADMAPPING framework, a tool developed to analyze and study EMEMUS, will serve as a partial background and inspiration for many parts of the present study including the questionnaire development and the discussion of results. The framework aims to promote the examination of the Roles of English in relation to Other languages (RO), Academic Disciplines (AD), language Management (M), Agents (A), Practices and Processes (PP), Internationalization and Glocalization (ING) as they are present and involved in policy making, common practices, and more when looking at English Medium Education (EME) at Higher Education Intuitions (HEIs) (Dafouz & Smit 2016). The framework will help situate much of the information that is particularly relevant for the participants enrolled in EME programs.

Two vocabulary tests will be used to measure receptive knowledge of general vocabulary words and academic vocabulary words. The two vocabulary tests are Schmitt, Schmitt and Clapham's (2001) Version 2 of the Vocabulary Levels Test (VLT) as well as Ghamarian-Krenn's (2023) newly developed Form Recognition Test of Academic Verbs (FoRTAV). The VLT will serve as a test of general vocabulary knowledge with one specific section focusing on academic vocabulary from Coxhead's (2000) Academic Word List (AWL). The FoRTAV will serve as a test of academic verbs, also taken from Coxhead's (2000) AWL. In addition to the vocabulary tests, the participants will also complete a short questionnaire with items inspired by several ROADMAPPING concepts such as academic discipline, English exposure in the current study programs, study abroad experience, languages used inside and outside of the classroom, as well as some short biographical information such as age, gender, and semesters of university attended.

With the collected information from the vocabulary tests and questionnaire, interesting insights into the vocabulary knowledge of students in Austria will be observed and reported. Several statistical significance tests will be run in addition to graphical summaries of all the

relevant results to the research questions. Before reporting on the results, a literature review will be presented in the following section that will cover several key concepts: EMEMUS, a framework for examining EMEMUS (the aforementioned ROADMAPPING), Holland and Biglan's approaches to academic disciplines, vocabulary acquisition, academic vocabulary and academic word lists, and vocabulary testing. Then the methodology of the present study will be covered, followed by a section presenting the collected data and results from the participant surveys and test results. Finally, a discussion of the results and short interpretation will conclude the inquiry.

2. Literature Review

2.1. EMEMUS

Universities across the world are integrating English into their curriculum more and more as the role of English as a Lingua Franca (ELF) continues to grow (Richards & Pun 2021: 12; Galloway 2017). With the rise of programs being offered in English as well as the general increasing role of English in the world, the need to understand and research these new programs is key to their success. Without taking the extreme stance that English will eventually replace all other languages in academic settings, a more moderate position taken by some researchers states that eventually and inevitably, "the world will become diglossic with one language for local communication, culture, and expression of identity, and another – English – for wider and more formal communication, especially in writing" (Costa & Coleman 2013: 11). This type of claim demonstrates the indisputable role English has and will continue to have at universities and in education in general outside of English-speaking countries no matter the tension or opposition that comes with it. English Medium Instruction (EMI), English Medium Learning (EML), Content and Language Integrated Learning (CLIL), immersion education, theme-based language teaching, bilingual education, and multilingual education, can all be considered relevant research fields when examining the role of English in these educational settings. However, some of these labels are too precise for the present study and others are too broad.

English Medium Education in Multi-lingual University Settings (EMEMUS) as a term is preferred over any of the others since the "label is semantically wider" as explained by Dafouz & Smit (2016: 399) and "does not specify any particular pedagogical approach or research agenda". With this term, an unbiased and all-encompassing approach can be taken when studying the programs. The term English Medium Education (EME) is used as a shorthand for EMEMUS which include "higher education scenarios in which English is used as an additional language for education or for related purposes, such as administration or

management" (Dafouz & Smit 2020: 4). The focus on education, instead of solely learning or instruction, broadens the scope of this research field to include both aspects, "instead of prioritizing one over the other" (Dafouz & Smit 2020: 3). Additionally, multilingual is meant "in the widest sense" (Dafouz & Smit 2020: 3) thus recognizing that English is only one of the many languages present in these environments which reflects the multilingual reality that exists at many institutions of higher education.

Other labels that are often used instead of EME include EMI, CLIL, and bi/multilingual education. These terms are sometimes, seemingly contradictorily so, too precise while being too vague in their scope if the goal is to capture and describe the true linguistic realities that exist at many Higher Education Institutions (HEIs). English Medium Instruction fails to recognize the give-and-take nature of teaching and education by instead focusing the attention specifically on the teacher and their contribution to the classroom (Dafouz & Smit 2020: 3). The reality is that many classroom settings strive for a better balance between teacher and student interaction and pedagogically, student-student as well as teacher-student interactions are being seen as a vital part of education and the learning process more and more. Teacher-centered classrooms are still a reality, but even when researching these settings, it is important to include information about the students, the student perspective, and other factors that happen outside the classroom since these all impact the educational sphere.

Content and language-integrated learning (CLIL) is again too narrow and broad in its definition to truly represent the type of education that is a reality for many students in HEIs. CLIL emphasizes content and language learning in some form at the same time or in the same setting. While this may be a goal for some programs, the parallel learning of English (be it implicit or explicit) and content relevant to the field of study, is not always a goal, and the motivations for teaching in English are varied and vast. While CLIL could happen in EMEMUS, it is not necessarily going to happen. Lastly, the term CLIL can be applied to education at any level, not just in HEIs which broadens the scope too wide since education at the primary level is vastly different from at the secondary level, which is again vastly different from the tertiary level.

Following similar lines of argumentation, other labels do not appropriately delineate EME nor create a space for this subset of educational research. It is for this reason that Dafouz and Smit coined the term EMEMUS nearly a decade ago and continue to use it to describe and investigate these types of situated learning experiences. To fully grasp the concept as a whole and what it does and does not include, a definition of the label will be provided next.

One of the defining factors of EME is indeed their great diversity. To highlight this diversity, EMEs have been described as a setting where "variation is part and parcel to the uniqueness" as well as being "unique constellations of […] global forces and locally relevant factors," "intricate and multi-layered," changing at "a faster pace than the surveys trying to measure them," "dynamic and highly situated," coming "in different shapes and forms" with "diverse, dynamic and possibly conflicting realizations" (Rose et al. 2021: 4) (Dafouz & Smit 2019: 4, 5, 13, 42). The variety that exists within the bounds of what can be considered EME is vast with degree programs that run the gamut from little English exposure to essentially only English exposure. Even when comparing other factors such as the number of languages present and the amount of English used in these settings, some of the differences that can exist are noteworthy.

Dafouz and Smit highlight some of this diversity in their 2020 publication about EME by "offering a 'collage' of current EME realities" (Dafouz & Smit 2020: 12). In this chapter of their book, they look at a few examples of EME in different countries. The sites of focus are the University of Stockholm in Sweden, the University of Maastricht in the Netherlands, the University of the Basque Country in Spain, Zayed University in Dubai, UAE, Waseda University in Tokyo, Japan, as well as Universidad del Norte in Barranquilla, Colombia. These EME sites vary in several ways, and they do not offer a fully representative look at the variety of types of programs that can exist. However, the variety that is presented in these few case examples is remarkable and demonstrates some ways in which these programs can be similar and different ranging from programs at private and public universities, with English at the center or English acting as an additional, even third language (Dafouz & Smit 2020: 13). Thus, reflecting the "individuality and complexity" (Dafouz & Smit 2020: 26) of EMEMUS. The best way to demonstrate this diversity will be to further illustrate it through the framework that is presented by Dafouz and Smit that will be discussed next. Each aspect of the framework will be defined and elaborated upon with a small example from the case studies mentioned above to highlight the diversity that exists within EMEMUS.

2.2. ROADMAPPING Framework

The complex and diverse nature of EME goes hand in hand with the need for a systematic yet flexible approach to studying these sites. One such approach that has been developed by researchers Smit and Dafouz, is the ROADMAPPING framework. Drawing from sociopolitical, linguistic, and psychological factors, the framework goes "beyond discipline boundaries" (Dafouz & Smit 2019: 6) encouraging researchers and university staff members to step outside

of their departments and create a dialogue between disciplines and researchers regarding EME; the hope being that this framework will be one tool to facilitate and guide interdisciplinary conversations that surround EME. Illustrated in Figure 1 below, with discourse at the middle of the diagram and the Roles of English (in relation to Other languages), Academic Disciplines, (language) Management, Agents, Practices & Processes, and INternationalization and Glocalization surrounding it, the many factors that can guide these constructive interdisciplinary conversations around EMEMUS are represented (Dafouz & Smit 2023: 3).

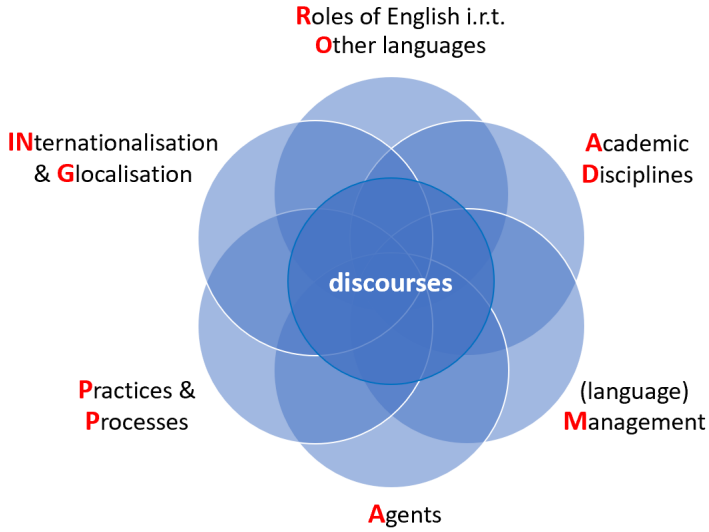


Figure 1. The ROAD-MAPPING framework for EMEMUS (Dafouz & Smit 2023: 3)

Discourse is placed at the middle since it is seen as the method by which these other factors can be examined (Dafouz & Smit 2016: 403).

The order in which the dimensions are presented does not reflect any sort of order or importance, each dimension can play an equally important role when stepping back and examining the larger reality of EMEMUS. Dafouz and Smit explain this relationship by stating that when considering one single dimension "the remaining dimensions are not irrelevant; on the contrary they play an important role in the explanatory factors in the interpretations of the findings" (Dafouz & Smit 2023: 12). Many of the examples will highlight that, even if stakeholders do support the cause and recognize the benefits of EME, it is not friction-free and the conversations concerning EME deserve attention (Tatzl 2011: 263).

2.2.1 Roles of English in relation to other languages (RO)

The Roles of English (in relation to other languages) dimension is aimed at defining and explaining "the communicative functions that language fulfills in HEIs" (Dafouz & Smit 2022: 5). With this dimension, all the languages present, in official and unofficial capacities, in the

given program are examined along with the intimate relationship that connects each language to the other ones. The four factors that describe the "interrelatedness" and "respective context" (Dafouz & Smit 2016: 299) are the societal, institutional, pedagogical, and communicational factors (Dafouz & Smit 2020: 49). Sometimes English serves as a complementary language to other languages and other times it is seen as more conflicting. Occasionally, some languages do not have the resources, such as textbooks or scientific publications, to support higher education for students, and therefore, resources in English are used to add additional support. In some Swedish universities, it is common for the lectures to be completely offered in Swedish, yet the reading materials are only offered in English (Richards & Pun 2021: 19). This approach to EME is called parallel language use since the languages are not seen as intersecting but existing side by side.

This approach can be seen as supplementing the Swedish education plan but at the same time as conflicting and contributing to a vicious cycle. If there is no need for textbooks to be published in Swedish, since all the reading will be done in English anyway, then no textbooks will be produced. If demand for textbooks in Swedish wanes, then publishers will not have the monetary incentive to publish books in Swedish. There is a fine balance for many Swedes between protecting the Swedish language, while at the same time being realistic and accommodating foreign students who may not speak Swedish but want to study in Sweden. This phenomenon is present in many EMEMUS when considering the desire by many to preserve a local culture and language, while hoping to reach out and include members of the larger international research community. This illustrates one way that the roles of various languages, including English, can form and shape academic spheres.

2.2.2 Academic Disciplines (AD)

The next dimension relates to academic disciplines in the sense of academic disciplinary culture and academic literacies. Academic disciplinary culture refers to "subject-specific conventions, norms, and values that define different disciplinary areas" (Dafouz & Smit 2020: 60). Academic literacies refer to the various types of spoken and written academic products that are created and "conform to socially conventionalized situated practices" (Dafouz & Smit 2022: 5) within the academic disciplines. These academic languages are highlighted as making up a unique part of academic groups since academic groups have a "distinct language or at least a distinct dialect and a variety of symbolic ways of demonstrating its apartness from others" (Becher 1994: 151). All the while, academic groups can communicate with other academic groups since there is a shared general academic language (Becher 1994: 151). Examples and

further explanations of this dimension in relation to EMEMUS can be found below in Section 2.3 since it is the most relevant dimension for the present study.

Essentially, the academic discipline dimension highlights the need for the lecturer and student "to acculturate in a second or foreign language and, at the same time, develop disciplinary discourse conventions" (Dafouz & Smit 2019: 52) which can be a difficult task for many when knowledge construction has been claimed to be "specific to particular cultures, settings, periods and the natural languages used to construct it" (Trowler in Smit & Dafouz 2020: 51). Given the example of textbooks only being offered in English for certain academic disciplines, this leads to the situation where "many students in these situations find L2 textbooks an additional obstacle to understanding already difficult material and may become less motivated to read them" (Busby 2021: 60). Some students have even remarked being so challenged by reading textbooks in English that they have changed their degree program to one that did not have as much or any required reading in English (Busby 2021: 75). The literary practices and norms of various academic disciplines can potentially lead to students deciding not to enter certain fields of study.

2.2.3 Language Management (M)

In the framework, the language management dimension deals with the different policies that are present that impact which languages are used in which settings and to what extent. One university may require that all administrative tasks are handled in the local language, even if the program of focus is fully offered in English. This can pose issues for international participants who may need to handle some of these administrative tasks before even arriving in the country or getting a chance to learn some of the local language if they decide to learn it at all. These policies can come from all levels of power including the universities themselves, ministries of education, departments, or on a class-by-class scenario. They can also exist in certain forms such as explicitly being stated in laws or procedural handbooks or only being apparent when present in these situations, in unwritten forms.

Again, we can take the example of parallel language use in Sweden to highlight this dimension since there are multiple opportunities for policies at different administrative levels to be implemented to set the language of textbooks in different university classes if that is desired. Additionally, the University of the Basque Country in Spain has a long history of bilingual education with Spanish and Basque traditionally being the languages of focus. However, more frequently, English is being introduced to the university through several programs being offered in English and now the university is facing a potential identity crisis. Basque is in the name and

at the center of the university, but the introduction of English appears to correspond with a direct disappearance of the Basque language from the university (Dafouz & Smit 2019: 20). Seemingly, the bilingual nature of the university is shifting from Basque and Spanish to Spanish and English instead of the trilingual environment that was originally the goal when introducing these English programs. To combat this growing issue, there is potential for several policies, formally or informally, to promote the use of English, Spanish, and Basque such as mandatory supplementary Basque lessons, a Basque entry exam, Basque cultural events, or stricter rules on when and where each language should be used. Since the proclamation of Basque as an official language, the university and the local autonomous government have "collaborated economically to guarantee the possibility of studying in Basque and/or Spanish at pre-university and tertiary levels" (Dafouz & Smit 2019: 20) showing the collaboration that exists between the governmental goals and university power. These would all belong to the discussion of how languages are managed and what governing levels these decisions are coming from as well as any attitudes or opinions towards these policies and how they impact the languages present at the university.

2.2.4 Agents (A)

The next dimension that comes into play for EMEMUS is the agents or "social players" who are involved at all levels. As mentioned earlier, even if the agents are supportive of EME, they still recognize the difficulties that can come with EME. In one interview, a tertiary English-medium instructor described classroom conversations and EME "like driving with a hand brake on" (Tatzl 2011: 263) meaning that the conversations felt forced and difficult at times due to language barriers and additional concerns that do not exist when an instructor is teaching to students in their shared common first language. Some of these players take on multiple titles or roles such as professors with additional administrative obligations who may be responsible for making curriculum decisions, accepting students into programs, or creating programs in English from scratch. The many roles one person may have, along with their "hierarchical status, their professional concerns, and/or their English language proficiency" (Dafouz & Smit 2022: 5) are all relevant factors to defining what they do. Depending on these factors, if one member of the staff can only speak English or does not speak any English, the languages in use, or the number of people needed to communicate may change. Furthermore, these actors or participants are not only impacted by higher education policies, but they also help create some of these policies. Agents of both micro- and macro-levels need to be considered when implementing higher education policies.

To understand the impact that micro-agents, or more locally based agents, can have on EMEMUS if they are resistant to the proposed policies or changes one just needs to look at the example of Waseda University in Japan. The micro-agents did not fully agree with the reasoning for a new policy that had been implemented and started to question and prevent policy changes from being implemented since they felt overlooked or forgotten by the macro-agents (Dafouz & Smit 2020: 55). It is not only important to look at the policies in action then, but to also consider the people and their reactions to changes that may impact their daily work-lives and systems they are very familiar and comfortable with. As mentioned earlier, even if the Agents recognize the potential benefits for all entities, they still have their interests to consider, and this can impact many factors in EME.

2.2.5 Practices and Processes (PP)

In the Practices and Processes dimension, the various "administrative, research, and educational activities" are in focus as they contribute to and are created by EMEMUS. This includes activities at the classroom level, such as an examination of the types of discourse or classroom activities at use, as well as out-of-classroom activities by administrators and professors. This is one aspect where the traditional classroom focus of pedagogical studies is expanded and the practices and policies of the researchers, professors, and administrators are also examined to include but not limited to various "teacher professional development or stages of internationalization" (Dafouz & Smit 2022: 5).

Translanguaging, or the act of using any linguistic resources at an individual's disposal to interact with each other and texts to "create meaning and deepen learning" (Palfreyman & Van Der Walt 2017: 12) in a larger educational context among other education practices is one of the research foci that can help researchers inform instructors and professors of ways to better adapt to the unique needs of these multi-lingual classrooms that are part and parcel of EMEMUS (Dafouz & Smit 2020: 57). This would include using classroom teaching materials from various languages if it is to be assumed that the players present in the classroom can interact with these materials. Additionally, it might appear in the form of letting students communicate with each other in whatever common languages they might have. This method is better for communication and learning, as well as a better reflection of how communication works outside of classroom settings. If these resources exist and are available for use, the argument stands to let participants use any and all of the resources that they can. Conversations and studies into the impact of phenomena like translanguaging are a part of the practices and processes aspect.

2.2.6 Internationalization and Glocalization (ING)

The last dimension is the Internationalization and Glocalization dimension which highlights the role and importance of external international factors on the local universities as well as the internal, national, and local factors that also need to be considered (Dafouz & Smit 2022: 5). Cited as the "tensions and synergies," (Scott 2011: 61) this quote highlights the duality of this dimension. Sometimes local and international forces collide in moments of conflict with hopes of preserving local culture or languages while also wanting to expand and diversify opportunities for those who are not from the host country. The balance, again, is a fragile and delicate topic with many identities and beliefs at the center of the conversation as to what it means to be an international institution while also still honoring the local identity. These identities are profoundly personal and can lead to strong debates. Some common concerns relate to the potential "curricular homogenization and the imposition of Westernized approaches and paradigms on local academic cultures" (Leask, 2015 in Dafouz & Smit 2019: 58). Additionally, fear exists that by requiring English knowledge of local students and academics, these agents will lose first-language academic literacy skills and potentially halt or lose the development of disciplinary language in these languages.

Generally, two types of EME emerge in globalized contexts. The first are internationalized EME programs "which are driven by neoliberalist internationalization forces and are often designed to attract international students" (Rose et al. 2021: 8). This can be seen as a method of bringing the outside world and international students into the local community. Meanwhile, the glocalized EME programs "mainly cater to local needs to build-up a domestic student knowledge base and to develop graduates into globally competitive human resources" (Rose et al. 2021: 8) which can be seen as a way to bring local cultures and discoveries to the outside world. In this way, giving a voice to students and making sure they possess the global skills to be able to inform and spread information that has been developed locally. This aspect focuses on the push and pull forces from the outside world on local academic practices and the impact of local academic practices on the outside world.

2.3. Academic Disciplines

The most logical approach to discussing academic disciplines would be to start with a definition of what exactly qualifies as an academic discipline. Unfortunately, groupings of academic disciplines are not as straightforward and clear-cut as perhaps initially imagined nor are they

equal in value for every situation. When conceptualizing academic disciplines for research purposes, it is important to use categorization schemes that are theoretically grounded instead of ones that follow the individual university organization schemes such as "social sciences," "life sciences," and "classics" (Pike & Killian 2001: 430). While some universities may place some subjects in certain departments, and other universities use a different organizational structure, the essence of these subjects remains the same. In nearly every field, the disciplinary culture transcends institutional and national boundaries meaning no matter the department or college the subject is placed in, it will always have commonalities with other iterations of the subject at different universities across the world (Becher 1994: 153). This is why physicists educated in Japan can research and collaborate with physicists in Austria; the disciplinary culture extends much further than the department doors or national borders as it connects knowledge.

Academic disciplines when confined to a simple and strict taxonomy, such as the departmental structure of universities or the labels "natural science" and "social sciences," are not fairly represented. These types of taxonomies fail to consider additional outside factors that can impact the expression of these disciplines in various academic settings. As stated by Trowler (2014a: 1728) "internal politics and policies of universities" can have a significant impact on the ways disciplines "are presented and named in their organization institutions". The departmental structure of one university may be more reflective of the funding structure, chronological order of when the subjects were introduced to the university, or simply room availability, instead of the academic similarities of certain subjects. Furthermore, courses of a similar nature offered at different universities may be classified differently "for reasons of image" as is the case when universities label a class as "economics" when portraying an "academic" image and the same class as "economics and business" at a different university when portraying "a more vocational orientation" (Simpson 2017: 1530). There are various motivations for departmental organization schemes, and not all of these are relevant to the research of academic disciplines.

The "changing nature of knowledge domains over time" is crucial in the impact it has on the "identities and cultural characteristics of disciplines" (Becher & Trowler 1991: 43). This is how a discipline, such as economics, can now find itself shifting from initially a more theory-based social science to now incorporating advanced applied statistics. With the advancements in certain modeling and technological tools, some subjects are changing and evolving with different approaches that answer main questions while also centering or removing themselves from one central paradigm – two factors that will be highlighted later as the key to defining

academic families in one approach. However, universities do not respond quickly to the shifting and changing nature of academic disciplines and will keep economics in the same department and let it be slightly different from the other members. Quite simply, following the departmental structure of one university to define academic disciplines would be arbitrary and inaccurate for the aims of this study. Instead, the present study will group academic disciplines together on factors outside of departmental structures. With this approach, there is then a system in place for distinguishing the academic disciplines instead of other factors that can impact which department a subject belongs to. Without an organized and theoretically based approach, simply relying on the organizational structure or long-held "natural" and "social sciences" categories does not make room for the inevitable shift of academic disciplines. The basis for the two discipline-defining approaches is explained below.

To combat the different university structuring methods, Trowler (2014a: 1725) proposes a "family resemblance" approach to academic disciplines instead, based on Wittgenstein's (Fox 2014) idea where groupings are not strict but rather are based on commonalities that may or may not be found in each family member. This is also why academic disciplines that have experienced a shift in approach or technology in the past are still recognizable to present-day researchers in that discipline but are contextually contingent on the time and location of practice. Trowler (2014a: 1724) additionally emphasizes that family resemblances and therefore academic disciplines can vary in terms of different relationships that are present in these familial interactions. This means that the ways researchers interact within an academic discipline can be different from how learning and teaching take place in these same academic disciplines. In other words, in a purely hypothetical example, it may be determined that professors in anthropology are more likely to engage in interactive learning, but this does not mean that the students in anthropology majors are equally likely to be interactive learners. What holds for one aspect of an academic discipline, does not automatically transfer to another aspect.

Trowler (2014b: 20) elaborates by saying "[t]his approach moves away from the conceptually disabling strong essentialism of the earlier work to a weak form of essentialism which gives intra-disciplinary difference room to breathe" thus emphasizing the differences that exist within and between academic disciplines. When compared with genetics, Becher makes a metaphor between academic disciplines and the genotype-phenotype distinction. In this metaphor, the differences that can exist within a subject, at different locations of practice or at different points in time, can be seen as phenotypical variations, or variations that arise on the surface and are expressions of the same genotypical culture that is "endemic to each discipline"

(Becher & Trowler 1991: 153). In this sense, the commonalities that exist within academic disciplines are the genes that are inherited and the differences within a discipline are due to varying expressions of the same genes when they interact with their local environments. It is why sometimes people with genetic predispositions for certain features, diseases, or other phenomena, do not ever end up getting these diseases or displaying these features. Academic disciplines will express their core features and genes in different manners depending on various environmental factors, again explaining why a physics program in Japan and Austria will have many commonalities but also some differences.

The main methods for grouping academic disciplines outside of organizational structures were thoroughly examined in two systematic reviews, initially in 1996 by Braxton and Hargens and then again in 2012 by Jones[1]. In 1996, Braxton and Hargens identified 11 different models for classifying academic disciplines but decided to focus on 5 of them that they found most relevant to research in higher education. In his 2012 study, Jones (2012: 11) decided to focus on only two schemes, one that was previously examined in the Braxton & Hargens 1996 review by Biglan (1973a, 1973b) and Kolb (1981), and a new scheme that had risen in popularity within the higher education research community after 1996 by Smart et al. (2000, 2006, 2007, 2009) based on Holland's work (1969, 1976, 1989, 1997). Jones references the 1996 paper by Braxton and Hargens in his analysis of analytical frameworks and research relating to academic disciplines since 1996, starting essentially where the Braxton and Hargens' review left off. The initial analysis by Braxton and Hargens (1996: 35) concludes that "the differences among academic disciplines are profound and extensive". In Jones' report (2012: 21), he comes to a similar conclusion that "there are substantial differences among academic disciplines in a number of areas with regard to the faculty" and that there appeared to be little change in classifying schemes in research since the 1996 analysis. Jones (2012: 21–22) states that this might be because "the classification schemes created prior to 1996 have been sufficiently accepted" and that "scholars see no need to modify them," emphasizing the relative stability in the field since 1996. Jones (2012) found that the most widely used classification schemes in use since 1996 have been the Smart et al. classification scheme and the Biglan and Kolb schemes, which will both be elaborated on in the following subsections. In this study, each of these schemes will be used to compare differences among different groups of students based on their different academic groupings.

[1] This section mainly summarizes the findings from two systematic reviews. The initial review is from Braxton & Hargens from 1996 and the second review is from Jones from 2012. In Jones' (2012) review, he mentions the findings from the initial systematic review, which mainly explored Biglan & Kolb's approach, and then expands upon them to include more recent approaches, such as Smart et al.'s method using Holland's approach.

2.3.1 Smart et al. and Holland's Approach

The first classification scheme that will be used is the one developed by Smart et al. (2000, 2006, 2007, 2009), which is based on Holland's classification scheme. While the schemes are largely the same, it is important not to confuse the two as there are a few key differences that will be highlighted. To begin, it makes most sense to look at Holland's classification scheme first. Then the adaptations made to Holland's classification scheme by Smart et al. to adjust the approach specifically for higher education studies will be explained.

In Holland's theory, he separates people into six main personality types which are: Realistic, Investigative, Artistic, Social, Enterprising, and Conventional personalities. Each of these six personality types has specifically attributed "salient attitudes, interests, and competencies" that were developed by Holland over a series of publications throughout his career (Holland et al. 1969; Holland & Gottfredson 1976; Holland 1997; Holland, Holmberg & Rosen 1989). In this theory, there is a psychological and sociological component that helps explain how these six personality types develop, how they are related to each other, and how people become or transform into their personality type. With each personality type, there is also a corresponding model environment where people who have one type of personality will thrive best. For example, in the realistic environment, those with a realistic personality type will, in theory, succeed because the environment fits their personality. The theory states that behaviors and skills that are possessed by people with these personality types are directly rewarded and encouraged in the matching environment, therefore rewarding, and encouraging them to maintain these behaviors. The six types, the salient attributes, their matching environment, and example careers are provided in the table below. These points have all been taken and combined from Smart's descriptions and careers list (Smart, Feldman & Ethington 2006; Smart 2000) which was adapted from Holland's previous work alone and with Gottfredson (Holland 1997; Gottfredson & Holland 1991).

Table 1. Holland's six personality type, salient attributes, model environments and matching careers and college majors (adapted from Smart, Feldman & Ethington 2006; Smart 2000)

TYPE and Salient Attributes	Model Environment	Careers or college majors
REALISTIC people prefer activities that involve the explicit, ordered, and systematic manipulation of objects, tools, machines, and animals, and avoid educational and interpersonal activities. Realistic people perceive themselves as practical and conservative, having mechanical, technical,	The behavioral tendencies in Realistic environments lead to the acquisition of mechanical and technical competencies and to a deficit in human relations skills. Realistic environments reward people for the display of conforming	Electrical engineering, Mechanical engineering, marina science, drafting/design, military science, computer engineering, forestry, surveyor, poultry science, and farmer

and athletic abilities, and as lacking ability in social skills.	behavior and practical accomplishment.	
INVESTIGATIVE people prefer activities that involve the observational, symbolic, systematic, and creative investigation of physical, biological, and cultural phenomena in order to understand and control such phenomena, and avoid persuasive, social, and repetitive activities. Investigative people perceive themselves as cautious, critical, complex, curious, independent, precise, rational, and scholarly.	The behavioral tendencies in Investigative environments lead to the acquisition of analytical, scientific, and mathematical competencies and to a deficit in persuasive and leadership abilities**.** Investigative environments reward people for skepticism and persistence in problem solving, documentation of new knowledge, and understanding solutions of common problems.	Biology and life sciences, health/medical technologies, economics, geography, math/statistics, physical sciences, finance, aeronautical engineering, civil engineering, chemical engineering, premedicine, anthropology, ethnic studies, geography, sociology, astronomy, earth science, pharmacy, chemist, physicist, geologist, anthropologist, laboratory assistant, and medical technician
ARTISTIC people prefer ambiguous, free, and unsystematized activities that involve the manipulation of physical, verbal, or human materials to create art forms or products, and avoid routine activities and conformity to established rules. Artistic people perceive themselves as expressive, original, intuitive, nonconforming, introspective, independent, emotional, and sensitive.	The behavioral tendencies in Artistic environments lead to the acquisition of innovative and creative competencies—language, art, music, drama, writing—and to a deficit in clerical and business system competencies. Artistic environments reward people for imagination in literary, artistic, or musical accomplishments.	Architecture, fine arts (art, drama, music), foreign languages, English, music, speech, theater, and environmental design, composer, musician, stage director, dancer, interior decorator, actor, and writer
SOCIAL people prefer activities that involve the manipulation of others to inform, train, develop, cure, or enlighten others, and avoid explicit, ordered, systematic activities involving materials, tools, or machines. Social people perceive themselves as cooperative, empathetic, generous, helpful, idealistic, responsible, tactful, understanding, and warm.	The behavioral tendencies in Social environments lead to the acquisition of interpersonal competencies and to a deficit in manual and technical competencies. Social environments reward people for the display of empathy, humanitarianism, sociability, and friendliness.	Ethnic studies, home economics, humanities (history, philosophy, religion, rhetoric), library science, physical and health education, psychology, social sciences (anthropology, political science, social work), education, teacher, speech therapist religious worker, counselor, clinical psychologist, and nurse
ENTERPRISING people prefer activities that involve the manipulation of others to attain organizational goals or economic gain, and avoid scientific, intellectual, and abstruse activities. Enterprising people perceive themselves as aggressive, ambitious, domineering, energetic, extroverted, optimistic, popular, self-confident, sociable, and talkative.	The behavioral tendencies in Enterprising environments lead to an acquisition of leadership, interpersonal, speaking, and persuasive competencies and to a deficit in scientific competencies. Enterprising environments reward people for the display of initiative in the pursuit of financial or material accomplishments, dominance, and self-confidence.	Business, communications, computer/information science, law, public affairs, journalism, marketing, industrial engineering, buyer, sports promoter, television producer, business executive, salesperson, travel agent, supervisor, and manager
CONVENTIONAL people prefer activities that involve the explicit, ordered, systematic manipulation of data—such as keeping records, filing, and reproducing materials, and organizing written and numerical data according to a prescribed plan—and avoid ambiguous and unstructured undertakings. Conventional people perceive themselves as careful, conforming, orderly, and as having clerical and numerical ability.	The behavioral tendencies in Conventional environments lead to the acquisition of clerical, computational, and business system competencies necessary to meet precise performance standards and to a deficit in artistic competencies. Conventional environments reward people for the display of dependability, conformity, and organizational skills.	Accounting, secretarial studies, data processing, bookkeeper, financial analyst, banker, tax expert, and medical laboratory assistant

From the salient attributes and the environment descriptors, Holland and his colleagues, developed a few resources for matching the personality types to academic majors and occupations. These include *The College Majors Finder* (Holland, Holmberg & Rosen 1989),

The Educational Opportunities Finder (Rosen, Holmberg & Holland 1994), and the *Dictionary of Holland Occupational Codes* (Gottfredson & Holland 1996). A career or academic field can be determined to match multiple types but when this is the case, there is generally a more dominant type that matches best and then a less dominant type with fewer similarities. With this information, individuals can self-sort themselves or be sorted by a career counselor into their best-fitting personality type. From this, they can look at potential career options or fields of study that will best match their personality. Some of these careers and university degree programs can be found in the last column in Table 1.

To explain how the personality types and environments are related there are three key assumptions that Holland proposed which are all supported by "voluminous empirical evidence" (Smart 2006: 12). They are as follows:

1. *The self-selection assumption* is the psychological aspect that assumes that individuals aim to pick occupational and educational environments that match their personality type as this will provide them the opportunity to thrive in these jobs and settings that allow them to partake in tasks that they enjoy doing. In short, individuals want to succeed and tend to pick environments in which they will flourish.
2. *The socialization assumption* is the sociological aspect that supposes those model environments (e.g., workplace or academic major) will encourage the skills and traits in individuals that align with the environment's needs and actively discourage the skills and traits that do not align with the dominant personality in the model environment. Over time, this will cause incongruent personality types to match their environment.
3. *The congruence assumption* "suggests that vocational and educational stability, satisfaction, and achievement are a function of the "fit" or congruence between individuals and their environments" (Smart 2006: 12). Therefore, the more someone is succeeding in terms of fit, the better they feel they belong in their environment.

Further elaborating on the congruence assumption, the six personalities or model environments can be placed on the vertices in a hexagonal diagram, as seen below in Figure 2, in a clockwise manner starting with Realistic and working around with Investigative, Artistic, Social, Enterprising, and Conventional (Holland et al. 1969). As mentioned previously, a personality type will always have the highest level of fit with its matching model environment. For example, the social personality type will thrive in the Social model environment. Furthermore, the personality types will continue to do relatively well in its neighboring model environment. Continuing to follow the example, this means the Social personality will have a "moderately high level of fit" in the Artistic and Enterprising model environments since they

are the neighboring personality types and environments in the hexagon (Smart 2006). Continuing this trend, a person with a Social personality will have a "moderately low level of fit" in the Investigative and Conventional model environments (Smart 2006). Lastly, they will

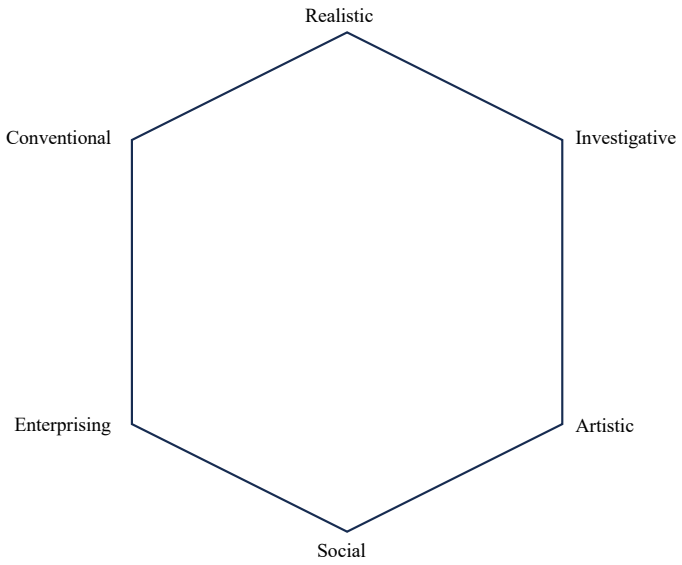


Figure 2. Holland's Congruency Hexagon

have the poorest fit in the Realistic environment since it is at the opposite end of the hexagon and the farthest distance away from their model environment.

Smart (2006: 14) claims that most research, when applying Holland's theory, focuses only on the psychological aspect, thus failing to acknowledge the socialization abilities of academic disciplines. In Smart's (2006: 13) opinion, this is a huge oversight. He has found, with support from multiple studies, that the socialization assumption is the reason why incongruent personality types were found to adapt and eventually thrive due to the hypothesized strong socialization forces when they entered a field of study or career that was not a perfect match with their dominant personality type. Instead of being "doomed to failure" when picking paths, Smart (2006: 17) instead found that "people learn what they are taught" and they are encouraged through reward and punishment (be it social pressure, motivation for good grades, etc.) to conform and match the personality that best matches the environment they are in. Smart (2006: 28) states, "our findings suggest a stronger socialization than psychological dynamic at work in Holland's theory given the consistent and pervasive effects of academic environments on both congruent and incongruent students" even going as far as to say that the socialization assumption can dually account for the success of congruent personality types and incongruent personality types. This means that "the influence of academic environments appears greater than the effects of the individual's own predispositions" (Smart 2006: 33).

Another difference that Smart et al. make to Holland's original theory when working specifically with academic environments is that they exclude the Realistic and Conventional

categories “due to the fact that very few college students and faculty fit into these categories” (Jones 2012: 13). It is not to say that people who have these personality types cannot go to university and cannot thrive in other settings, but simply based on the numbers and academic environments that fit into these two categories, Smart et al. decided to exclude them. Finding a representative sample for these categories from students enrolled in higher education programs would be much more difficult than the remaining four categories. By the time the students reach the master's level in their academic field, if the socialization assumption is applied, the students will be conditioned to belong to the academic discipline and display the main characteristics of the model setting that the discipline belongs to. This is part of the reason why the present study will look at students enrolled in master’s programs. By this point, they should decisively be members of their academic communities and personality environments, especially when the bachelor studies of the students will also be known. If a student completed a bachelor’s program in the same field as their master’s studies, they should firmly be a member of their academic community at that point.

As mentioned earlier, the empirical evidence that supports this claim is sufficient to argue for its use and further investigation in this study. Several studies have looked at differences that exist between the four groups in the context of higher education studies with varying results. The evidence supports the claims that the groups hold significance when considering various factors such as teaching methods, goals and expectations for students, and faculty disputes. The results from the various studies will be highlighted but it is important to keep in mind that the results relating to the Realistic and Conventional types are minimal since these types make up a small minority of academic fields within higher education settings. Therefore, they will not be discussed since they are not the focus of the current study. Each of the four remaining types will be highlighted and the findings that relate to this group (in terms of what they do and do not do in course planning, teaching methods, and departmental procedures) will be discussed.

Investigative faculty members, which include fields such as biology, math, earth sciences, and pharmacy, were found to encourage certain behaviors that align with the investigative environment, such as analytical abilities, mathematical skills, and scientific abilities (Jones 2012: 14). The faculty also ranked students’ ability to analyze qualitative problems as more important than communicative skills and understanding people (Smart et al. 2009: 491). Furthermore, in research dating back to the 70’s the Investigative faculty were found to value structured and formal approaches such as lecturing, had a stronger emphasis on grades, and maintained more formal and distanced relationships with their students when

compared to faculty from the enterprising, artistic, and social types (Morstain & Smart 1976: 1208; Peters 1974: 62). The faculty were additionally found to discourage active and collaborative learning techniques, higher-order learning, and leadership, managerial, and persuasive abilities in comparison to the other types (Umbach 2006: 334; Jones 2012: 14).

Enterprising faculty members, which include fields such as business, law, and journalism, were found to encourage and promote communication-based skills. In general, the faculty members designed courses and rewarded behaviors that related to forming a deeper understanding of people (Smart & Umbach 2007: 189). They also promoted leadership and managerial skills, along with persuasive abilities (Smart & Umbach 2007: 185). Lastly, they focused on communicative abilities that would progress their students' careers more than the other types (Smart & Umbach 2007: 185).

Social faculty members, which include fields such as education, religion studies, and psychology, were found to design courses to help their students understand themselves and others from different backgrounds (Smart & Umbach 2007: 189). Following this same idea, they were also found to include the greatest number of activities based on understanding diversity (Jones 2012: 15). The departments in this category were found to have the highest level of disagreement when it came to tenure votes and decisions regarding promotions (Hearn & Anderson 2002: 510). This is insightful information on some of the inner workings of the departmental function. Activities in the classroom tend to be unstructured and independent for both the professors and students, heavily relying on student-centric group discussions and small group activities (Morstain & Smart 1976: 1208; Peters 1974: 62). Lastly, it was found that faculty in the social fields were most likely to use effective educational practices in general, encourage active collaborative learning, and have close student-faculty relationships when compared to the relationships in the investigative fields (Milem & Umbach 2003: 615; Hearn & Anderson 2002: 510).

Artistic faculty members, which include fields such as architecture, fine arts, music, and foreign languages, were found to have a lot of similarities to the Social faculty members. To a lesser degree, but still significantly, faculty members in the artistic fields were found to employ effective educational practices, encourage active collaborative learning, and have close relationships with their students (Milem & Umbach 2003: 617; Umbach 2006: 338). They also tended to employ unstructured and student-centric approaches in the classroom (Morstain & Smart 1976: 1208; Peters 1974: 62). The faculty was also found to encourage real-world problem-solving skills as well as understanding oneself and others from different backgrounds (Smart & Umbach 2007: 198).

Even though the above-mentioned findings are based on faculty members and their actions; these actions impact their students. Smart and Umbach (2007: 190) concluded from their findings that "the distinctive preferred norms, values, and practices of faculty in the respective academic environments transcend the institutional settings in which they are organizationally housed," showing that the institutional influence is less than the influence of the academic environment. They also found similarities across different institution types across the US, which further supported this conclusion (Smart & Umbach 2007: 190). Additionally, Umbach (2006: 338) concluded that "academic discipline plays an important role in the use of effective instructional practices," thus showing that different teaching practices exist between the four different types.

This is relevant since this further shows that Holland's socialization theory (number two) is the most powerful as opposed to the self-selection assumption (number one) (Smart 2006: 17). Smart (2006: 17) further states that throughout their past studies, including over 2,300 student participants at over 300 colleges and universities, all three assumptions have been validated but there is more evidence supporting the validity of the socialization assumption than the other two. Thus, concluding that "the influence of academic environments appears greater than the effects of the individual's own predispositions (Smart 2006: 33). These conclusions could be made since they found that students with congruent and incongruent personality types were equally socialized by their academic environments (Smart 2006: 16). These findings and "discipline-specific socialization process" were also found to take place at the graduate level (Smart 2006: 14, 19). Smart (2006: 28) combines his findings with his previous work to definitively claim that academic environments, as defined by Holland, are "a primary influence on the longitudinal change and stability" in terms of abilities, interests, and values. Thompson (2003: 422) also reaches a similar conclusion, claiming it is the academic environment where students are encouraged to fit in and belong that grants them the ability to foster and grow the values that align with their academic environment.

2.3.2 Biglan and Kolb Section

The second scheme that was highlighted by Jones (2012: 16) as being relevant in the field of higher education studies is two schemes that have been combined. One was created by Biglan (1973a; 1973b) and the other by Kolb (1981). Biglan and Kolb did not work together on their schemes but there is significant overlap and often in research the two schemes are connected as having the same resulting groups of academic disciplines. Kolb (1981: 243) even cites Biglan in his paper and admits that his "indexes reveal a pattern of relationships among academic fields

that is highly consistent with Biglan's study and the managerial learning style data". For all intents and purposes relevant to the current study at hand, they will be considered as the same approach going forward.

Kolb and Biglan both have two spectra in their models. For Kolb (1981: 236), student learning can be somewhere on the abstract and concrete spectrum as well as on the reflective and active spectrum. The concrete/abstract dimension refers to which degree the subject is based in theory. Subjects focused on "the experiencing of events" are thus labeled as concrete and subjects focused on "conceptualization" are on the abstract end (Kolb 1981: 236). In the reflective/active dimension, the reflective subjects are "observers" and express "general analytic detachment" while active subjects are "actors" and deal with "specific involvement" (Kolb 1981: 237).

For Biglan, academic disciplines and the paradigms that define them can be placed on three independent spectra hard to soft, pure to applied, and life to non-life. Biglan (1973a: 12) found that the life/non-life dimension was the least impactful when explaining and predicting behaviors, so he favors only classifying subjects using the first two dimensions. The Hard/Soft dimension reflects the existence or lack of a central paradigm that rules the subject and the relevant research. Hard subjects are accepting of one central paradigm that is determined to be "correct," meanwhile soft subjects are accepting of multiple theories and different approaches that each have their appropriate applications (Biglan 1973a: 12). The Pure/Applied dimension relates to the applicability of the subject knowledge with pure fields being highly theoretical, and applied fields, as the name might suggest, are seen as directly applicable to real-life situations (Biglan 1973a: 12). Becher (1994: 152) highlights the overlap of the Kolb and Biglan categories below and gives an example for each of the four groupings which are included in Table 2 below.

Table 2. Kolb and Biglan academic disciplines from Becher (1994: 152)

Kolb	**Biglan**	**Example Subject Area**
Abstract Reflective	Hard Pure	Natural Sciences
Concrete Reflective	Soft Pure	Humanities and Social Sciences
Abstract Active	Hard Applied	Science professions
Concrete Active	Soft Applied	Social professions

Biglan is said to have "discovered" these spectra because his research involved asking the opinion of faculty members at institutions of higher education on which disciplines should belong together. He mailed them all small note cards with different academic departments

written on them and asked faculty at a small private university and a large public university in the US to group them in whatever way they saw fit. The faculty members then mailed the groupings back to Biglan and this is where he started finding trends, larger groupings, and assigning features and names to these groups. Over time, there has been a lot of criticism on the validity of this method, but consistently in research since the development of these groupings, they have shown to have a significant impact when considering the behaviors of students and staff. The specifics of the supporting research will be highlighted below.

Kolb developed his groups in an entirely different manner by using the results gathered during the Carnegie Commission on Higher Education, a 1969 study of representative American colleges and universities. The data examined was the result of over 30,000 graduate student questionnaires at 158 institutions and over 60,000 questionnaires from faculty at 303 institutions (Kolb 1981: 241). From his analysis of this data, Kolb determined that the classic separation of scientific and artistic fields may not be precise enough. From this, he suggested his fourfold typology which, as mentioned earlier, has significant overlap with Biglan's resulting paradigms (Kolb 1981: 243) but comes from perhaps a more suitable and reputable dataset.

Even though these two schemes were developed in different ways and used different criteria, both resulted in nearly identical groups. Kolb's Concrete/Abstract dimension matches Biglan's Hard/Soft dimension. Kolb's Reflective/Active dimension matches Biglan's Pure/Applied dimension. Some of the overarching characteristics of each group were provided by Becher (1994). Becher (1994: 152) describes hard/pure subjects as being "cumulative, atomistic, and concerned with universal". Soft/pure subjects are "reiterative, holistic, and concerned with particulars" (Becher 1994: 152). Hard/applied subjects are "purposive, pragmatic, and concerned with mastery of the physical environment" (Becher 1994: 152). Lastly, soft/applied are "functional, utilitarian, and concerned with the enhancement of professional practice" (Becher 1994: 152).

When considering the dimensions, Biglan along with several other researchers (Biglan 1973a; Biglan 1973b; Nelson Laird et al. 2008; Hiller 2021: 19) found that the hard/soft spectrum accounted for the most significant differences between academic groups. While it can be agreed upon that the hard/soft dimension is the most dominant, several studies show further support for the remaining two dimensions, but the connections between these findings are not evident. Doberneck and Schweitzer (2017: 78) make the point that in their research, there were statistically significant disciplinary variations associated with all three of Biglan dimensions. However, in reviewing the studies mentioned in the Jones (2012) report, only soft/hard

dimension will be explored in this literature review since it has received the most amount of attention in research and seems to have the greatest number of findings related to it. After that a dedicated section will explore the relevant findings at non-US based institutions (anglophone and non-anglophone) as they hold an additional pertinence to the present study since all the participants are attending Austrian institutions of higher education. Some of points of inquiry concerning the differences that may exist between soft and hard disciplines include teaching differences, leadership practices, departmental functions, research methods, and career satisfaction. The findings from the studies that have explored these disciplinary differences will be summarized below.

Faculty members from soft disciplines were found to use active learning techniques in their teaching, value diversity, have contact with their students, and have higher expectations of their students (Braxton, Olsen & Simmons 1998: 308). On a similar note, professors from soft disciplines were found to prioritize student development as their main teaching goal and encourage deep learning more than members of hard disciplines (Barnes et al. 2001: 460; Nelson Laird et al. 2006: 18). Furthermore, Colbeck (1998: 665) found that when integrating their own research into their courses, professors were likely to do this through their teaching. Alise (2008: 137) determined that members of soft disciplines were likely to use qualitative methods and mixed methods in their own research instead of quantitative methods. In terms of faculty disputes over promotions and tenure, they were more likely to occur in soft disciplines. Leaders and leadership in soft disciplines tend to be seen as a democratic method with a dislike for "efficient, hard results-oriented" leaders (Kekäle 1999: 233).

Faculty members from hard disciplines were more likely to see grades as a gatekeeping method for students when looking at academic success (Barnes et al. 2001: 461). This is since subject matter facts were seen as the primary teaching goal for faculty members of hard disciplines (Barnes et al. 2001: 462). They were also likely to incorporate their own research when instructing students how to carry out research in their field (Colbeck 1998: 665). Furthermore, members of the hard disciplines preferred leadership that made decisions based on measurable features along with following the logic of linear thinking when forming conclusions. Kidwell and Kidwell (2008: 149) additionally found that staff members in business schools who belonged to harder sub-disciplines were more likely to view potentially controversial activities as less ethical and did not see a purpose in including these types of activities in their teaching.

Several items were also investigated through research and found to not be impacted by academic discipline. One such item is faculty decisions to leave the profession wherein

disciplinary differences were determined to have little impact on this choice (Barnes, Agago & Coombs 1998: 460). It was also found that "moral turpitude and lack of attention to course planning were inviolable norms across disciplines," which means that for all disciplines, these are seen as unacceptable behaviors by members of the faculty (Braxton & Bayer 1999: 81). Additionally, little difference was found in professors' attention to providing timely feedback, encouraging cooperation, and emphasizing time on task when disciplines were compared (Barnes et al. 2001: 462).

Favero (2006) explored dean behaviors between the hard and soft disciplines as well as between the applied and pure disciplines. From this work, she determined that the behavior of deans was impacted by which discipline they belonged to. In applied disciplines, deans were found to consider problems from multiple perspectives and use these perspectives to inform their decisions instead of only considering one viewpoint. Favero (2006: 298) highlights that this finding is consistent with Becher's (1989) conclusion and states that the "predisposition to understand others might influence a tendency to more readily explore alternative perspectives in administrative decision-making" for deans in applied fields, thus highlighting the potential impact of academic discipline on administrators' behaviors. It was also found that deans in the pure fields were "more likely to remain more rooted to their disciplines, even as deans" (Favero 2006: 304).

Lastly, several studies were conducted to see how well Biglan's groups held up at non-US-based institutions of higher education. As mentioned earlier, these findings are especially relevant since the present study takes place in Austria. There was only one study conducted at a non-Anglophone university. In this study, Ylijoki (2000: 350) investigated the core beliefs, values, and norms of an academic department at a Finnish university and concluded that there were differences due to academic discipline grouping. Some of these differences include the faulty view of professional orientation, the reference point of study, professional aims, and primary activities (Ylijoki 2000: 350). In the examination of the culture of the sociology and social psychology department, the "[t]heoretical orientation, individualism, and dedication to one's discipline" we found to match the logic of the soft-pure domain (Ylijoki 2000:356). Meanwhile, the subjects selected to represent the applied domain were found to indeed have a professional orientation, as would be expected from applied disciplines (Ylijoki 2000: 356). The ways in which the individual subjects expressed this orientation towards the work world differed, as the hard subjects "aim at control and new applications" meanwhile the soft disciplines aim at "serving and developing the practice" (Ylijoki 2000: 356). Furthermore, these

faculty differences were found to change what the students in the different categories viewed as the benefits and downsides to studying in general.

Several more studies were carried out at non-US but still Anglophone universities. While these are also non-US-based institutions they might have more in common with US institutions than HEIs in non-Anglophone countries. Ballantyne, Bain, and Packer (1999: 237) looked at teaching practices present at Australian universities. They found that the results followed what was already known about the academic disciplines from other studies (Ballantyne, Bain & Packer 1999: 242). The specifics of their findings stated that in general, lectures were the most common teaching method. In soft disciplines, they used more cooperative and collaborative teaching methods while including more tutorials in the curriculum (Ballantyne, Bain & Packer 1999: 242). In hard disciplines, they tended to use more computer-based techniques and more "practical sessions" (Ballantyne, Bain & Packer 1999: 242). When looking at 697 faculty members based at four UK universities, Norton et al. (2005: 540) concluded that there was a significant difference between faculty members when considering teaching intentions. They determined that these differences were due to different conceptions of teaching and found that members of hard disciplines did not believe in interactive teaching techniques but did believe in training students for jobs. Lastly, Simpson (2017: 1520) investigated how well Biglan's academic groupings would fare at UK universities. Seemingly doubtful in the initial goal, Simpson (2017: 1525) concludes that the dimensions outlined by Biglan "reemerge with high accuracy as the critical dimensions". Furthermore, the reluctant acceptance of the results continues when Simpson (2017: 1529) states, "[t]his analysis suggests, surprisingly, a classification scheme developed forty years ago from some stapled sheets of paper still represents a powerful underlying organizational mechanism in higher education". This quote highlights the possibility that the scheme might be able to be used in a different setting, including a new country and a completely new time, and still potentially maintain its relevance (Simpson 2017: 1528).

2.3.3 Biglan & Holland Approaches Compared

Not many studies have compared the effectiveness of Smart et al.'s approach to Biglan's approach within the same study. Biglan's approach has received a lot of criticism for its methodology in creating its group, as hinted at previously by Simpson (2017: 1530) when Biglan's methodology was boiled down to consisting of "some stapled sheets of paper". Kolb's groups were determined to be similar to Biglan's, but his methodology has not been as widely criticized. Smart et al.'s methodology in general seems to be more widely accepted with less

criticism. However, beyond the methodology, the results that have compared the two approaches reflect a preference for Biglan's approach. For example, Hiller (2021: 2), when comparing the two approaches' ability to explain the disciplinary differences of the emphasis on deep approaches to learning, found that "while showing some overlap, […] Biglan's dimensions explain slightly more variation, with a slightly greater range of magnitude in some effect sizes," hence showing the preference in the research community toward Biglan's approach. Some of the overlaps that were found were attributed to the overlapping of the groups. Hiller (2021: 20–21) illustrates this claim by pointing out that Holland's Realistic disciplines are predominantly categorized into Biglan's hard-applied-nonlife disciplines, while Holland's Enterprising disciplines appeared in Biglan's soft-applied-nonlife disciplines. Some of the results, though, were different for groups that tended to consist of similar faculty members which Hiller (2021: 20–21) credits to "the complexities within these groupings" and further states that these apparent inconsistencies illustrate the need for further exploration of the "nuances in values, interests, and practices amongst faculty within the same Biglan or Holland categories".

Even though the effectiveness of both schemes is not yet fully known, Hiller (2021: 18-19) asserts that once "the vast complexity of faculty teaching in different contexts" has been fully taken into consideration "the ability of both Biglan and Holland conceptualizations to explain even moderate levels of variance is significant". Hutchings & Garraway (2010: 4) further warn that by ignoring the role of disciplines, researchers are at risk of overlooking important nuances between the groups. The research outlined in this section touches on some of the findings relevant to Smart et al. and Holland's approach, along with Biglan and Kolb's approach to academic disciplines. It is clear from this brief overview that the impact of academic disciplines, when grouped by larger themes and considered as members of a family with resemblance to other members instead of belonging to strict and bounded groups, is often investigated when studying higher education programs. Thus, meriting the further use of these disciplinary groupings in the current study to further the investigation of their usefulness.

2.4. Vocabulary Acquisition

2.4.1 Academic Disciplines and Academic Discourse

The present study relates academic disciplines to the learning of academic English, specifically vocabulary. One source of inspiration for this point of inquiry was a quote that reads: "[g]iven the potential for the disciplinary difference to influence teaching and the clear distinctions in disciplinary practices [...] there is likely to be disciplinary specificity in the approaches to

learning, including the learning of academic English" (Lau & Gardner 2019: 258). The resulting impact on vocabulary acquisition of academic English by students enrolled in masters' programs taught in English will be investigated in the empirical part of this study. Durrant (2014: 354) points out that "even novice students of EAP start out with large amounts of vocabulary knowledge; knowledge, which is likely to differ from student to student, depending on their educational background, first language, and personal preferences and interests," thus emphasizing the potential impact of academic background.

Fostering disciplinary literacy is cited as the goal of any degree program (Kuteeva & Airey 2014: 539). With this goal, disciplinary literacy is defined as "the ability to appropriately participate in the communicative practices of a discipline," and is made up of academic, societal, and vocational literacies which are further split into local and international forms (Airey 2011: 2). These local forms tend to take place in local languages, meanwhile the international forms usually take place in English. When discussing "one-size-fits-all" language policy for higher education in Sweden, it is argued that a key issue with these types of policies is that they do not consider "fundamental disciplinary differences and their potential impact on language use" (Kuteeva & Airey 2014: 533). With this failure to consider disciplinary differences, there is a lack of nuance in the policies and all disciplines must conform to norms that may or may not be beneficial for their disciplinary needs.

Sometimes, as Hutchings & Garraway (2010: 10) point out, academics are so embedded in their disciplines, that they might not even be able to identify and notice the norms and practices of their discipline. However, they continue by explaining that professors' and instructors' expectations for their students are "situated entirely within these very norms and practices" (Hutchings & Garraway 2010: 10). This leads to an interesting situation where there is an expectation to perform by professors and agents higher in the hierarchy, meanwhile, they are potentially unaware of the extent of the norms and practices that they are expecting. This is especially true for literary practices due to "the discourses that unconsciously construct each educator's notions of academic literacy," again highlighting the unconscious nature of these expectations and notions (Hutchings & Garraway 2010: 10). Thus, the discourses that are produced because of this are "socially constructed by the discipline's members and 'intimately related to the rhetorical processes underlying the reading and writing of texts'" (Geisler 1994: 221 cited in Hutchings & Garraway 2010: 12). Due to this relationship between discourse, the unconscious notions of academic literacy, and the expectations that are constructed within each discipline and differ among disciplines, vocabulary use, and vocabulary knowledge became an interesting point of inquiry. With an assessment of the vocabulary knowledge of university-

level students, who are just becoming members of their disciplines and still learning the norms, some of the nuances between the disciplines may be revealed. Furthermore, how vocabulary plays an undeniable role in discourse, which is undeniably shaped by academic disciplines when in the context of higher education and academic literacy, is crucial and a key aspect to starting to have a deep understanding of students' abilities. To continue to see student growth and improvement, it is key to know where their abilities lay at that moment and use that to develop instruction and feedback.

This is especially relevant since it has been found that academic vocabulary is one point where students consistently struggle (Wang, Liang & Ge 2008: 443). Nation (2001: 11) divides vocabulary into four levels which include "high-frequency words, academic vocabulary, technical vocabulary, and low-frequency words". Each classification of vocabulary should be taught differently which reflects the differences between the types. Nation (2008: 13) explains in his research that high-frequency words should be taught directly with the use of vocabulary activities, direct teaching of the words, pointing out the words in intensive reading, and using them to develop further vocabulary learning strategies. Low-frequency words, on the other hand, should not be explicitly taught, since they do not offer a lot of reward for the effort it can take to learn a new word to the level of being able to understand it and use it spontaneously (Nation 2008: 14). Instead of spending time teaching these words in class, teachers should encourage students to use the vocabulary learning strategies such as guessing from context, learning words from cards, using word parts, and using a dictionary (Nation 2008: 14). Learners should not put much effort into acquiring the low-frequency words that are not relevant to the contexts they find themselves interacting with, but they should focus on mastering the strategies to learning low frequency words. This will be particularly relevant for the present study since if students show a potential lack of high-frequency word knowledge, those could be taught in a classroom setting, but if the participants show a tendency towards missing knowledge of low-frequency words, this should be developed through learning vocabulary learning skills and additional independent work. This can have potential implications for future steps when considering developing goals around building vocabulary knowledge in English for Austrian students.

Technical terms are words that are extremely contextually contingent with large variability for the subject matter (Wang, Liang & Ge 2008: 443). Technical terms tend not to pose a large problem for students, as they are "the focus of the discussion in the classroom or are glossed in the textbook," meaning that time is spent by teachers to be assured that the students know these words specifically as they are often tied to the ability for students to be

able to perform in their subjects and they tend to be mentioned and defined in student textbooks (Strevens 1973: 226). Technical words might be names for specific tools, devices, and processes that people outside of a specific subject may have never heard before (i.e., *altimeter, aileron, rudder,* and *camber* in aviation or *neurohypophysis, neuroma, bradycardia,* and *craniopharyngioma* for neurosurgeons).

Academic vocabulary is also known as non-subject-specific semi-technical vocabulary, sub-technical vocabulary, and semi-technical vocabulary, and is seen as a class of words that has both "technical and non-technical implications" since the words seem to exist in this gray space between technical and non-technical words (Wang, Liang & Ge 2008: 451). Academic words were found to be the words that students struggle with the most out of the four different types (Wang, Liang & Ge 2008: 448). Anderson and Freebody (1981: 82) also found that these words were the most likely to be identified by students as unknown when reading academic texts. Farrell (1990: 11) explains that academic words are "formal, context-independent words with a high frequency and/or wide range of occurrence across scientific disciplines, not usually found in basic general English courses; words with high frequency across scientific disciplines". Even if learning these words has shown to be particularly difficult, it is vital to know these words to understand academic texts.

Laufer (1992: 126), as well as Laufer & Ravenhorst-Kalovski Field (2010: 24), found that for general comprehension of a text, L2 readers needed to understand somewhere between 95-98% of the words present. If these thresholds were not met, the reader would spend too much time focusing on the unknown words to really have a deeper comprehension of the running text. The more words a reader knows, the more of the text they will understand. In L1 studies, comprehension was not measured but the difficulty was. It was determined that an "easy" text contains approximately 0% unknown words, meanwhile, a "difficult" text contains approximately 2% unknown words (van Zeeland & Schmitt 2013: 458). That means the difference between an easy and a difficult text is only a few words per paragraph. While this does not exactly measure comprehension, L2 studies have determined a few additional key milestones. Hsueh-Chao & Nation (2000: 411) replaced random words in a text with nonsense words and found with 80% understandable words and 20% nonsensical words, none of the participants were able to comprehend the text. At 90 and 95% understandable words a few of the readers could comprehend (Hsueh-Chao & Nation 2000: 422). At 100%, a majority of the readers could comprehend the texts (van Zeeland & Schmitt 2013: 459). This led to the idea that the ideal lexical coverage rate, or the percentage of words known by the reader to ensure comprehension, was somewhere between 95 and 100%. After this study, Schmitt, Jiang, et al.

(2011: 35), discovered a linear relationship between coverage level and degree of comprehension; however, on average, 98% coverage led to 70% comprehension of the texts and 95% coverage led to 60% comprehension. Lastly, as mentioned earlier, Laufer & Ravenhorst-Kalovski (2010: 26) determined that two cutoff points are relevant for L2 studies: 98% is seen as "optimal" and 95% as the minimal (van Zeeland & Schmitt 2013: 459). In their terms, 95% leads to "acceptable comprehension" and 98% leads to "successful comprehension by most learners" (Laufer & Ravenhorst-Kalovski 2010: 26).

Coxhead (2000: 213) determined that 10% of words in her academic corpus were academic in nature. Santos (2010: 20) determined separately that academic words make up approximately 16% of the running words in academic textbook samples across different disciplines. Both figures are above the 2-5% of words that have been determined to be the percentage of unknown words a reader can cope with through guessing and context clues and still have sufficient comprehension (Wang, Liang & Ge 2008: 443). If a reader wants any chance at comprehending and succeeding in understanding academic texts, they need to have adequate knowledge of the academic vocabulary present in these texts since it makes up a significant percentage (>5%) of the running words in the texts (Wang, Liang & Ge 2008: 443).

2.4.2 Word Knowledge

When considering vocabulary acquisition in non-native speakers (NNS), a valid benchmark to relativize the discussion is to first look at the vocabulary knowledge of native speakers (NS). As will be explained later, this is not in hopes of setting a "goal" or "target" for NNS; however, it will serve as a number to look at and can help put things into perspective. To begin to do this, first looking at the number of total words in English is helpful, not as a realistic learning goal but to put the numbers into context and understand what percentage of words in a language the average user knows.

The estimates for how many words exist in English vary drastically. These estimates depend on the unit of counting for words, which will all be explained and detailed later in this section. One explanation in the wide range of estimates is due to varying answers to questions such as if proper nouns should be counted, if different versions of the same word should be counted (ex. *dog* and *dogs*), and if words that are related morphologically and in meaning should be counted as one word (ex. *explorer* and *exploring*).

Depending on the answers to questions like these, the estimates for the number of words in English range anywhere from tens of thousands to ten million words (Schmitt & Schmitt

2020: 8–9). Further studies estimate that the range of the number of words in English is somewhere between 88,500- and 114,000-word families (Nation 2011: 59–60). One certain thing is that "it is impossible to say precisely how many words there are in English due to the ever-changing nature of language and counting methods" (Schmitt & Schmitt 2020: 9). Furthermore, another conclusion that can be made from this information is that it is unrealistic to imagine that any speaker of English (native and nonnative alike) would be able to learn all these words to a meaningful level (Nation 2011: 60; Goulden et al. 1990; Nagy & Anderson 1984). Second language (L2) learners and first language (L1) learners instead will learn a small portion of these words in their lifetimes.

To help standardize the method for counting words, Bauer and Nation (1993) created a word family scheme to explain the levels of defining and grouping words into their clusters, or families. This scale "is a set of levels based on frequency, productivity, predictability, and regularity of affixes" (Dang, Coxhead & Webb 2017: 970). Word estimates can then be based on counting methods using one of these levels as the defining level for the word families. For example, a word list can suggest a book contains 20,000 words in it at a Level 4, meanwhile, with a Level 2, the word count is much higher at around 30,000. This scheme is summarized below in Table 3. In this scheme Level 1 is the most restrictive and states that each word form is a different word (i.e., *dog* and *dogs* would be counted as two words). The following levels always include the previous levels in their definitions of word families, moving from the most transparent forms to the least transparent members of word families (Dang, Coxhead & Webb 2017: 970). Level 7 is the most wide-ranging and inclusive in the scheme including classical roots and affixes as well as all the regular and irregular affixes that are included in the previous levels.

Table 3. Bauer and Nation's (1993) Word Family Scheme

Level	Description	Number of affixes	Examples of affixes	Example of forms
1	Each form is a different word			
2	Inflectional suffixes	8	*-ed*, *-s*	*heat-ed* *waste-s*
3	The most frequent and regular derivational affixes	10	*-able*, *-er*, *non-* (each without restricted uses)	*heat-er*
4	Frequent, orthographically regular affixes	11	*-al, -ful*, *in-* (each with restricted uses)	*waste-ful*
5	Regular but infrequent affixes	50	*-age*, *-ally*, *ante-*	*wast-age*
6	Frequent but irregular affixes	12	*-ee*, *-ic*, *pre-*	*pre-heat*
7	Classical roots and affixes			

Which level to use and the implications of using different levels to define a word are debated. On the extreme account, Level 1, it is assumed that learners cannot understand that words like *dog* and *dogs* are related in meaning and the *-s* suffix in this case is just indicating plural *dogs*. For many, this feels like too extreme of an approach as it has been shown that even beginning learners have some basic understanding of morphological transformations in English independent of their L1 backgrounds and learning contexts (Dang, Coxhead & Webb 2017: 972). On the other end of the scale, level 7 assumes that learners know all forms of the same stem or base word that are included in a word family. For example, that implies that if someone knows the word *formula* then they also know the meaning of *formulae, formulas, formulate, formulated, formulating, formulation, formulations, reformulate, reformulated, reformulating, reformulation,* and *reformulations*. Many researchers argue that this is too generous of a counting system for learners of English and native speakers cannot even be assumed to know all these meanings (Hyland & Tse 2007: 236).

This method of creating levels also solves a commonly debated topic of using lemmas or word families as counting units. Nation (2016) points out that lemmas are defined in his word family scheme as they correspond to a Level 2 word family on the list. He continues to make the argument that "the question is not whether lemmas or word families are better, but which word family level is the most suitable for a particular group of learners" (Nation 2013, 2016; Nation & Webb, 2011). Other word forms (ex. two meanings of *produce*) are only considered to be part of the same word family if the meaning of the base in the derived word is closely related to the meaning of the base when it stands alone or is combined with other derived forms. Since this method allows for a more flexible approach, some vocabulary learning lists are presented in multiple-word family levels, depending on the target audience and purpose. When users are more advanced, a higher word family can be used to avoid heavy repetition in the teaching and learning (Dang, Coxhead & Webb 2017: 971). L2 learners' knowledge of affixes and word families increases as they advance their knowledge of English and "instruction about word parts helps to expand learners' vocabulary knowledge" (Dang, Coxhead & Webb 2017: 972).

This word family theory starts to touch on the idea of the extent to which words can be known. Word knowledge and what it means to really know a word can vary drastically dependent on the context. Daller et al. (2007: 8), refer to these types of word knowledge as the lexical space, or the 3-D space that includes depth, breadth, and fluency of word knowledge. These factors highlight the number of words a user knows (breadth), the extent to which the user knows these words (depth), and the ease at which these words can be accessed (fluency).

As was done in Busby's (2021: 61) study of vocabulary knowledge in university students, vocabulary size will be investigated "since this is the aspect most often discussed when providing a quantitative picture of vocabulary used for recognition and understanding".

The extent (or depth) to which vocabulary can be known can range from the most basic form of simply recognizing a word to the most complex knowledge including common collocations and idioms using the word. Nation (2016: 49) again summarized how a word can be known into 18 aspects that cover three main areas: knowledge of form, knowledge of meaning, and knowledge of use. Each of these three areas are split further into receptive and productive knowledge. The form category is split into knowledge of spoken, written, and word parts. The meaning category is split into knowledge of the form & meaning, concepts & referents, and associations. Lastly, the use category is split into the knowledge of grammatical functions, collocations, and constraints on use. This makes 18 categories, 9 of which involve passive or receptive knowledge, and the other 9 involve productive or active knowledge. In general, the receptive knowledge of words is much higher than the productive knowledge of words, and "the ratio between the two aspects of vocabulary knowledge can be anywhere between 50% and 70%" (Nizonkiza 2016: 169). The 18 aspects can be found in Table 4 below along with a question a user could ask themselves to determine if they possess that aspect of the word's knowledge or not.

Table 4. Nation's 18 aspects of word knowledge (2016: 49)

Form	Spoken	R	What does the word sound like?
		P	How is the word pronounced?
	Written	R	What does the word look like?
		P	How is the word written and spelled?
	Word parts	R	What parts are recognizable in this word?
		P	What word parts are needed to express this meaning?
Meaning	Form and meaning	R	What meaning does this word form signal?
		P	What word form can be used to express this meaning?
	Concepts and referents	R	What is included in the concept?
		P	What items can the concepts refer to?
	Associations	R	What other words does this make us think of?
		P	What other words could we use instead of this one?
Use	Grammatical functions	R	In what patterns does the word occur?
		P	In what patterns must we use this word?
	Collocations	R	What words or types of words occur with this one?
		P	What words or types of words must we use with this one?

Constraints on use (register, frequency, …)	R	Where, where and how often would we expect to meet this word?
	P	Where, where and how often can we use this word?

With multiple estimates on how many words exist in English, along with the many possibilities of how an individual can know a word, some numbers can be presented in terms of estimates of the number of words for both NS and NNS. It is estimated that the average L1 speaker of English "knows broadly 10,000 - 13,000-word families" but the extent to which these word families are known is unspecified (Schmitt & Schmitt 2020: 11). This means, it is unknown to which level mentioned above these estimates corresponds. As Nation (Read 2000: 26–27) states, this is an "idealized account, rather than a realistic description of what native speakers know". A different estimate states that "well-educated native speakers know around 20,000 words families" with an estimate that they learn about 1,000 words a year until the age of 20 (Nation 2011: 60). Again, the extent to which these words are known is still not mentioned. It has also been observed that non-native speakers who go to school in an English-speaking environment seem to display the same vocabulary growth trends and sizes as their native-speaking counterparts (Nation & Webb 2010: 196). While these figures present quite a wide range of vocabulary size, what can be said is that "vocabulary acquisition occurs at an impressively fast rate from childhood throughout the years of formal education and at a slower pace on into adult life" (Read 2000: 43).

It has also been found for nonnative speakers of English that knowledge of 8,000-9,000 word families would equate to approximately 98% coverage of written texts and knowledge of 6,000-7,000 word families would be sufficient for 98% coverage of spoken texts (Nation 2006: 79). These estimates do not include knowledge of proper nouns. It is important to point out that the 98% lexical coverage rate was only investigated for written comprehension and spoken language may require a higher percentage to "cope effectively with the transitory nature of spoken language" (Nation 2006: 79). Another study estimated that knowledge of 8,000-word families including proper nouns would be sufficient for 98% lexical coverage of written texts and only 4,000-5,000 word families for 95% lexical coverage (Laufer & Ravenhorst-Kalovski 2010: 26).

It was also determined that "L2 learners with a knowledge of the most frequent 10,000 word families in English can be considered to have a wide variety" (Schmitt, Schmitt & Clapham 2001: 56). When these words are not known, Hazenberg and Hulstun (1996: 158) found that "individuals with a vocabulary of fewer than 10,000 base words run a serious risk of not attaining the reading comprehension level required for entering university studies". They

also emphasize that due to the potentially challenging content of academic texts in terms of new concepts and new information, vocabulary coverage might be even more important than in everyday life when it comes specifically to academic texts (Hazenberg & Hulstun 1996: 158). Some estimates even put the figure between 15,000 - 20,000 word families to undertake L2 education or to succeed in professional settings but more specific details have not been investigated related to these higher figures (Busby 2021: 62). For the present study, a vocabulary size of 10,000 word families will be deemed sufficient for academic reading. Some accounts place the number of words required for advanced degrees at a lower figure (i.e. around 8,000 – 9,000 word-families) but in the present study, a higher level will be required simply due to the fact that the vocabulary tests that will be used do not cover the 8,000 and 9,000 word family level (it covers the 5,000 and 10,000 word family levels). Furthermore, since receptive vocabulary tests have been known to over-estimate the abilities of participants in some cases as much as 20%, taking a higher level as a sufficient vocabulary knowledge level, will assure that this over estimation does not have a large impact on the results of the present study (Nation 2006: 60; Kremmel & Schmitt 2016: 389).

While a larger vocabulary does not guarantee an immediate path to success, it does offer many advantages to English users. What can be said is that in first and second-language research, vocabulary knowledge is deemed to be of great significance in language competence (Laufer & Nation 1999: 33). Additionally, multiple researchers have found that vocabulary knowledge is "one of the best predictors of successful L2 academic reading," as well as for general reading abilities, reading comprehension, and listening abilities (Busby 2021: 59).

More specifically, the relationship between receptive vocabulary and L2 reading comprehension has been shown to strongly correlate with correlation scores in different studies ranging from 0.4 - 0.85 (van Zeeland & Schmitt 2013: 458–459). This shows that vocabulary knowledge is a key factor in comprehension and without sufficient vocabulary knowledge, there is little chance of comprehension. It is also "essential for learners to master the words that they are likely to encounter often in a wide range of academic written and spoken texts" if they have hopes to succeed in academic studies and comprehend the texts they are given, to the extent that is acquired in academic reading (Dang 2018b: 69). It has also been said that to succeed in higher levels of education, one invaluable key to success is a thorough knowledge of academic vocabulary (Coxhead 2000: 230).

It has also been determined that "the breadth and depth of a student's vocabulary will have a direct influence upon the descriptiveness, accuracy and quality of his or her writing" (Wang, Liang & Ge 2008: 442). This shows the direct impact that specific vocabulary knowledge can

have on an individual's success. Even though it is more difficult to prove, L2 proficiency has also been linked to the ability to learn new L2 vocabulary (Busby 2021: 61). Lastly, Grabe (1988:63) has even gone as far as to say that for L2 readers "the lack of a massive receptive vocabulary that is rapidly, accurately and automatically processed [...] may be the greatest single impediment to fluent reading"; therefore, the indispensability of adequate vocabulary knowledge is irrefutable when aiming for successful academic participation. Stær (2008: 148) states that "a regression analysis suggested that 72% of the variance in the ability to score above average in the reading test could be explained by the variance in the vocabulary score" when he looked at the L2 reading abilities of Danish high school students. Academic vocabulary is especially vital since academic texts contain such a large percentage of academic words.

2.5. Academic English and Academic Word Lists

Following the assumption that L1 and L2 users alike will never practically be able to master all the words in a language, the question then becomes which words an individual should learn to maximize knowledge and minimize effort. This is reflective of the "cost-benefit principle" which emphasizes prioritizing learning certain words over others, most commonly the most frequently encountered word families over the less frequently encountered word families. Therefore, at the beginning, learners should focus efforts on the most frequent words and then as they improve, move their focus to less frequent words (Nation & Webb 2010: 132). Following this principle, it was determined that "the top 2000 words make up around 80% of typical English texts" (Schmitt & Schmitt 2020: 6). If beginner learners focus their efforts on these words, instead of highly specified words they will seldomly come across outside of specialized contexts, they will have a jumpstart on understanding large proportions of texts and therefore be motivated by their success. Multiple factors have been determined to impact the learnability of individual words. It has been determined that more frequent words, words with fewer syllables, and concrete words were most likely to be known by a wide variety of learners (Read 2000: 79). Furthermore, words that are hard to pronounce tend to be harder to learn (Read 2000: 40). It is important to note that pronounceability is an individual factor as other known languages and learners' L1s (or first languages) impact the ease of pronunciation and help determine the linguistic distance from other concepts they are familiar with. Lastly, nouns are easiest to learn since "learners can form mental images of them more readily," followed by adjectives and then verbs and adverbs rounding out the list (Read 2000: 40).

To further help with vocabulary acquisition and to teach some of these less frequent word families, researchers and teachers have found it useful to create word lists with specific words for users to focus on. Instead of learning words in order from most frequently used to least frequently used from general corpora, they try to guess which words will be most frequently encountered by the learners in specific contexts most relevant to their lives by using specific corpora. These word lists can serve multiple purposes and range from general knowledge to very specific subject-based lists. Nation (2001: 11) defines the first two-thousand word families as "high-frequency words" of the language and they can be useful for all contexts. Any knowledge outside of these most frequent word families is dependent upon the individual and how applicable an individual word family may be to the individual's life.

Special purposes word lists aim to serve a specific purpose, or in other words these more specified knowledge domains are relevant only to some individual's lives. Some such word lists include academic word lists, engineering word lists, computer science word lists, as well as other word lists for more special uses. When using these lists, it is important to keep in mind that "compared to other words, a special purposes vocabulary has words of limited range and higher frequency" (Nation & Coxhead 2013). One such list, the Academic Word List, offers "a coverage of around 10%" in academic texts but only "a coverage of less than 2%" in novels (Nation & Coxhead 2013). Determining which lists to use, which lists will benefit the learners most, and users' current level is "part of the curriculum design process called needs analysis" (Nation & Macalister 2010) and something the instructors can use to meet the needs of learners and provide clear goals for language courses (Macalister & Nation 2011). The creation of word lists and their use in classrooms is "an attempt to apply the very useful teaching principle that work on today's text should have the goal of making tomorrow's text easier" (Nation & Coxhead 2013).

In addition to learning the most frequent words, another method to maximize learning efforts which was mentioned earlier includes breaking up vocabulary into topics and smaller more relevant groups. By focusing on the topics most relevant to a learner's life and experiences, the individual or their teachers, can tailor the words they learn to be more beneficial for contexts they believe they will come across. For students partaking in EME, academic vocabulary is a very practical and applicable subset of words to focus on. Since students spend a lot of time interacting with and producing academic texts, it makes the most sense if they familiarize themselves with the words, they will most likely encounter. Furthermore, when it comes time to produce these texts, knowing the vocabulary that is in use, as well as the appropriate terms for certain situations, will be one factor in their success when

creating academic texts. Other groups of English users will find other vocabulary groups to be more beneficial for their life experiences, such as medical professionals learning every day and medical terms for certain medical phenomena. Hotel and hospitality employees will also find certain words and vocabulary terms to be more relevant to their careers than perhaps certain academic or medical terms. Certain subgroups will find it beneficial to learn certain words outside of the typical frequency order as they apply more directly to their careers or lives.

The next section focuses on continuing to define academic English as it pertains to academic settings with non-native speakers with some research even questioning its existence at all. Additionally, the academic word lists that have been created for written contexts will be summarized along with a short look into spoken academic word lists due to their relevance to the present study when addressing academic discipline differences in English vocabulary use.

2.5.1 Academic English

As already highlighted, academic vocabulary has been defined in multiple ways. What is consistent is that academic vocabulary is related to the discourse centered around and created in university and higher education settings, involving texts produced by students, and researchers, and present in textbooks, classroom settings, and academic journals to name a few examples. Additionally, words belonging to this category do not share “any intrinsic values" but instead they are found by determining which words are used most frequently in academic discourses compared to more general discourses (Schmitt & Schmitt 2020: 116). While academic words are not necessarily high or low-frequency words (Schmitt, Schmitt & Clapham 2001: 56), what can be said is that understanding Academic English is an area of concern for students (Coxhead 2017: 89). Maxwell (2013:14) also emphasizes that academic vocabulary is not only a challenge for L2 learners, but instead “nobody is a native speaker of Academic English, " emphasizing the fact that Academic English is challenging for everyone and worthy of special attention from L1 and L2 users (Coxhead 2017: 97).

2.5.2 Academic Word Lists (AWL and others)

Coxhead, a former student of Nation and expert in her own right, compiled a list of academic vocabulary words to create the Academic Word List (AWL). She did this by creating a corpus of academic texts sourced from a variety of places “including but not limited to university textbooks, laboratory manuals, academic journal articles, and other corpora,” (Coxhead 2000: 219) which is now referred to as the Academic Corpus. The initial corpus included 3.5 million

words from written academic texts (Coxhead 2000: 219). From there she narrowed the corpus down to a list of 570 word-families that met her three specific criteria. The first was that the words could not appear in the 2000 most common words on the General Service List (GSL), thus making the words a part of what she calls having a "specialized occurrence". Secondly, since the corpus was compiled using texts from 28 subject areas and split into 4 main sections, a word had to "occur at least 10 times in each of the sections and in 15 or more of the subject areas" (Coxhead 2000: 222). Lastly, the word needed to occur at least 100 times in the corpus (Coxhead 2000: 222). These criteria ensured that the words were not also general-purpose words, ensured they were "generally useful academic words," and made sure that the words were "common enough in academic text to make it worth learning" (Nation & Coxhead 2013). For this list, "all word families were defined as the stem plus all closely related affixed forms" corresponding with Level 6 of the scale previously mentioned by Bauer & Nation (1993) in Table 3 since this is the "most common unit of counting in specialized wordlists" (Dang 2018a: 53).

With these criteria, the Academic Word List was created and since this point, has been used for a variety of purposes involved in the teaching, assessing, and learning of academic words (Coxhead 2011: 357). Townsend & Collins (2009: 1003) found in their intervention study that by focusing efforts on learning the words present on the AWL, students experienced better comprehension of academic written texts and higher academic achievement. In a survey of teachers about their feelings towards the AWL, it was found that the AWL has been "widely used and perceived as a useful instrument for L2 learners across a wide range of subjects by EAP teachers" (Dang, Coxhead & Webb 2017: 961).

2.5.3 Criticisms of the AWL

While the AWL has become a widely accepted tool, it has faced its fair share of criticism. One of the main points of criticism of the AWL was raised by Hyland & Tse (2007) and Martínez et al. (2009). The main claim of this point of criticism is that they believe a singular academic discourse does not exist. They instead argue that there are subject-based academic discourses with varying uses for vocabulary present in different settings (Hyland & Tse 2007: 251). Hyland (2002: 390) refers to the concept as "specificity" and states that "disciplines and professions are largely created and maintained through their discourses" (Lau & Gardner 2019: 258). The main conclusion by the AWL critics from this point of criticism is that a general academic word list is not useful for students to learn from, and instead, they should be learning from subject-specific word lists.

The response to this argument is structured in four main parts. First and foremost, Ming-Tzu and Nation (2004) investigated the claim that words in the AWL change meaning based on the different sub-divisions in the Academic Corpus. They did this by looking at homography, which is the idea that one word form can have multiple meanings. Hyland and Tse (2004: 309) make the argument that words appear on the AWL not because they are equally common across all disciplines, but because they are more common in some disciplines than others and that the same word has multiple meanings and in disciplines can represent different concepts. They give the example of *project* existing as "a piece of work" and "to go forward" which are not related in meaning or origin even though they share the same base word (Hyland & Tse 2007: 295). Ming-Tzu and Nation (2004: 309) investigated this claim but in the end were able to conclude decisively that "words in the AWL do not significantly change their meaning" in the sub-divisions. Additionally, they concluded that there was very little homonymy (Nation 2011: 534) and that any "specialist meanings" did not impact or influence which words were added to the AWL or not (Ming-Tzu & Nation 2004: 309). This does not mean that they denied the fact that there were words with multiple meanings present on the list, but just that this did not have an impact on which words appeared on the list in the first place.

Nation and Coxhead (2013) also further elaborated upon this argument later by stating that the core meaning of a word and its discipline-specific meanings are to be seen together, as one entry in a ditionary with multiple definitions. Crossley, Salsbury, & McNamara (2010: 574) support this by pointing out that knowledge of the core meaning of a word provides support for the acquisition of further meanings. Dang, Coxhead, and Webb (2017: 962) continue this line of argumentation by stating that "knowledge of the core meaning of an academic word will help learners to gradually become aware of its discipline-specific meaning if they encounter the word often in texts from their specific subject areas". This argument shows that even if multiple meanings do exist of a word, which was established and accepted by Ming-Tzu and Nation (2004), learning the general meaning and as many of the other meanings as possible will help with further understanding and deepening of knowledge in the learners' own mental lexicons (Dang, Coxhead & Webb 2017: 962).

The second part of the response relates to the lexical coverage of the list or the percentage of words in a given text that is present on the AWL. Since the AWL was claimed to be only a "general service list" by skeptics (such as Hyland and Tse 2007), rather than an academic word list, the lexical coverage of fiction works was determined and compared with the coverage of academic texts using the AWL. The AWL gave a 1.4% coverage rate for the fiction collection and an 8.5-11.6% coverage rate (depending on the study) for a wide variety of

academic texts (Pecorari, Shaw & Malmström 2019: 59; Ming-Tzu & Nation 2004: 291). When looking at the coverage in specific individual disciplines, the results are all within the range provided in other studies of mixed disciplines. Chen & Ge (2007: 502) found a 10.07% coverage rate of medical research articles. Vongpumivitch, Huang, & Chang (2009: 22) found an 11.17% coverage rate of applied linguistics research papers. Martínez, Beck & Panza (2009: 183) found a 9.06% coverage of agricultural research articles. Ward (2009: 171) found an 11.3% coverage of engineering textbooks. Lastly, Li & Qian (2010: 405) found a 10.46% coverage of the Hong Kong Financial Services Corpus. Only one study has found a coverage rate outside of the 8.5-11.6% range and that was Konstantakis' (2007: 86) results which showed the AWL offered a 4.66% coverage rate of Business English course books. It is suspected the rate was so low due to the nature of language present in coursebooks as opposed to other academic texts as textbooks can be seen as a unique subset of academic texts that have their own features, rules, and perhaps vocabulary needs (Dang & Webb 2014: 68). This clearly shows that the words present on the AWL are indeed academic in nature, independent of the discipline an individual may be studying, and it would be beneficial to all learners who interact with academic texts in some form to study the words on this list. Furthermore, these percentages (with Konstantakis' results as an exception) are all higher than the 2-5% of words that can potentially be unknown in a running text while still being able to achieve acceptable comprehension of the text. Thus, making these words seemingly vital to learn for members of all academic subjects.

The third response is from Nation (2016: 149) where he argues that this criticism does not even need to exist when an "inclusive" approach is taken to learning and teaching instead of focusing on one solution or principle to guide classroom instruction. The point he makes is that both general academic vocabulary and discipline-specific vocabulary are vital to one's success (Nation 2016: 149). This is illustrated further since semi-technical vocabulary words (or academic vocabulary words) have been shown to be the words that second language learners struggle with most, meanwhile, the "fully technical words" or words that are subject-specific with a specialized meaning are difficult for all learners (Coxhead 2017: 99). Thus, the argument can be made that there is a need for some type of extra attention just on semi-technical vocabulary words for L2 learners, and then all learners need to focus on highly technical words which are content words and include "proper nouns, acronyms, and abbreviations" all of which are not included in the AWL (Coxhead 2017: 99).

Lastly, the AWL serves a unique purpose that cannot be fulfilled by subject-specific word lists. A valid consideration to point out is that students beginning their university studies

may not be exactly sure which discipline they will end up majoring in or they may be taking courses in multiple subject areas (Dang, Coxhead & Webb 2017: 961). Furthermore, some ESP classes are for a general audience with no specific subject focus. In these cases, a general Academic English word list is beneficial to have as a resource (Dang, Coxhead & Webb 2017: 961). General ESP courses can make it easier for universities to hire staff since many ESP professors may not have specific subject knowledge and one ESP faculty member could teach many classes with students from many different backgrounds. Finding specialists with ESP, EAP, and subject-specific knowledge is extremely niche and therefore hard to come by. Peters & Fernández (2013: 240) also emphasize the point that technical vocabulary is needed for architecture students but also knowledge of the words that makeup "this shared lexis that is shared with other academic areas of study and disciplines," as well as everyday words for items such as building materials to communicate with experts in other disciplines and nonexperts (Coxhead 2017: 107). Even in homogenous classrooms where students are sure of the direction of their study, they still need skills to be able to communicate and collaborate with colleagues and academics from all types of disciplines. For these reasons, Coxhead's AWL is a great resource for professors when considering a range of academic environments.

With this in mind, it is perfectly acceptable to see the AWL as one tool that can be used in a much larger toolbelt when someone is preparing or participating in the academic communities that exist and discourses that happen in English. Furthermore, since this research investigation aims to compare academic disciplines, it is the best tool to use as a neutral non-subject-specific word list (Durrant 2014: 329).

2.5.4 Other Written Academic Word Lists

Other academic word lists that have been developed over the past fifty years can be seen as alternative options for academic vocabulary learning. Some lists will fit certain users' needs better, but that is up to the discretion of the teacher when considering the class needs based on a needs assessment. These lists can be split into two categories, those that are focused on academic written texts which are based on corpora including written texts for academic purposes (i.e., journals, textbooks, etc.), and then other lists that are academic spoken word lists which are based on corpora of spoken texts in academic settings (i.e., lectures, presentations, etc.). Written word lists are far more prevalent, currently, than spoken word lists but there has been a recent wave of publications that have started to develop more spoken word lists.

The academic written word lists that have been developed include Campion & Elley's (1971) Academic Vocabulary List which was developed as a subtest for a university's English

entrance exam and pulled words from 19 different disciplines. Another list is Praninskas' (1972) American University Word List which pulls words from 10 textbooks from 10 different academic disciplines. Lynn's (1973) English for Academic Purposes Word Lists and Ghadessy's (1979) New Academic Word List were lists based on textbook annotations by foreign students, thus highlighting the words they struggled with the most. Xue & Nation's (1984) University Word List compiled the words from the previously mentioned 4 lists while removing any word that appears in the first 2000 words of the General Service List (GSL), a vocabulary list meant for general purposes. Coxhead's (2000) AWL was developed to update and improve the University Word List and therefore followed a similar approach by not including words that served a general purpose, but only an academic purpose. Browne, Culligan, & Phillips' (2013) New Academic Word List harkens back to the AWL but uses the New General Service List in hopes of updating the list even further. Lastly, Gardner & Davies' (2014) Academic Vocabulary List rejects the notion that learners have previous knowledge of more general vocabulary and therefore includes words from a more general background if they are commonly found in academic contexts. Some additional subject-specific written word lists that have been developed include lists for engineering students (Mudraya 2006; Ward 2009; Todd 2017), medical students (Wang, Liang & Ge 2008), and computer science students (Lam 2001).

2.5.5 Spoken Academic Word Lists - HSWL and SSWL

Additional academic word lists that focus on spoken contexts instead of written contexts have also been developed. Since they are sources from a different discourse type, they are not as relevant to the present study; however, two spoken word lists that are particularly interesting to investigate are the Hard science Spoken Word List (HSWL) and the Soft science Spoken Word List (SSWL) since they split the word lists based on Biglan's Hard/Soft paradigm (Dang 2018a; Dang 2018b). For this reason, they will be introduced below. Some of the observations and findings from these two word lists may prove to be relevant when considering the findings of the present study due to this hard/soft split consideration.

Dang (2018a; 2018b) initially began investigating the differences in vocabulary use between different academic disciplines. This research led to further investigations into potential differences in vocabulary use between soft and hard sciences. She sees this shorthand as useful "when attempting to explain the complexity and diversity of academic discourse" (Dang 2018b: 70) while still recognizing that the hard/soft division "cannot fully reflect the complexity and variation in inquiry processes and knowledge structures in various disciplines" (Nesi 2002:

351). She also recognizes that spoken word lists are to be used in combination with written word lists in EAP programs to "support L2 learners' comprehension of academic English" thus highlighting a need for both written and spoken lists, not one over another (Dang 2018b: 69). Additionally, as will be highlighted later in this subsection, the overlap between words that appear on both the HSWL and SSWL illustrates the fact that general academic word lists, such as the AWL, can also serve as a useful tool for teachers in EAP courses.

Multiple sources compared various features between soft science and hard science spoken and written texts. Dang & Webb (2014: 66) looked at the number of word families that need to be known to understand hard science speech and soft science speech. Coxhead & Dang (2019: 120) looked at the number of word families that need to be known to understand speech in a specific hard and soft science-specific event, for hard science it was speech in a lab setting and for soft science it was speech in a tutorial. Hsu (2011; 2014) determined the number of word families that need to be known to understand business and engineering textbooks. Chung and Nation (2004) compared the number of word types that were present and the percentage of technical vocabulary in an applied linguistics text and an anatomy text. Lastly, Dang (2018b: 80) discussed the differences between two-word combinations that are used in written and spoken texts in the hard and soft sciences. The results from all these studies have been combined into Table 5 below to help summarize and compare the findings.

Table 5. Trends and findings in spoken and written hard and soft science texts

				Observations	
Source	**Spoken/ Written**	**Hard/Soft**	**Event**	**Word families needed for 95% lexical coverage**	**Word families needed for 98% lexical coverage**
Dang & Webb 2014	Spoken	Soft	General speech	3-4k	5-7k
		Hard	General speech	4-5k	10-13k
Coxhead & Dang 2019		Soft	Tutorial speech	2k	4k
		Hard	Lab speech	3k	7k
Hsu 2011	Written	Soft	Business textbook	3.5k	5k
Hsu 2014		Hard	Engineering textbook	5k	10k
		Soft	Applied Linguistics text	20% of the 835 word-types were technical	

Chung& Nation 2004		Hard	Anatomy text	31.2% of the 4,270 word-types were technical
Dang 2018b	Spoken & Written	Soft	Two-word combinations	More likely to be the combination of two high frequency words (ex. *demand curve, fat man, green revolution, invisible hand*)
		Hard	Two-word combinations	More likely to be the combination of two low frequency words (ex. *harmonic oscillator, amino acid, radial velocity*) or a high frequency word and low frequency word (ex. *hash function, convex problem*)

From the findings summarized in Table 5, two main conclusions can be made. First, spoken text has a lower lexical burden than written text. This fact is supported time and time again in many areas of research, but it is emphasized here again. Furthermore, a comparison of the results helps further support the assertion from Dang (2018a: 46) that "comprehending written and spoken texts of hard sciences is more challenging than those of soft sciences in terms of lexical coverage". In the soft sciences findings, there were fewer word families needed in all instances, there was a smaller percentage of technical words in use, fewer word types in use, and two-word combinations consisting of more frequent words. In the hard science findings, there were more word families needed in all instances, a higher percentage of technical words in use, more word types in use, and two-word combinations consisting of infrequent words or a frequent and an infrequent word.

Other research has led to different conclusions about the nature of academic vocabulary within different academic disciplines. Durrant (2014: 353) found that vocabulary use does differ across disciplines but focused on the types of words by saying "[p]hilosophy students use specialized adjectives, engineers use specialized nouns, and science students use specialized verbs". Nation (2016) claims that fields like medicine and botany are fields with "large technical vocabularies" therefore requiring "L2 users to possess a large vocabulary to cope with their studies in academic or professional environments". Lastly, Lau and Gardner (2019: 265) claimed that the "soft-pure field typically demands much higher linguistic repertoires from its learners in order to negotiate meaning and convey academic thoughts".

While the differences between soft and hard science speech and writing appear to be significant, there is still quite a large overlap between the HSWL and SSWL. There are 1,344 word families that appear on both lists leaving 620 word families unique to the SSWL and 251 word families unique to the HSWL (Dang 2018b: 78). Furthermore, 98.81% of the words appearing on both lists are among the most frequent 3,000 British National Corpus (BNC) and Corpus of Contemporary American English (COCA) word families. Additionally, 93.55% of the words unique to the SSWL and 45.42% of the words unique to the HSWL are in the most frequent 3,000 word families of the BNC and COCA (Dang 2018b: 78). Due to the "nature of academic speech," these results "indicate that knowledge of the most frequent 3,000 word families of general English is important for the comprehension of spoken texts from both hard- and soft- sciences (Dang 2018b: 79). Another interesting finding is that 90% of the words on the HSWL and 82.28% of the words on the SSWL appear on the spoken version of the AWL, the Academic Spoken Word List (ASWL) (Dang 2018b: 80).

The duality of a difference existing between the demands of hard science vocabulary and soft science vocabulary while at the same time, an overlap between the two lists existing is what validates the research questions and point of inquiry of the present study. It will be the main point of focus to determine if, and to what extent, vocabulary knowledge of members of different academic disciplines varies. To test vocabulary knowledge, two developed and tested vocabulary tests will be used as measurement tools. The next section will deal with these vocabulary tests.

2.6. Vocabulary Testing

In the present study, the main measurement tools will be two vocabulary tests. These vocabulary tests were already developed and tested before this study. Many vocabulary tests that are used for measurement, reporting, and detailing vocabulary knowledge have been developed using various vocabulary lists as the source for their word families. There is not a vocabulary test that corresponds to every vocabulary list that has been developed, and in the present study's case, vocabulary tests that use the AWL as a source for word families have been chosen. There are not currently any vocabulary tests that use the HSWL, SSWL, ASWL, AVL, or other academic word lists that have been as well tested and studied.

Efforts to measure vocabulary size through tests started in L1 language settings with hopes of tracking progress as children aged. Researchers also hoped to be able to relate vocabulary size to other features of language use and production such as reading abilities (Read 2000: 76). Eventually, researchers began to realize that vocabulary size measurements past a

certain level became very difficult and the results they were able to obtain were hard to verify. This is when vocabulary tests started being used in L2 settings instead. Since L2 learners start with a very low level of vocabulary knowledge and build up their knowledge over time, only sometimes reaching the level of L1 users, it is much easier to measure L2 users' vocabulary knowledge. For teachers, knowing their learners' vocabulary sizes has been shown to be a useful tool since instructors were found to "not have good intuitions about what vocabulary their learners know" (Nation 2011: 530–531). Knowing the vocabulary size of learners can help set appropriate and attainable goals for learners as well as track progress and report gaps in L2 users' vocabulary knowledge (Nation 2011: 530).

While many different types of vocabulary tests have been created and used in research in the past, the two tests that will be used in this study are Schmitt, Schmitt, and Clapham's revised version of Nation's Vocabulary Levels Test (VLT) as well as Ghamarian-Krenn's Form-meaning Recognition Test of Academic Verbs (FoRTAV). Both tests measure the receptive knowledge, or passive knowledge, of the form and meaning of words which is highlighted by Nation (2018) as being two of the eighteen aspects of knowing a word which can be found in Table 4.

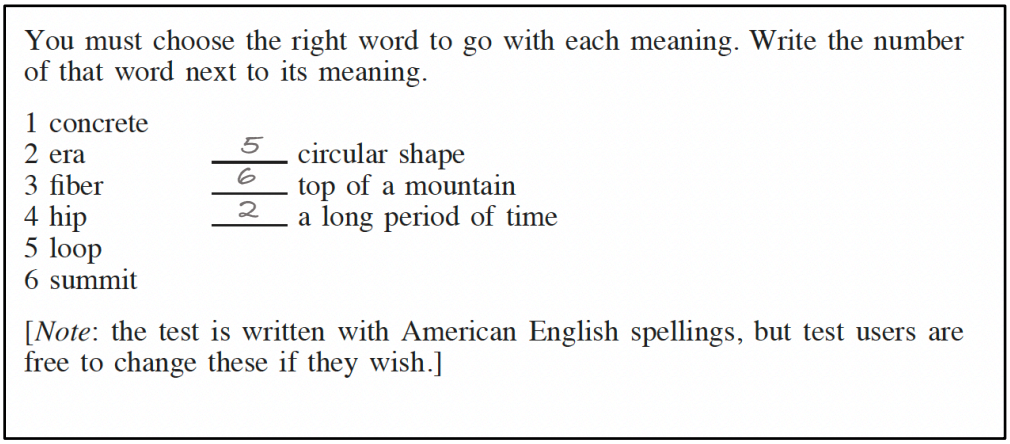
2.6.1 Vocabulary Levels Test

Nation initially created the VLT in the 1980s to be used as a tool in classrooms by teachers in the hope of making more effective and efficient evidence-based vocabulary teaching and learning goals accessible to all (Read 2000: 118). He wanted a fast and effective way for teachers to set goals and track progress without tests that would demotivate students or take too long. What is nice about the VLT, is that for beginners, the test can be shortened to prevent frustration and adjust it to lower-level learning goals. The VLT is regarded as "the nearest thing to a standardized vocabulary test currently available" and is seen by many as the standard for vocabulary tests (Kremmel & Schmitt 2016: 1). While many agree that the VLT has its issues, it "has proven validity and reliability" and has been shown to significantly distinguish different proficiency levels (Milton & Fitzpatrick 2013: 62).

Nation's original version of the VLT was split into five independent levels. These levels were based on 2000, 3000, 5000, and 10000-word frequency bands (respectively the 2k, 3k, 5k, and 10k levels) and then an academic vocabulary section that used word families from Xue and Nation's (1984) University Word List (Milton & Fitzpatrick 2013: 61; Webb, Sasao & Ballance 2017: 34). The higher levels, such as the 5k and 10k level can be excluded for beginner learners and the academic word list can also be left out when there is minimal use of knowing a

learner's academic vocabulary knowledge. Nation claims that knowledge of the words in the 2k and 3k levels is mandatory "in order to function effectively in English," while knowledge of the 5k band is the upper limit of words that are worth focusing on in a classroom setting, and they represent the upper limit of "general high-frequency vocabulary" (Read 2000: 119). The academic section is most relevant for learners looking to partake in university-level education. The scores on the individual sections are more revealing of a participant's actual vocabulary knowledge than the overall score on the test. The overall score should not be mistaken as a precise measure of vocabulary size; meaning if a student answers half of all questions correctly, they do not have a vocabulary size of 5,000 word families (half of the top level, 10,000 word families) (Busby 2021: 65).

The questions in the VLT are multiple-matching questions. This means that participants are given 5 or 6 words and then 3 definitions. They must then match the definitions to the correct word and leave out the others that have no match. An example of this type of question format with its correct answers is included in Figure 3 below:


You must choose the right word to go with each meaning. Write the number of that word next to its meaning.

1 concrete
2 era ___5___ circular shape
3 fiber ___6___ top of a mountain
4 hip ___2___ a long period of time
5 loop
6 summit

[*Note*: the test is written with American English spellings, but test users are free to change these if they wish.]

Figure 3. An example question from Schmitt, Schmitt, and Clapham's (2001) Vocabulary Levels Test

This type of question focuses on the word knowledge of the "form-meaning link learning" and since there are more words given than definitions, instead of more definitions than words or another word knowledge type, it is one of the most basic and initial stages of word knowledge (Kremmel & Schmitt 2016: 1). This type of link is the most efficient at providing an "indication of the potential for understanding and using words" (Nation & Webb 2010: 224). Each level contains 18 items with a fixed ratio of word classes at 5 nouns : 3 verbs : 1 adjective (Kremmel & Schmitt 2017: 1) and each of the items only features other answers of the same word class to "avoid giving any grammatical clues" (Read 2000: 120). Words with similar meanings were not included in the same items as this is intended to be a "broad measure of word knowledge" and not focused on precise meaning (Read 2000: 120). Lastly, the question type was designed with the hopes of minimizing reading time and minimizing the possibility to guess an answer

correctly; therefore, minimizing effort while maintaining the test's integrity and reliability (Read 2000: 119).

The updated versions of the test by Schmitt, Schmitt, and Clapham were created in 2001 and one of these versions is used in the present study. These versions are nearly the same as the original VLT except the number of items was increased to 30 items per level instead of 18 and new words were taken to increase statistical reliability (Schmitt, Schmitt & Clapham 2001: 71). Additionally, the academic section now used words from Coxhead's AWL, which was not available at the time that Nation initially created the VLT (Webb, Sasao & Ballance 2017: 34). Lastly, the ratio of word classes among the items was updated to reflect a better representation of the actual distribution of word classes in English (Kremmel & Schmitt 2017: 1–2). The ratio in the 2001 versions is 3 nouns : 2 verbs : 1 adjective (Kremmel & Schmitt 2017: 1–2). This update was successful in limiting some of the issues that had been found with the original VLT. Furthermore, by making these changes, the 2001 versions were found to be more reliable than the original VLT, meaning due to a higher number of questions the scores are more likely to reflect the actual vocabulary knowledge of participants. As sample size increases, in terms of the number of participants and number of questions, statistical reliability also increases. The four frequency sections (2k, 3k, 5k, and 10k) were found to have very high statistical scalability meaning "an implication scale can also be assumed for the frequency levels" (Webb, Sasao & Ballance 2017: 34). Mastery of the higher-order levels implies knowledge of the lower levels, even those that were not explicitly tested, i.e., the 8,000 is known if the 10,000 word family level is passed and the 4,000 word family level is mastered if the 5,000 word family level is passed (Kremmel & Schmitt 2017: 2).

When determining what will be considered mastery, multiple sources have cited different cutoff points for the VLT. The main question when determining a cutoff point for mastery comes down to the goals of the experiment and the expectations of mastery. Nation interpreted this initially as test takers needing to score 66% on the test sections to be considered for mastery. Schmitt, Schmitt et al. (2001), suggested 26/30 items correct originally, but 24/30 items correct has also been suggested by Schmitt (2010) as the mastery cutoff. Other researchers have concluded that these scores are arbitrary and that there should be a flexible scale with higher cutoff points for the lower levels, such as 29/30 correct for the 2k and 3k sections, and lower cutoff points for the higher levels such as 24/30 for sections above 3k (Webb, Sasao & Ballance 2017: 56). Since the VLT measures a very basic level of word knowledge, a higher cutoff point could be the best measure that reflects real life results (Webb, Sasao & Ballance 2017: 56).

Some of the issues brought up with the VLT (newer and older versions) have to do with score inflation and the ease of the test. To minimize this effect, a higher mastery cutoff point of 26/30 per section will be taken as Busby (2021: 65) did in her research. All the sections will have the same mastery cutoff point for consistency throughout the analysis, but implications of this decision will be considered in the discussion and interpretation of results section. Some researchers have claimed that the scores on the VLT are overestimating the receptive abilities of the test takers and are inflated from anywhere between 16 - 20% (Kremmel & Schmitt 2016: 383). Furthermore, Kamimoto (2008), Webb (2008), and Stewart & White (2011) point out that blind guessing can inflate the scores by around 17%. However, since the participants of current study are expected to perform quite well on the tests, this is less of an issue since Stewart & White (2011) conclude the impact of guessing reduces after 60% of the words are known. This means that low-scoring test takers are more likely to benefit from guessing (Kremmel & Schmitt 2017: 3). With these factors in mind, the participants in this study will be encouraged not to guess. Additionally, since the grade inflation is "relatively systematic and in one direction" it could be possible that a score adjustment might be developed to compensate for this at some point in time (Kremmel & Schmitt 2016: 389). Lastly, due to these reasons, participants will need to perform very well on the test (i.e. mastery of the academic level and 10k level), to be considered to have adequate receptive vocabulary knowledge for academic studies. In this study, the VLT will serve as a measure of general vocabulary knowledge with a small insight into the participants' academic vocabulary[2]. The full tests that were used can be found in Appendix II.

2.6.2 Form-Recognition Test of Academic Verbs

To further examine the academic knowledge of the participants, a second and more extensive academic vocabulary test will also be used. This test is the FoRTAV which was developed by Ghamarian-Krenn as one of a series of academic vocabulary tests. This is the form-meaning recognition test, and it is one of four tests that were created to measure the development of academic verb knowledge in students just beginning their studies at the University of Vienna's English bachelor's program (Ghamarian-Krenn 2023). The vocabulary tests developed by

[2] Since productive vocabulary tests are more likely to be affected by sociocultural conditions at the time the production is measured, using a receptive knowledge test is the best fit for the present study as receptive knowledge is more constant and less dependent on sociocultural conditions. Furthermore, other vocabulary tests were considered to be used in place of the VLT, such as the Vocabulary Size Test (VST) which tests the 14,000 and 20,000 word levels as well and would be less likely to lead to a ceiling effect. However, this test was not used due to the exclusion of an academic vocabulary section and because the VST has been "criticized for having an insufficient sampling rate" (Busby 2021: 65).

Ghamarian-Krenn (2023: 162) focus on academic verbs "due to the diverse nature of words which co-occur with academic nouns", which would have made it "difficult to create suitable and regular items" for the collocation recognition test, another of the tests developed by Ghamarian-Krenn (2023:162).

The FoRTAV is labeled as a "discrete, selective and context-dependent test" but "context-dependent" is relative since the linguistic context given in the test is "not very rich and simplified" (Ghamarian-Krenn 2023: 161). The verbs being tested on the FoRTAV were taken from Coxhead's AWL. The design of the FoRTAV is based on the VLT and the format is very similar. Participants are given a few definitions and then several academic verbs and they must match the definition to the appropriate word. This is the opposite format of the VLT. An example of an item featured on the FoRTAV is shown below in Figure 4:

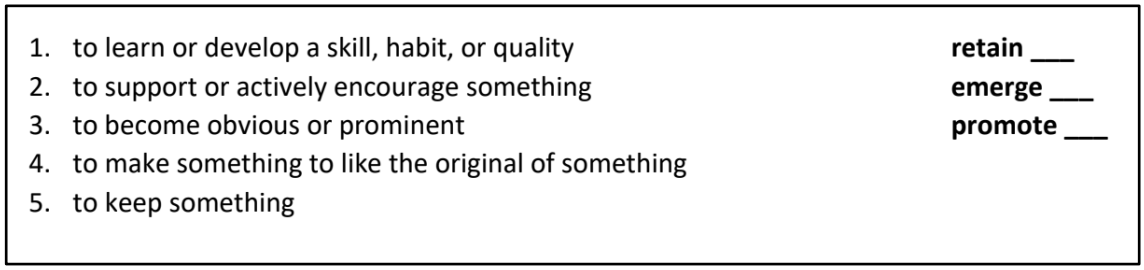


Figure 4. An example from Ghamarian-Krenn's Form Recognition Test of Academic Verbs (FoRTAV)

The FoRTAV has not yet been used with groups of participants outside of the English department; however, it has been suggested that "it could be suitable for less restrictive target groups with slightly lower proficiency levels as well" (Test Manual FoRTAV). Due to the relatively new nature of the test along with only having been used for students enrolled in English programs, this test could not be used alone to determine student academic vocabulary knowledge. Hopefully, in tandem with the VLT, the results from both tests will be informative about the types of vocabulary knowledge possessed by the students who participate in the present study.

3. Methodology and Research Design

The present study involved creating and distributing a short questionnaire along with the two vocabulary tests (VLT and FoRTAV) to students enrolled in master's programs across Austria. Details on how the questionnaire was developed, some of the concepts included, the vocabulary tests, the participants, and pre-testing will be given in the coming section. For all statistical analyses and graphs, the open-source statistical programming, R, was used (R Core Team 2020). The following packages were also used as additional analysis and graphing tools: readr, tidyr, dplyr, tidyverse, gapminder, reshape2, ggpubr, ggstatsplot, RColorBrewer, paletteer,

ggrepel, and stats (Wickham et al. 2023; Bryan 2023; Kassambara 2023; Slowikowski 2023; Hvitfeldt 2021; R Core Team 2020; Neuwirth 2022; Wickham, Hester & Bryan 2023; Wickham 2007; Wickham, Vaughan & Girlich 2023; Patil 2021; Wickham et al. 2019). All code created for the data analysis has been included in Appendix III.

To try to learn about master's students enrolled in Austrian universities vocabulary knowledge, the present study will aim to address four research questions. The first research question will aim to assess how well the participants performed on the two vocabulary tests.

> **RQ1:** What vocabulary knowledge do master's students enrolled in Austrian universities possess at different frequency levels according to the Vocabulary Levels Test (VLT) and the Form Recognition Test of Academic Verbs (FoRTAV)?

The second research question will zoom in on the vocabulary knowledge of the students in relation to Biglan's and Holland's groupings. Statistical testing will be performed to determine if there are significant differences or similarities between the groups of students.

> **RQ2:** In what ways does vocabulary knowledge vary between Biglan's academic groups when considering master's students in Austria? In what ways does vocabulary knowledge vary between Holland's academic groups when considering master's students in Austria?

The third research question will examine the results from the questionnaire pertaining to the students' current master's programs and determine which levels of English exposure in their degree programs may appear to correlate with vocabulary knowledge. Other factors, such as participant age, and semester progress will also be examined to determine if these factors show any correlation as well.

> **RQ3:** How does vocabulary knowledge compare between master's students with different proportions of English as part of their curriculum in Austria? What other factors might impact the vocabulary knowledge of master's students in Austria?

The last research question will compare the participant results from the AWL to the participant results from the FoRTAV. Since the FoRTAV test has not yet undergone significant validation testing it will be interesting to compare the results and see if students perform equally on the two tests.

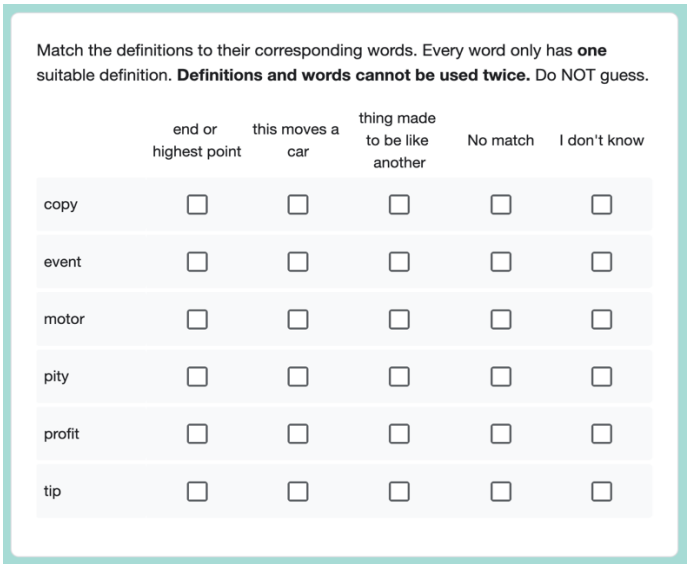
> **RQ4:** How do the results from the Academic Word List (AWL) section of the VLT compare with the results from the FoRTAV?

To attempt to answer these four research questions, several tools were necessary. The modifications made to the vocabulary tests, the development of the questionnaire, pre-testing,

variable development, and an introduction to the study participants will be given in the rest of this section.

3.1. VLT and FoRTAV

The use of the VLT and FoRTAV was determined to be the best fit for the present study due to the reliability of the VLT and the specific nature of the FoRTAV that is relevant to the participants of this study. The combination of the two tests together will provide the chance for results that are reflective of the participant's general English vocabulary knowledge and specific academic vocabulary knowledge. Furthermore, there will be an opportunity to compare each participant's results on the FoRTAV to their results on the academic section of the VLT. The full tests that were used, in the format they were given in, can be found in Appendix II.



Match the definitions to their corresponding words. Every word only has **one** suitable definition. **Definitions and words cannot be used twice.** Do NOT guess.

	end or highest point	this moves a car	thing made to be like another	No match	I don't know
copy	☐	☐	☐	☐	☐
event	☐	☐	☐	☐	☐
motor	☐	☐	☐	☐	☐
pity	☐	☐	☐	☐	☐
profit	☐	☐	☐	☐	☐
tip	☐	☐	☐	☐	☐

Figure 5. Modified grid format example from the VLT

Since both vocabulary tests are provided openly in pen and paper formats with little ability for easy distribution and a simple way to collect the results, the tests were converted to an online format using Google Forms and paired with the questionnaire. With this approach, the questionnaire and both tests could be accessed through the same link, and it would not require the participants to send their own results for processing. The results were automatically compiled into a spreadsheet where the results could be processed. Busby (2021) did a similar thing when administering the VLT to university students in Norway. She states that by presenting the clusters in a grid formation, "the instructions are as transparent as possible" thus reducing the potential for confusion when facing a potentially new test question type (Busby 2021: 65). This is especially beneficial since the tests and questionnaire were distributed remotely and the students were not able to contact anyone for immediate troubleshooting if they

did not understand. Figure 5 (previous page) shows an example of a grid format cluster for the VLT questions and Figure 6 (below) shows an example of the grid format cluster for the FoRTAV. At the beginning of the first test, the participants were provided with one example cluster of basic words so they could see what an appropriate set of answers was supposed to look like.

Match the definitions to their corresponding words. Every word only has **one** suitable definition. **Definitions and words cannot be used twice.** Do NOT guess.

	retain	emerge	promote	no matching word	I don't know
to learn or develop a skill, habit, or quality	☐	☐	☐	☐	☐
to support or actively encourage something	☐	☐	☐	☐	☐
to become obvious or prominent	☐	☐	☐	☐	☐
to make something to like the original of something	☐	☐	☐	☐	☐
to keep something	☐	☐	☐	☐	☐

Figure 6. Modified grid format example from the FoRTAV

Participants were also informed in the instructions to take the tests as quickly as possible since passive knowledge is best measured in a quick time span. This time pressure is placed in hopes of reflecting the immediate demand for meaning that is required to understand and comprehend text while reading. There was no actual time limit on the response time, but the hope was by including a short note in the instructions, the participants would complete it more quickly. Furthermore, the participants were given the option to select “no matching word” and “I don’t know” in addition to being informed that any blank answer would be recorded as “no matching word”. The instructions were repeated at the beginning of each section of the test and shorter more precise instructions were repeated with each cluster.

Scoring for the two vocabulary tests was relatively simple. Only the words that had a match would be counted towards the total score, leaving the VLT with 5 subsections of the test to grade out each out of a possible 30 points and the FoRTAV as one longer test with 72 possible points. The subsections of the VLT tested the 2,000 word family level, 3,000 word family level, academic level, 5,000 word family level, and the 10,000 word family level. Any individual level score including and above 26/30 was considered as passing on the VLT to reflect common scoring practices among other studies using the test as a vocabulary measurement tool as well, especially since this was the passing cutoff used by Busby (2021).

Since there is no established passing score for the FoRTAV, the minimum from the VLT (26/30 or 86.6%) was transferred to the FoRTAV thus requiring a minimum score of 62/72 (86.1%) to be determined as passing. It was important to set a pass/fail threshold since the results on all the tests will be analyzed in terms of total score as well as trends in passing and failing. In addition to the two vocabulary tests, participants were also required to complete a questionnaire. Details about the creation of the questionnaire are provided in the following section.

3.2. Questionnaire

The questionnaire developed for this study was to provide background information about the participants, find out what roles certain languages have in their lives, determine the amount of English they have been exposed to in their studies, and find out which study program they are currently enrolled in. A pre-exiting questionnaire that matched the research aims of this thesis could not be found, therefore, this questionnaire was created from scratch. The information collected will be used in creating multiple variables for the data analysis. More information on these variables and how they've been developed will be included in Section 3.4. The full questionnaire that was distributed is included in Appendix I. In general, the tradeoff for turning words into numbers is granularity; however, for the initial study of some of these topics, the tradeoff was made. The questionnaire captured many different topics since from the beginning it was known the analysis would be mostly numerical, but other less quantitative research studies could look at several of the topics covered in more depth if they create further questions or require a deeper look. Some other topics could involve the impact of the topic of the participant's bachelor's study or a different measurement for vocabulary proficiency.

The participants were found by word of mouth, sharing the questionnaire link on various social media pages of university study groups, snowball sampling by asking participants to share the questionnaire link with classmates, and by contacting professors who teach classes in EME programs, asking if they would be willing to send the surveys to their students. This led to an initial total of 144 participants; however, only 85 of the responses qualified as valid participation due to the target group of this study.

The participants were initially given a short introduction to the study and asked to consent to the use of their results. After that point, they were also asked in the questionnaire if they would like to receive their results and score breakdown. They did not have to provide this if they did not want to and without providing an email address, they were still able to participate in the study. Receiving their results back served as a motivating factor for many participants, as almost all of them left their email addresses. Participants were also asked to fill

in a class code if they were provided one by a professor since they were participating as a class. If enough participants provided a class code, their professors were sent an anonymized version of the results to illustrate the vocabulary knowledge of a portion of their students. If a student did not enter a class code, they could still participate, and their results would still be processed as a part of the study.

Several of the items included in the questionnaire were inspired by Dafouz and Smit's (2023) ROADMAPPING framework. Many of the items in the questionnaire deal with languages in use including details on when participants use certain languages in their lives. These questions were inspired by the role of English in relation to other languages' dimensions (RO). Participants were asked about their first languages, their self-reported level of English, the languages present in their bachelor's programs, and the languages they use in their master's program with various agents (A) including professors and administrators, as well as with classmates inside and outside of class. By determining which languages are potentially at play, more information about the multilingual reality of these HEIs will be investigated. Additionally, the students were asked what percentage of their master's and bachelor's programs contained teaching, reading, writing, and speaking in English (each aspect was a different question). The students were also asked to provide the name of their bachelor's and master's programs, and this was done with the aim of sorting the students into their academic disciplines, another dimension presented by Dafouz and Smit (AD). To reflect some basic information about the classroom practices and processes (PP) students were asked to report if their program offered an ESP class, if it was mandatory if it was offered, and if they had taken the class yet. Additionally, to peak into the internationalization and glocalization dimension (ING) the participants were asked to provide their nationality as well as if they had studied abroad at some point, and if so, they were requested to detail the duration and location of their experiences.

Additional items on the questionnaire were items that Busby (2021) discussed in her article where she tested the vocabulary sizes of Norwegian university students. The exact questionnaire she used was not provided in the article, but from her discussion of results, a few items can be inferred. She tested groups of students from three different subjects: English, IT, and Sociology. In a similar way, participants in this study will be split by their academic discipline grouping according to two separate approaches: the Biglan/Kolb approach and the Holland approach. She also asked her participants how far along they were in their studies so participants in the present study were requested to provide an approximation of the number of credits they had completed already in their program. Since some participants may be studying

full time and others might be studying in addition to working or other duties, credits were taken as a more representative counting unit to measure progress through a program rather than the number of semesters a participant had completed. Busby used this measure to represent the amount of English the students had been exposed to, making the argument that the longer the participants had studied, the more English instruction they will have encountered. To avoid assuming, the students in the present study were also asked to rank the amount of speaking/presenting, listening, reading, and writing they had been doing on average in their classes in English in their master's programs as well as in their bachelor's studies. This way there would be further information collected contributing towards quantifying an estimate on the amount of English they had been exposed to. Age was also collected to determine if vocabulary knowledge is impacted by age and to determine if studying time and exposure to English has more of an impact than age on the results.

Once the questionnaire and vocabulary tests were digitalized, the first surveys were sent to a group of pre-testing participants. Results and any changes that were made based on the results from these initial participants will be described in the next section.

3.3. Pre-testing

The questionnaire and two vocabulary tests were pilot tested on a group of 20 people. Of these 20 participants, 10 submissions were able to be used in the final analysis as valid participants. In addition to these 10 participants, there were 3 native speakers of English in the pilot group as well as 6 participants who studied in an English program, where English as a subject is the focus of the study. These 9 participants could not be used in the final analysis since they do not meet the criteria of the study, but they helped validate the questionnaire and the vocabulary tests. The last participant of the pilot group misunderstood part of the email consent question and disqualified themselves from the study by not agreeing to the disclaimer. The disclaimer was edited for clarity after this point, and it did not appear to be an issue for other participants.

The questionnaire did not need to be changed after the pilot participants completed the study except for a few small details. As mentioned earlier, the consent question was edited for clarity and slightly reworded. Additionally, a question was added to the end of the questionnaire to explicitly ask if the participants were currently enrolled in a master's or masters-level equivalent program. Prior to this, there was a way to interpret and answer the questions as someone who at some point had been enrolled in a master's program but by adding an additional question at the end, any possibility of these types of participants going undetected was removed.

After the pilot participants finished the questionnaire and vocabulary tests, their results were examined. A decision had to be made based on the results of the tests if these would be the vocabulary tests to be used in the study. The vocabulary tests would have to be used in their entirety to maintain consistency with other studies and editing them was not a possibility. The results were calculated, and averages were determined to help make this decision. Table 6 below shows the average scores on each part of the vocabulary tests of the subgroups of pilot participants, the average of all the pilot participants together, as well as the maximum scores that were possible on each section.

Table 6. Average scores of pilot testing subgroups

Group (Number of members)	2k	3k	Academic	5k	10k	FoRTAV
Native Speaker Average (3)	29.3	30	30	30	30	68
English Student Average (6)	28.2	28.8	29.4	28.8	22.2	63
Valid Participant Average (10)	28.4	28.4	29	26.8	20.9	49.8
All pilot study participants (20)	28.5	28.8	29.3	27.9	22.8	56.5
Maximum Possible Points	30	30	30	30	30	72

The decision was made to carry on with these vocabulary tests as the main measurement tools since the results from the pilot group were as expected. The native speakers performed very well, which was expected and validates that the tests are not meant for their vocabulary knowledge. The students enrolled in English-subject programs also performed well, but not to the same level as native speakers. The valid participants, or the participants who matched the target group for this study, performed the worst out of the subgroups but overall, the tests appeared to be good measures of their abilities in terms of vocabulary knowledge. The scores were not too low, which if it had been the case may have been a symptom of test fatigue or evidence the tests were too difficult.

With the promising results from the pilot group vocabulary test results as well as minimal issues with the questionnaire, the study was continued, and the surveys were distributed to as many students as possible. The next section will discuss how the participants were separated into their groups, other pre-processing steps that were taken to prepare the data for an analysis, and a breakdown of how the individual variables were created that will be used in the statistical analysis.

3.4. Data Preprocessing and Variable Development

To prepare the results from the survey to be processed, several changes needed to be made to the structure of the dataset. First and foremost, the scores were determined for the various sections of the VLT as well as the FoRTAV. To determine the individual scores, the results were turned into binary, one represented a correct answer and zero represented an incorrect answer. The distractors did not count towards the final scores since they do not count in the original versions of the VLT and FoRTAV due to the difference in formats of the tests[3] compared with the format that was distributed for the present study. The results from the distractor items could be analyzed for common errors or trends, but in the present study they were not examined. After each result was transformed into binary, the section scores were able to be determined. The five sections from the VLT were scored out of 30 points and the FoRTAV score was scored out of a total of 72 points. The total AWL scores were also determined (out of 150) since Busby (2021) also used the total scores in her data analysis. Additionally, another total score was calculated by adding all the AWL sections together along with the FoRTAV scores, resulting in a possible score out of 222. The scores composed from multiple test sections will be used in the analysis; however, it is important to note that the total scores do not correspond to a set vocabulary size in participants, as the VLT was developed as a levels test not a size test and the FoRTAV is measuring a non-leveled set of words, as academic words are not inherently frequent or infrequent by nature.

Based on the theoretical ideas that then were operationalized to create the questionnaire items, several variables were created to be able to separate the participants into groups as well as summarize several of their experiences into numbers. The vocabulary tests were lengthy; therefore, a less demanding approach was taken for the questionnaire items to encourage the participants to complete it in its entirety. A brief description of each variable that was measured, as well as how it was determined will be given below. Additional or new studies could consider reanalyzing the data in different ways with different interpretations of the questionnaires.

a. Self-reported English proficiency – students ranked themselves according to the CEFR language ranking. This is a self-determined variable by each participant based on short

[3] The difference in the formats of the tests refers to the fact that the present study included options for "I don't know" and "no match". Meanwhile, the original versions of the VLT and FoRTAV do not offer any alternative answer options besides the matching words or definitions. Including an "I don't know" options has been shown "to improve guessing behavior and exactness of the test score (e.g. Peters, Velghe & Van Rompaey, 2019; Stoeckel, Bennett & McLean's, 2016)" (Ghamarian-Krenn 2023: 300).

descriptions of the language levels (A1, A2, B1, B2, C1, C2, and "English is my first language"). Students who picked that English was their first language were disqualified from the main part of the study and their results are not included in the analysis since the vocabulary tests are not intended for L1 users.

b. Study abroad experience – the participants were sorted into two groups, those who reported they had studied abroad at some point and those who said they had not. Further separation was considered since those who did study abroad provided the location and length of their time abroad, but it is nearly impossible to conclude from this information how much English they potentially encountered due to the role of English as a lingua franca.

c. Teaching/Reading/Writing/and Presenting & Talking in English in current study program – the participants were asked to rank on scales of 0-5 the number of classes they have taken in their current study program that have been taught in English, as well as the amount of reading, writing, and presenting & talking they do in English in their classes in their current study program. These scores were kept on a scale of 0-5 and used as a measurement of their exposure to listening, writing, reading, and speaking activities in English in their current program.

d. Average EME Score– this variable was created by taking the mean of the previous four variables to create an average exposure score.

e. Commonly used languages – participants were asked to report what languages they use inside and outside of the university setting on a regular basis. To determine this information, they were asked to list which languages they use: in class with classmates, out of class with classmates, with university professors, with administrators, with their family, with friends, and at work.

f. Number of languages regularly used – from the previous variable, the number of unique languages a participant answered was determined to show the number of languages they use on a regular basis.

g. Biglan Categories – based on the master's program (or master's-level program in the case of medical and legal studies) the students were assigned to their corresponding groups (Hard or Soft; Pure or Applied). For studies that were difficult to determine which group they belonged to, several other studies were referenced for their sorting guides (Doberneck & Schweitzer 2017; Nelson Laird et al. 2006; Simpson 2017). When facing a program that had not yet been sorted into a group, the webpages of the various

programs were used to determine the group. Special attention was paid to the bachelor programs that guarantee admittance to the master's programs as well as the posted curriculum. Some programs were interdisciplinary in nature, leaving some programs as more than one category but this was tolerated to reflect the true distribution of disciplines among the participants.

h. Holland Categories – again based on the master's program the participants were assigned to their corresponding group based on Holland's categories. For studies that were again difficult to assign to a group, previous studies were referenced to determine their grouping as well as a few university websites that split studies into Holland's groups on university counseling pages (Gottfredson & Holland 1996; Smart, Feldman & Ethington 2000). Again, several programs fit into multiple categories and in these situations, the best fitting category was determined based on the program websites or when the program seemed truly interdisciplinary in nature, the program was assigned multiple labels. Only the categories with more than 15 participants were kept resulting in four Holland categories (Enterprising, Investigative, Realistic, and Social).

i. Semester progress – the participants were requested to estimate the number of ECTS or credits they had completed in their program so far. Since 30 ECTS is equal to one full-time semester of work, the ranges corresponded to the number of semesters they would have completed (including the current one) if they had work on their studies full-time. The range of 0-29 was transformed to 1 to indicate 1 semester of studying, 30-59 was transformed to 2 to indicate 2 semesters of studying, and so on up to 180 ECTS which is a complete master's program. This method of determining progress through one's studies eliminates the potential impact of part-time studying or pauses that one make take due to certain life circumstances.

From these main variables[4] the analysis was able to be undertaken. In the next section, information about the participants will be summarized.

3.5. Participants

The participants selected for the study were all master's students enrolled in studies at Austrian universities and "Fachhochschulen" (FHs), or universities of applied sciences. Some of the students were from other universities but studying abroad at the time in Austria and their results

[4] Additional variables could be created as some of the items that were collected on the questionnaires have not been turned into variables nor examined due to the scope of the present study.

were also included in the final group. In total, 85 participants completed both vocabulary tests and the entire questionnaire. As illustrated below in Figure 7, 52 of the participants identified as female, 33 as male, and most of the participants were between 22 and 29 years old. The gender distribution is illustrated in Figure 7 and is slightly skewed toward female participants; however, the age distribution is as expected as most students complete a master's degree after finishing a bachelor's degree. In Figure 8, the ages of the participants are represented with a histogram. Most of the participants fell between the ages of 22 and 29, which is to be expected from master's level students.

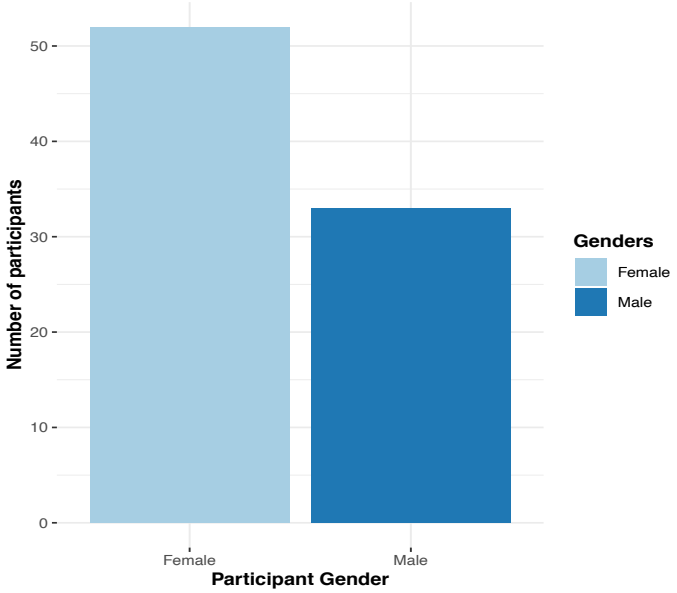


Figure 7. Gender histogram of participants (gender vs. count)

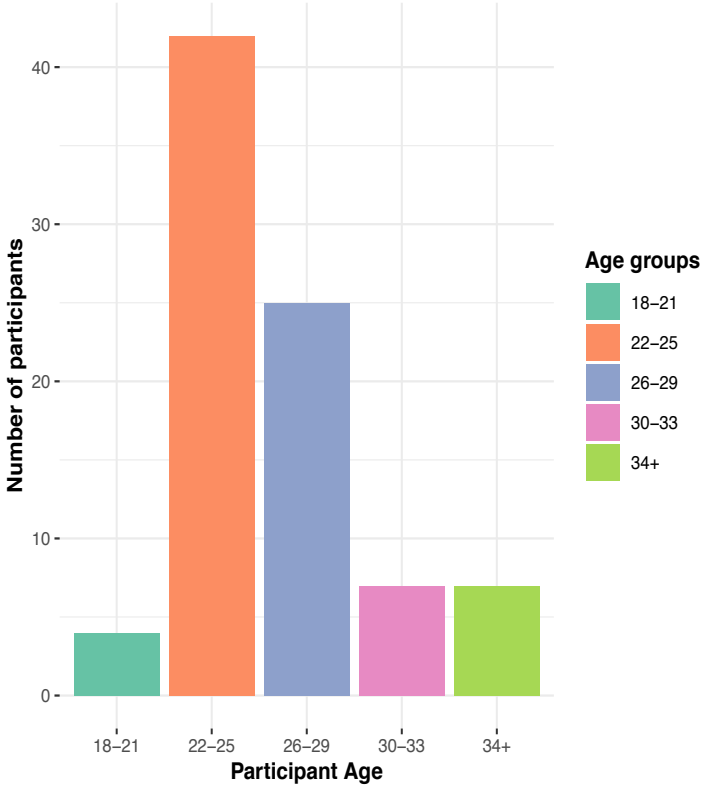


Figure 8. Age histogram of participants (age vs. count)

The students ranked their English abilities on the CEFR scales, and the results are illustrated below in Figure 9. These results are not reflective of the actual ability levels of the students but instead the perceived levels the participants believe they have. Even if the scores are inflated because of self-reporting, it can be concluded that a majority of the participants are at least intermediate users of English and have some familiarity with the language.

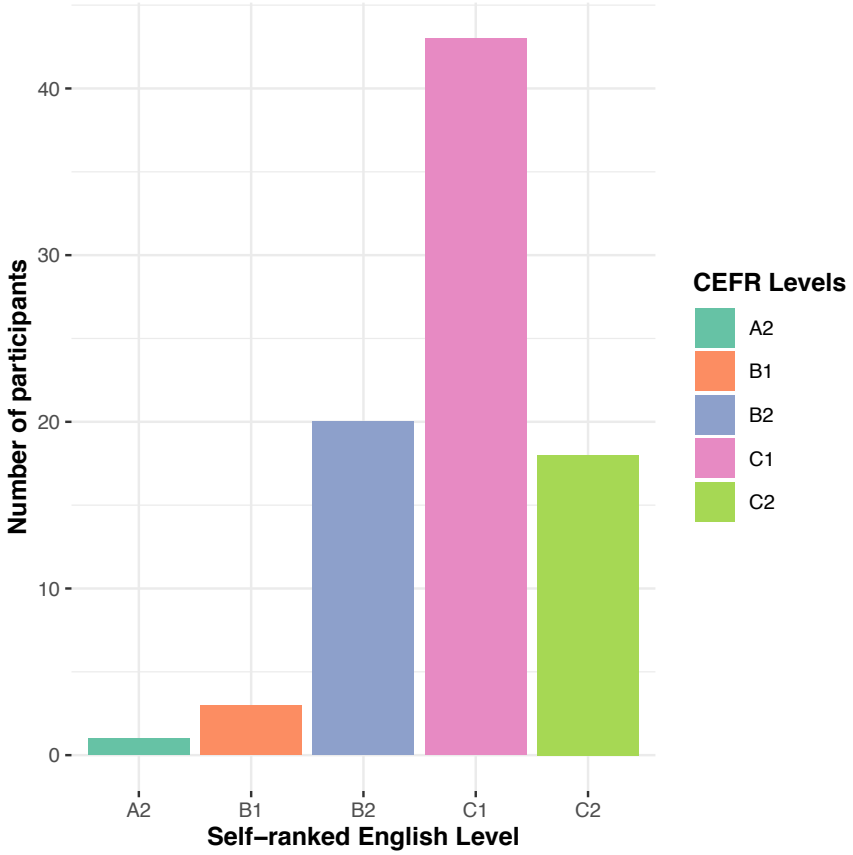


Figure 9. Self-ranked English level histogram (CEFR level vs. count)

From the valid participants, the students were split into their academic disciplines according to Biglan's approach and Holland's approach. As mentioned in the previous section, some of the academic disciplines were determined to belong to more than one group, so the group totals do not equal the final number of participants. Additionally, Holland initially split students into 6 groups, but as mentioned in the academic discipline section, Smart determined that only 4 were found to be compatible with academic studies. Smart does not include the realistic and conventional majors in his discipline research. In the present study, there were not enough participants with conventional and artistic groupings; however, there were larger groups of enterprising, investigative, realistic, and social students. Due to insufficient sample sizes, the conventional and artistic groupings were not included in the present study.

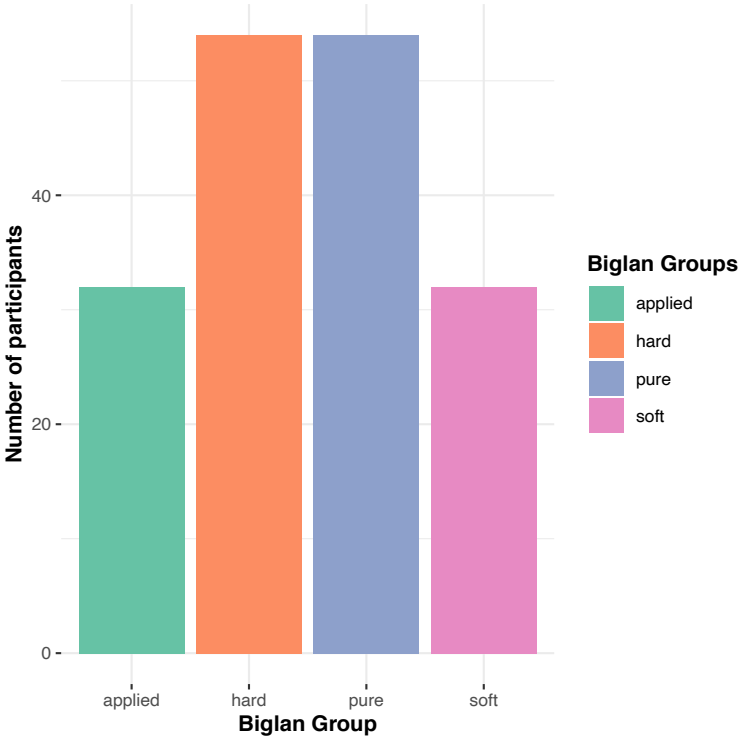


Figure 10. Participant distribution according to Biglan's grouping scheme

In figures 10 (above) and 11 (below), the number of participants according to the Biglan and Holland grouping schemes are shown. The groups representing the hard, pure, and investigative students are the largest in the present study. The investigative sample size is much larger than the other three groupings, and the results will be reflective of skewed data grouping, not a symmetrical one.

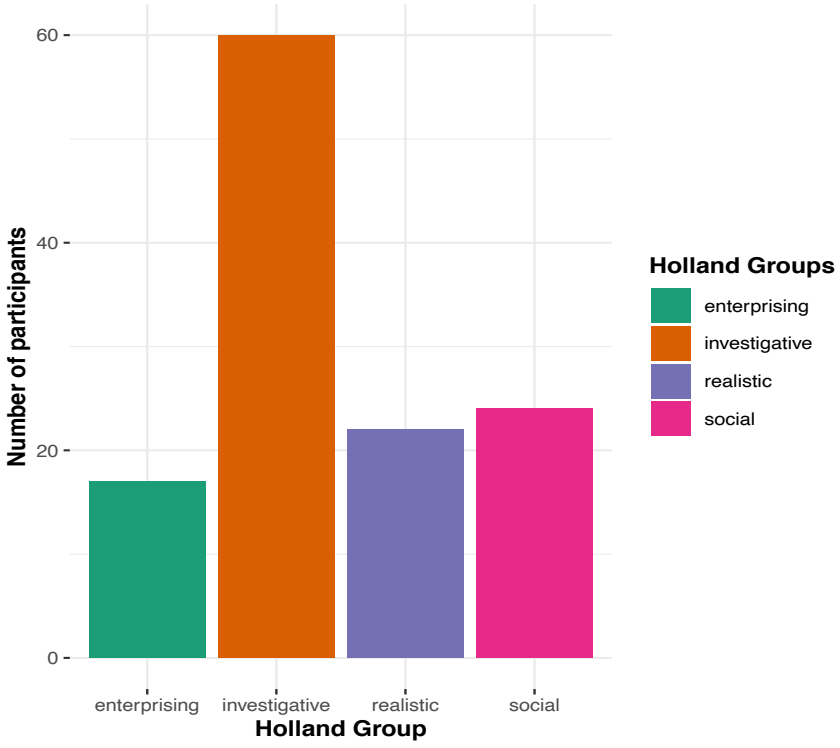


Figure 11. Participant distribution according to Holland's grouping scheme

The nationalities of the students are mostly Austrian, as shown in Table 7 below. The nationalities were added by the participants themselves, hence the vague answer of Southeast Asia. The second largest group of participants is German, thus reflecting the international student enrollment trend in Austrian universities where Germans make up the largest minority (Dannerer, Gasich, & Smit 2021: 290). Even though the data here is skewed to represent mostly Austrian students, this is expected from the data set since the study is being conducted in Austria at Austrian universities.

Table 7. Participant nationalities

Country	Number of Participants
Afghanistan	1
Austria	61
Bulgaria	1
Egypt	1
Germany	7
Hungary	2
India	2
Italy	1
Malaysia	1
Mexico	1
Russia	2
Southeast Asia	1
Turkey	2
Ukraine	2

There were 52 participants in the study enrolled at the University of Vienna, as illustrated in Table 8 below. The rest of the students were enrolled at other Austrian universities around the country. In the final group of participants, none were enrolled at a FH. The participants are therefore overly represented and skewed toward students at the University of Vienna.

Table 8. *Participant universities*

University Name	Number of Participants
Medizinische Universität Wien	1
Technische Universität Graz	8
Technische Universität Wien	2
Universität für Bodenkultur Wien	5
Universität für Musik und darstellende Kunst Wien	1
Universität Graz	4
Universität Klagenfurt	1
Universität Salzburg	5
Universität Wien	52
Wirtschaftsuniversität Wien	5
Other	1

With the above information about the participants, including their ages, nationalities, universities, self-ranked English level, gender, and academic grouping, the data analysis can begin. In the next chapter, the results from the tests and questionnaires will be summarized and compared.

4. Results & Statistical Analysis

For the data interpretation section, the subchapter organization will be based on the research questions. Each research question and the corresponding results will be provided. Research questions 2 and 3 have been split into 2a, 2b, 3a, and 3b for clarity and will each have their own subchapter.

RQ1: What vocabulary knowledge do master's students enrolled in Austrian universities possess at different frequency levels according to the Vocabulary Levels Test (VLT) and the Form Recognition Test of Academic Verbs (FoRTAV)?

RQ2a: In what ways does vocabulary knowledge vary between Biglan's academic groups when considering master's students in Austria?

RQ2b: In what ways does vocabulary knowledge vary between Holland's academic groups when considering master's students in Austria?

RQ3a: How does vocabulary knowledge compare between master's students with different proportions of English as part of their curriculum in Austria?

RQ3b: What other factors might impact the vocabulary knowledge of master's students in Austria?

RQ4: How do the results from the Academic Word List (AWL) section of the VLT compare with the results from the FoRTAV?

There will be two main statistical approaches to analyzing the relationships between the variables. All the statistical tests will be nonparametric due to the non-normal distribution of the test results (more details on this will be given in the next section).

The variables related to semester progress, age, amounts of reading, writing, speaking, and listening in English present in the participant's master's program, the number of commonly used languages, and the participant's self-reported English level are ordinal, meaning that they follow a ranking. When considering age, 34 is more than 26, which is more than 22. For the level of English, C1 is more than B2, which is more than A1. For the ordinal variables Spearman's ρ will be calculated to determine if there is a correlation between the variable and the test results, and if there is a correlation, how strong it is. Then if a correlation is determined to exist, Pearson's r will be calculated, and a simple linear regression to determine if the correlation relationship is potentially linear and the explained variance with R^2 as Busby (2021: 70) did in her statistical approach.

The other variables related to gender, study abroad experience, English in the workplace, academic discipline, and the vocabulary tests themselves broken into levels, are not ordinal variables since there is no natural ranking or order. One group is not inherently better, bigger, or more than another group. For the vocabulary tests, this is true because only some of the levels are based on the frequency of the words being tested. The combined scores and the academic vocabulary tests, take away the ordinal nature of the 2k, 3k, 5k, and 10k vocabulary tests. For these variables, following the methodology of Busby, first a Kruskal Wallis H test will be carried out to determine if there is a significant difference between the groups. When there are more than two groups, as is the case with the academic disciplines, and if the p-value is 0.05 or below, further pairwise Wilcoxon rank sum tests will be conducted using the Benjamini and Hochberg (1995) correction for multiple testing to show which groups in the sample are significantly different (Busby 2021: 68).

4.1. VLT & FoRTAV Test Results

RQ1: What vocabulary knowledge do master's students enrolled in Austrian universities possess at different frequency levels according to the VLT and FoRTAV?

In Figure 12 the results from the vocabulary tests have been normalized and transformed into boxplots for comparison. The 2k, 3k, AWL, 5k, and 10k sections were originally out of 30 questions. The FoRTAV was originally out of 72 questions. The VLT score comprised of all the test sections added together except for the FoRTAV, thus being out of 150 possible total points. By adding the VLT scores together, as Busby (2021) did in her article, a total VLT score was calculated. This score is not known to measure anything and is not the intended use of the test, but since it was done by Busby (2021), it was done here as a comparison point between the two studies and an additional measurement. Finally, the VLT & FoRTAV scores were added for a grand total of 222 possible points. Following a similar logic as the combined VLT score, the VLT and the FoRTAV scores were combined to give another measurement to evaluate, but this has not been done before thus the results are not conclusive. When the scores were added together, the tests were given weights based on how many items were tested and how many individual questions there were instead of giving the two tests the same weight (i.e., the total score was not out of 200 points where the weight of the FoRTAV and VLT scores would equal 100 points each respectively). The combination VLT + FoRTAV method resulted in a score out of 222 possible points. This method was taken since there have not been any studies yet on the "weight" each vocabulary test and respective test sections should hold.

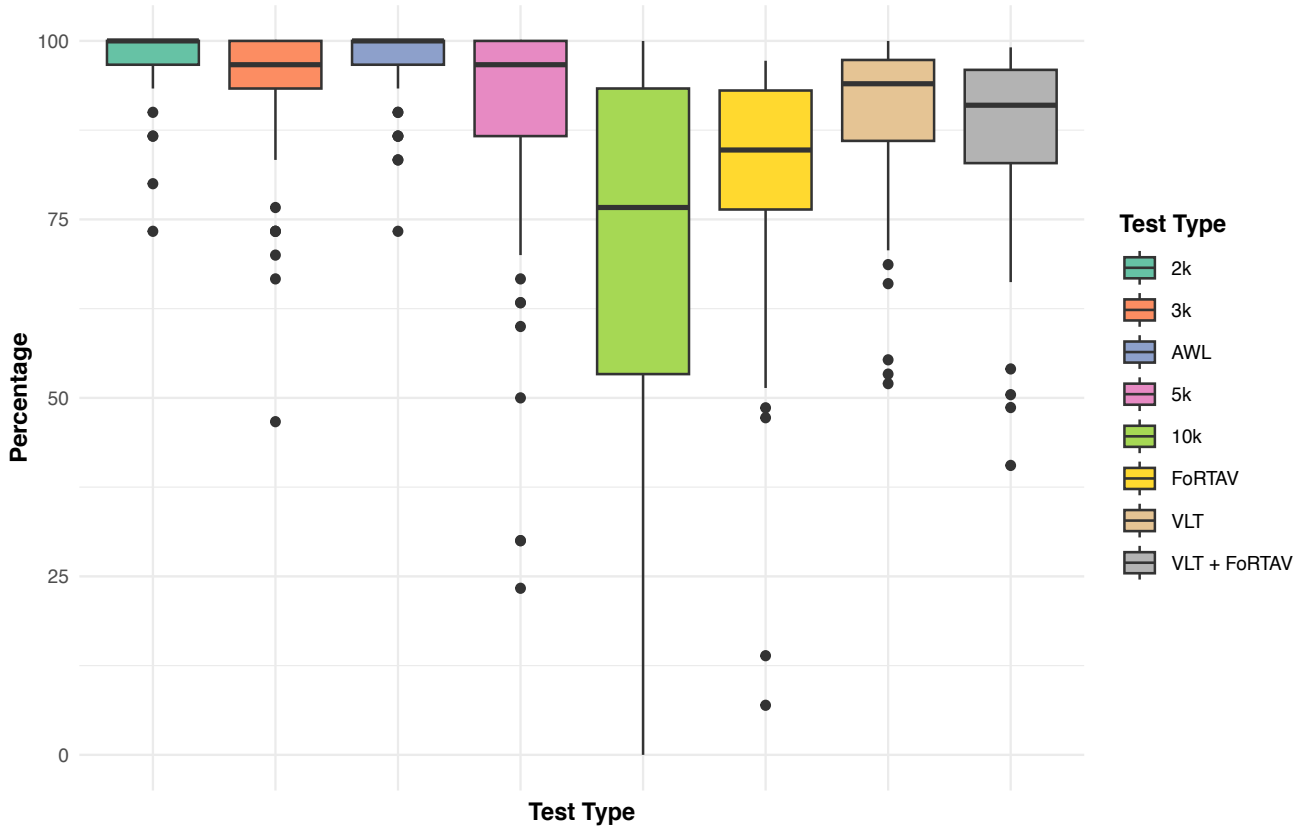


Figure 12. Box plots of scores (%) for each test type

In general, as is well illustrated in Figure 12, the participants performed very well on the different sections of the vocabulary tests. Most of the test sections show evidence of a ceiling effect, meaning the tests were too easy for the participants. This is because the higher frequency levels of the VLT are supposed to be quite simple when a participant has a high English level, as most of the participants have reported having. The ceiling effect is potentially also a side effect of the "passing rate" for each test being so high. The difference between passing and failing is only a few incorrect items in each section, as passing is equal to 26 or above out of 30, and failing is anywhere below 26. The ceiling effect is a good sign for the level of English for the students as this shows that they have mastered, on average, the 2k, 3k, AWL, and 5k sections of the VLT when looking at the mean and median scores. In Table 9, the big 5 statistics are provided along with the median scores for each of the test sections highlighting the differences between the means and medians of the data set.

Table 9. Test results by section including big 5 statistics and mean score

	2k	3k	AWL	5k	10k	FoRTAV	VLT	VLT & FoRTAV
Minimum	22	14	22	7	0	5	78	90
Q1	29	28	29	26	16	55	129	184
Median	30	29	30	29	23	61	141	202
Mean	29.38	28.19	29.2	26.95	21.75	58.72	135.5	194.2
Q3	30	30	30	30	28	67	146	213
Maximum	30	30	30	30	30	70	150	220
Total Possible	30	30	30	30	30	72	150	222

As is evident from looking at the boxplots, the generally high averages, and the Quantile-Quantile plots (which can be found in Appendix IV), the various test results are not normally distributed. The 10k section is the closest to having a normal distribution out of any of the sections; however, upon closer inspection of the Q-Q plot and the trend line (see Figure 13 on the next page), a significant proportion of the points are not on the trend line, making it impossible to claim there is a normal distribution of the results of the present study. Since there is not a normal distribution for a single section of the tests, the types of statistical testing that can be carried out and the conclusions that can be made from the results will be limited.

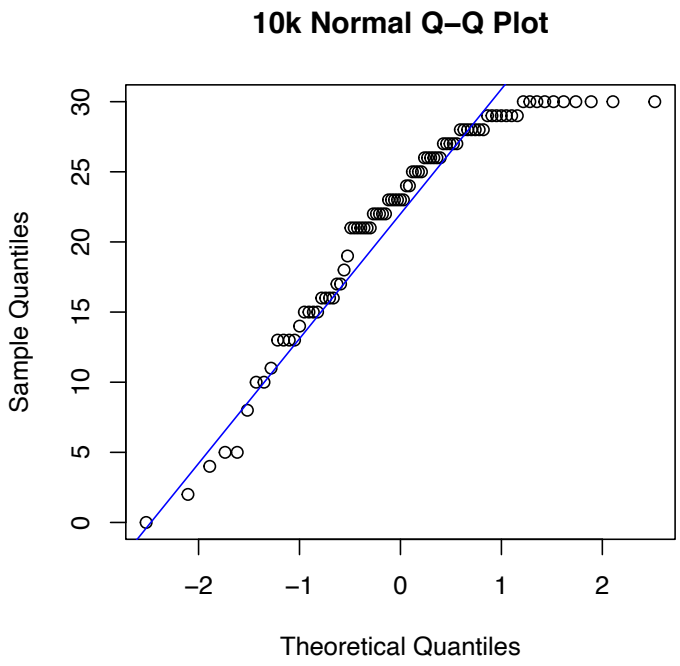


Figure 13. Q-Q plot for residuals of 10k level test

Even though the average scores are quite high and there is a ceiling effect, the percentage of participants who passed the individual levels of the tests varies drastically. As illustrated in Figure 14, 95.3% of the participants passed the 2k level, 90.6% passed the 3k level, 76.5% passed the 5k level, 41.2% passed the 10k level, and 96.5% passed the AWL section.

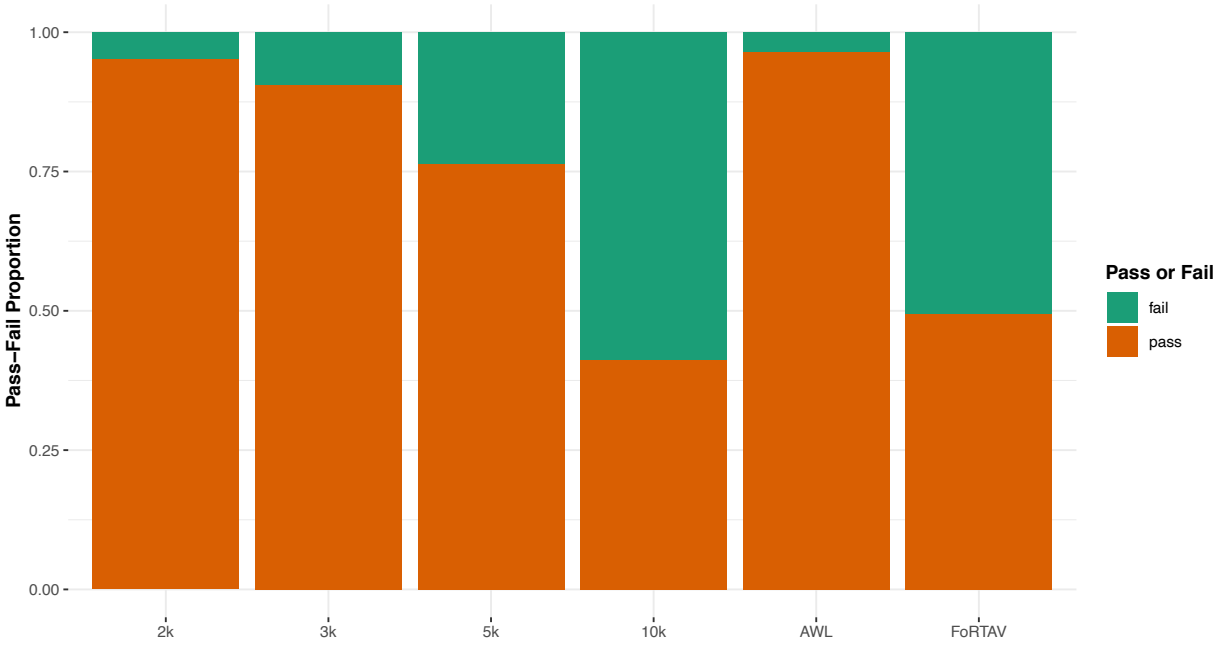


Figure 14. Pass/fail rates for all participants on each test level

This is a good sign for academic knowledge and the higher frequency levels; however, it is concerning that almost 25% of participants failed the 5k level, which is going to be critical for advanced studies in English. Additionally, nearly 60% of the participants failed the 10k level which could be a cause for concern if the higher estimates for vocabulary knowledge are taken as the standard. The FoRTAV does not have a pass/fail cutoff that has been tested, but when the same passing rate it taken as the VLT, slightly less than 50% of the participants failed the test. This is a far lower passing rate than the rate for the AWL section. The potential implications of these passing rates will be discussed in the interpretation of the results

In addition to a statistical analysis for the general results as well as the academic discipline results, the most frequently incorrectly answered items will be presented and compared between groups of participants. In Table 10, this information is shown along with the percentage of participants who were able to answer the item correctly in ascending order and the part of speech (POS) of the item. All the items from the FoRTAV are verbs, as it is specifically a test of academic verbs. The items featured on the list mostly came from questions on the FoRTAV but when they come from another test section, that is indicated in parentheses. Out of the 15 words, 8 of the words came from the FoRTAV, 6 from the 10k level of the VLT, and 1 from the 5k level of the VLT.

Table 10. Top 15 frequently incorrectly answered words

Words	% of participants who answered item correctly	Part of Speech (POS)
Indolent (10k)	35.29%	Adjective
Wily (10k)	38.82%	Adjective
Assume	40%	Verb
Allocate	41.18%	Verb
Derive	49.41%	Verb
Invoke	50.59%	Verb
Induce	50.59%	Verb
Assemble	52.94%	Verb
Conceive	54.12%	Verb
Cringe (10k)	55.29%	Verb
Found	57.65%	Verb
Eagerness (5k)	57.65%	Adverb
Brocade (10k)	57.65%	Noun
Mammoth (10k)	57.65%	Noun
Squirm (10k)	57.65%	Verb

Now that the initial results have been summarized, the next subsection will address the research questions focused on the potential differences between academic disciplines. Initially the results regarding Biglan's discipline structure will be examined and then the results regarding Holland's approach.

4.2. Biglan Groups and Vocabulary Scores

RQ2a: In what ways does vocabulary knowledge vary between Biglan's academic groups when considering master's students in Austria?

To determine if groups of participants when sorted on Biglan's groupings of academic disciplines performed differently on the VLT and FoRTAV the participant's results were run through a Kruskal Wallis H Test which will determine if there is a statically significant difference between the groupings. Each individual section has been tested as well as the VLT sections combined, and the VLT and FoRTAV scores combined. Figures 15-22 below illustrate the results separated by academic grouping. The x-axes show the 4 disciplines with the applied results being shown in the first column, then the hard results, pure results, then lastly, the soft results on the far-right side. The figures show the scores as dots with slight horizontal jiggering to better show the data points that are concentrated in clusters. On the y-axes, there is a range of test scores given and the range is dependent on the results optimizing showing detail in the higher ends of the results as well as zooming out far enough to view the full range of results. In each figure, there are traditional box plots drawn over the four distributions, but since the distributions were already determined to not be normal, an additional projected non-parametric distribution is added to the plots to demonstrate a possible potential distribution of the results. The Kruskal Wallis statistics are provided at the top of each figure, however the most relevant value to answer the research question is the p -value. If the p-value is 0.05 or below there is a statistically significant difference between the group means.

The results are spectacularly anticlimactic, with none of the test sections nor academic groups showing a significant difference between the means. All the p-values are well above the 0.05 threshold, therefore further investigations into the academic groupings are most likely not needed. The plots illustrate well the ceiling effect that was present for the 2k, 3k, 5k, and AWL sections of the VLT as none of the boxplots have whiskers at the top end meaning the results were clustered at the tops of the charts. An interesting observation that can be made is that it is all the results that are insignificant. At least there is consistency in the insignificance of the groups in their entirety, and not just selectively.

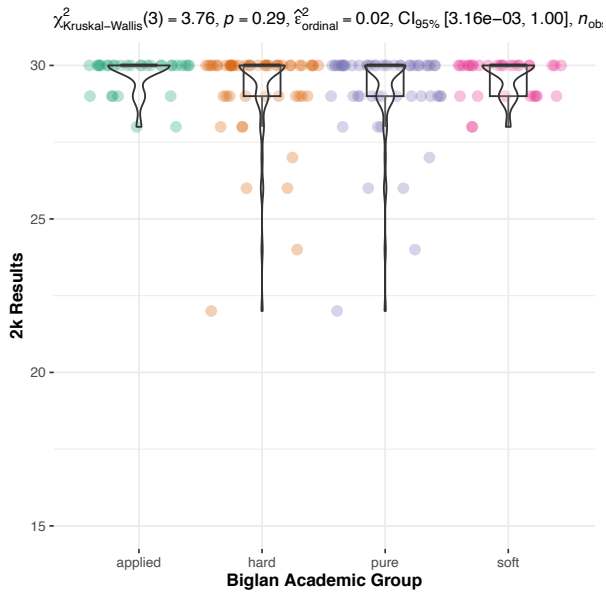


Figure 15. Boxplots for Biglan's groups vs. 2k results

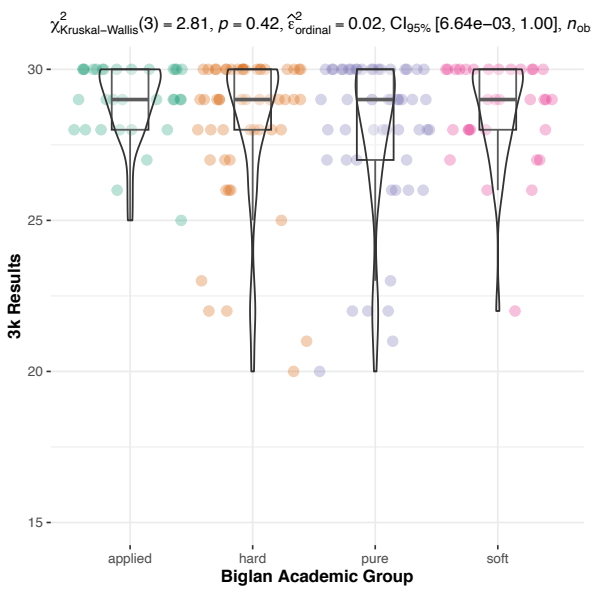


Figure 16. Boxplots for Biglan's groups vs. 3k results

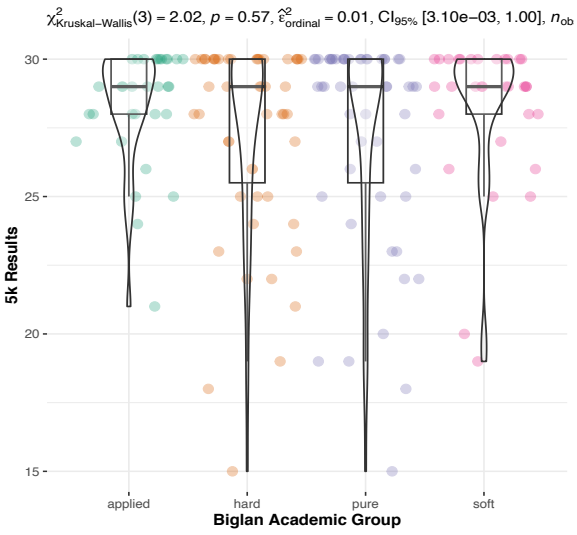


Figure 17. Boxplots for Biglan's groups vs. 5k results

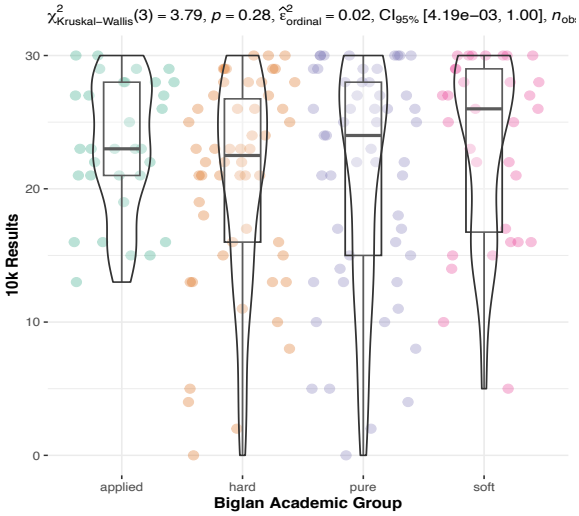


Figure 18. Boxplots for Biglan's groups vs. 10k results

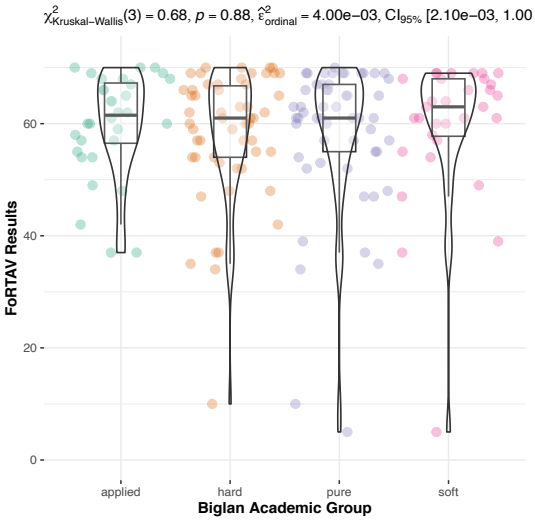


Figure 19. Boxplots for Biglan's groups vs. FoRTAV results

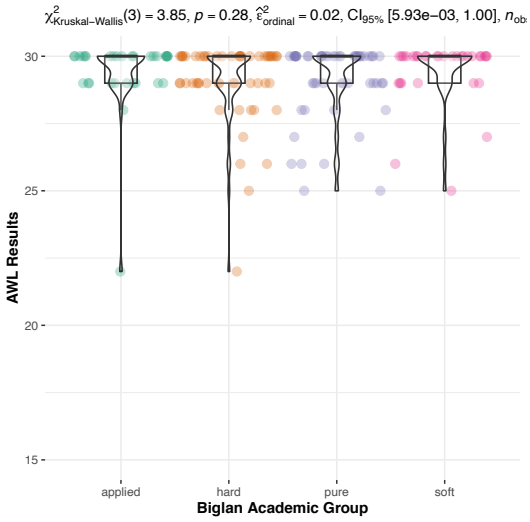


Figure 20. Boxplots for Biglan's groups vs. AWL results

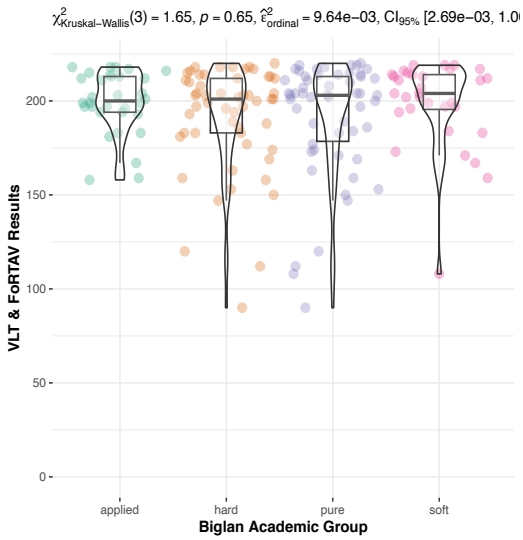


Figure 21. Boxplots for Biglan's groups vs. combined VLT & FoRTAV results

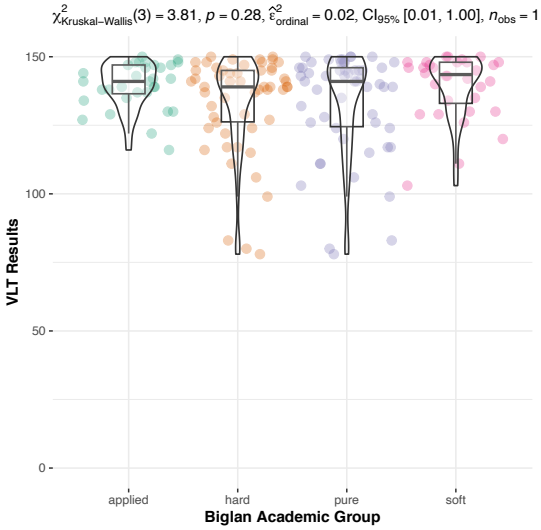


Figure 22. Boxplots for Biglan's groups vs. VLT results

In Figure 23 the percentage of participants who passed each level of the VLT and the FoRTAV split into the academic disciplines are presented. This is again an interesting measure to consider due to the near perfect scores needed for a participant to be considered to have mastery of the vocabulary level or section. At the 2k, 3k, and 5k levels of the VLT, there appears to be a consistent trend where more participants in the applied and soft groups pass the levels whereas the participants in the pure and hard groups have a lower percentage of students passing the tests. At the 10k level, the soft group had the highest passing rate at a bit over 50%, the pure and applied groups had similar passing rates to each other at around 40%, and lastly, the hard group has the lowest passing rate in this frequency range at around 35%. For the AWL

the four groups have remarkably consistent and high passing rates. Lastly, the four groups have slightly different passing rates on the FoRTAV, but the differences are not as drastic as the differences at the 3k, 5k, and 10k levels. From the figure, it is also clear that the soft discipline consistently tied for the highest passing rate or was the only discipline with the highest passing rate.

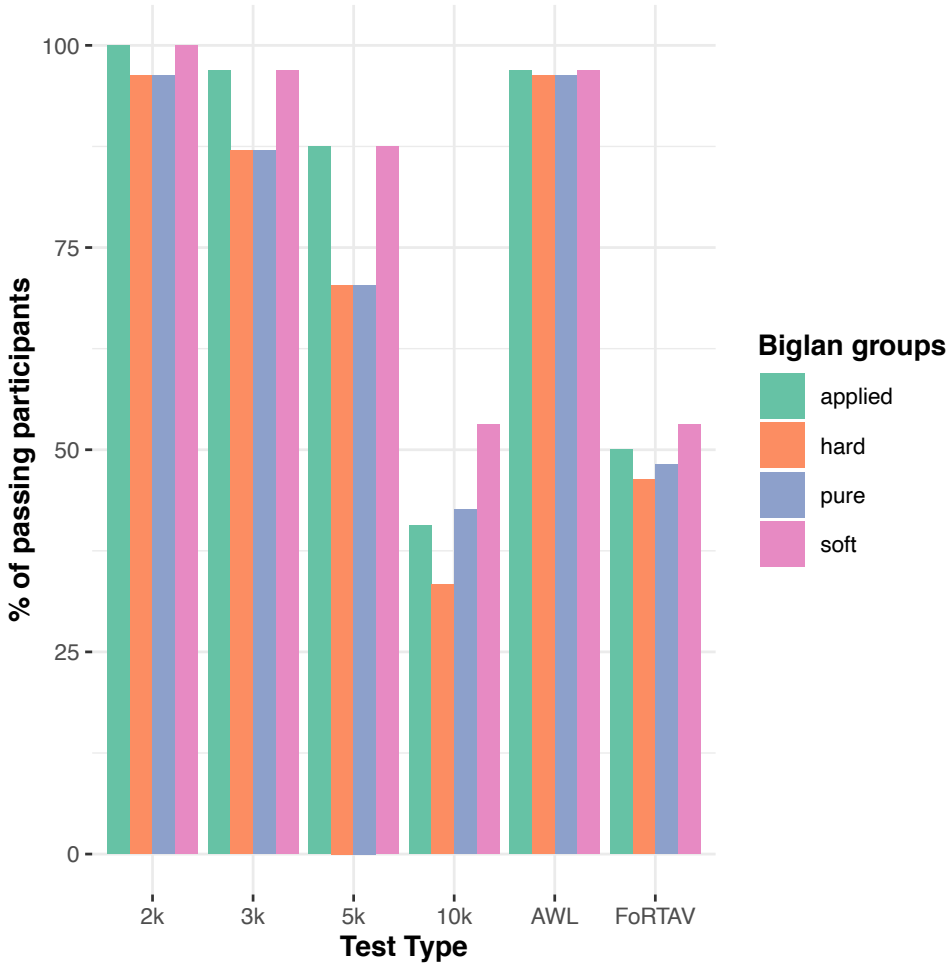


Figure 23. Passing rates as a percent of all students on each test section split into Biglan's groups

In the next table, Table 11, the 15 most frequently incorrectly answered items for all the participants are listed in the left most column along with the 10 to 11 most frequently incorrectly answered items for each of Biglan's four academic groups. The number of items vary depending on the list since sometimes the 10th most frequently incorrectly answered item was tied with another item and to remain consistent, both items have been included on the list. Again, most of the items are from the FoRTAV but when they are not, the test section is indicated in paratheses.

Table 11. Most frequently incorrectly answered items for all participants and Biglan groups

All Participants	**Hard**	**Soft**	**Pure**	**Applied**
Indolent (10k)	Indolent (10k)	Assume	Indolent (10k)	Indolent (10k)
Wily (10k)	Wily (10k)	Induce	Wily (10k)	Assume
Assume	Allocate	Wily (10k)	Allocate	Wily (10k)
Allocate	Assume	Allocate	Assume	Induce
Derive	Invoke	Derive	Derive	Allocate
Invoke	Assemble	Squirm (10k)	Invoke	Brocade (10k)
Induce	Derive	Invoke	Cringe (10k)	Conceive
Assemble	Brocade (10k)	Found	Mammoth (10k)	Eagerness (5k)
Conceive	Cringe (10k)	*Infer*	Assemble	*Infer*
Cringe (10k)	*Trigger*	Conceive	***Amend***	*Trigger*
Found		***Attribute***		
Eagerness (5k)				
Brocade (10k)				
Mammoth (10k)				
Squirm (10k)				

Note: words in *italics* are not featured in the top 15 words from all participants, words in **bold** are completely unique to the list when compared with the others

If the four groups behaved in the same way, the same words should appear on both lists in a similar order. However, that is not the case as there are items on the Biglan discipline group lists that are not on the original list. The items that are italicized did not appear in the general 15 most frequently incorrectly answered items, and if it is italicized and boldened the item is unique to only one discipline's list. The hard discipline list includes one item not featured on the general list. The soft discipline list includes two items that are not featured on the general list, one being unique to the soft list. The pure list has one item not found on the general list and it is unique to the pure list. Lastly, the applied list has two items not found on the general list. Fifteen items were given for the general list while ten items were given for the discipline lists to leave a slight buffer as the difference between the 10th and 11th place probably isn't significant but the difference between 10th and 16th place seems worth mentioning. In general, *wily* and *indolent* remain near the tops of all of the lists, except for the soft list where *indolent* does not appear at all in the top 11 most frequently incorrectly answered items.

4.3. Holland Groups and Vocabulary Scores

RQ2b: In what ways does vocabulary knowledge vary between Holland's academic groups when considering master's students in Austria?

The same approach was taken for statistically determining the significance of Holland's groupings on the vocabulary test results as was taken to determine the significance of the Biglan groupings. Figures 24 – 30 show the results on the VLT and FoRTAV split by Holland's academic groupings and split into the various levels and sections. Again, Kruskal Wallis H tests were carried out, and again none of the p-values were found to be statistically significant, or below 0.05. This means that there does not appear to be any significant difference between the student vocabulary scores on the VLT and FoRTAV, at any level, according to Holland's groupings. So, across the enterprising, investigative, realistic, and social groupings, the mean test scores are similar. It is important to note that the sample size for the realistic and enterprising groups was lower than 25 participants, so a future analysis could aim for a study with more participants. However, realistically, there seems to be such a small difference that more participants might not make any difference.

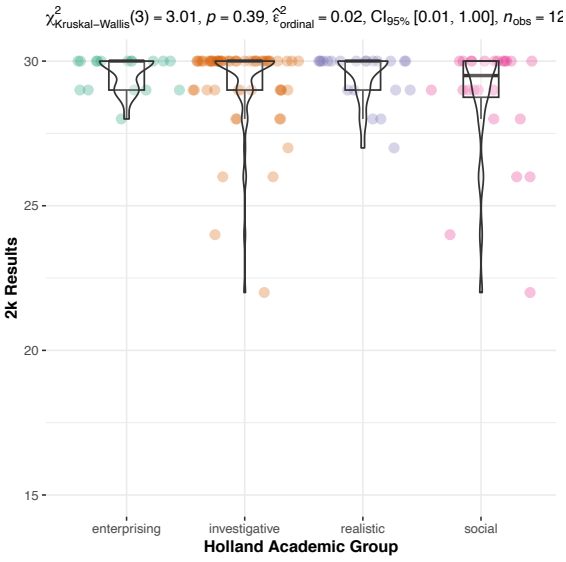


Figure 24. Boxplots for Holland's groups vs. 2k results

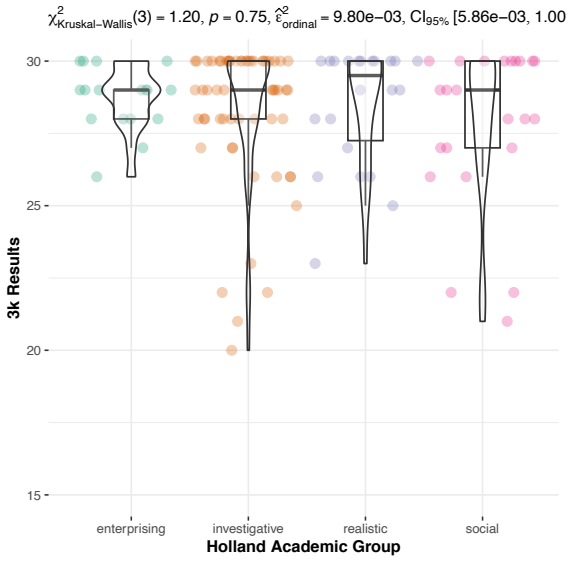


Figure 25. Boxplots for Holland's groups vs. 3k results

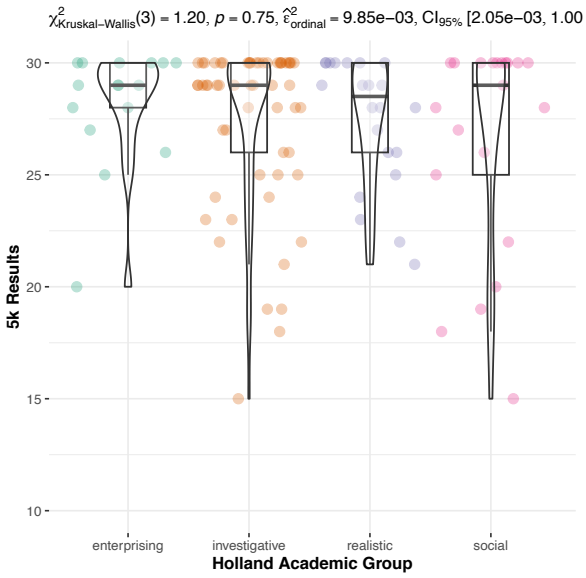


Figure 26. Boxplots for Holland's groups vs. 5k results

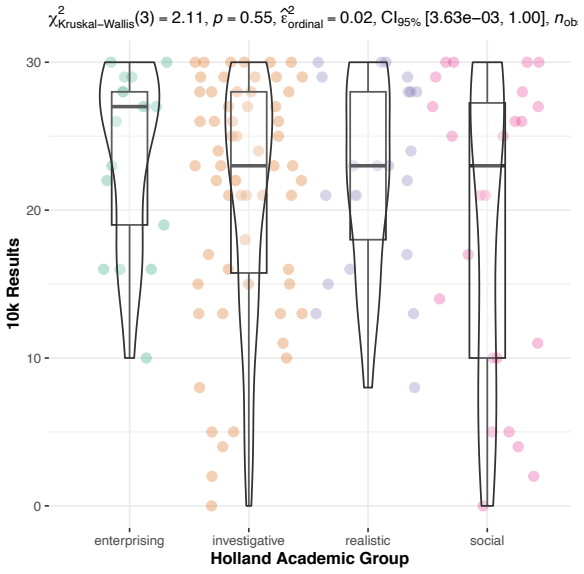


Figure 27. Boxplots for Holland's groups vs. 10k results

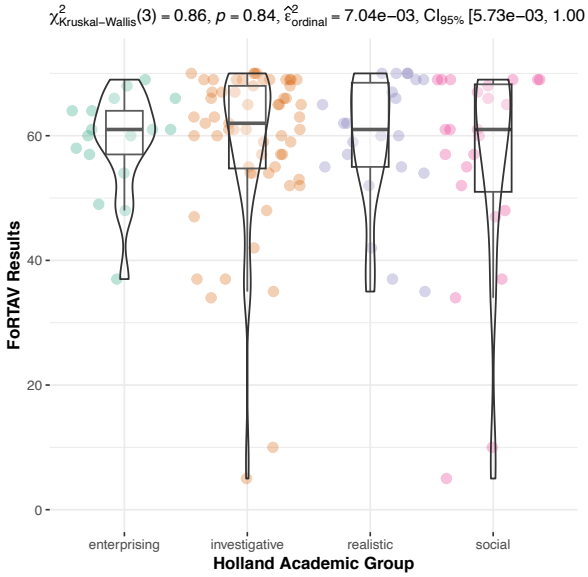


Figure 28. Boxplots for Holland's groups vs. FoRTAV results

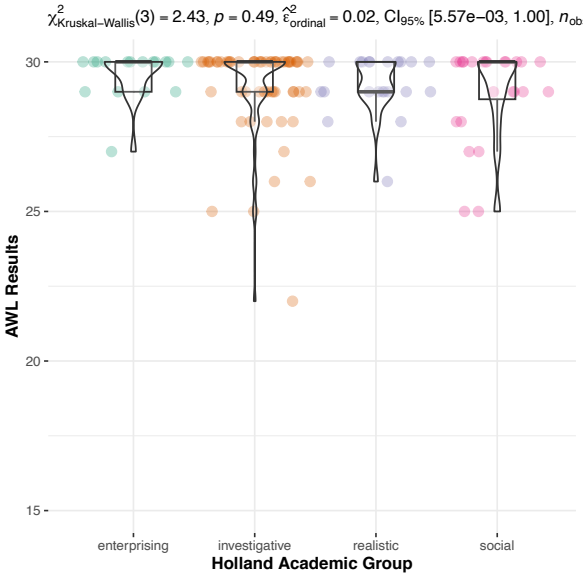


Figure 29. Boxplots for Holland's groups vs. AWL results

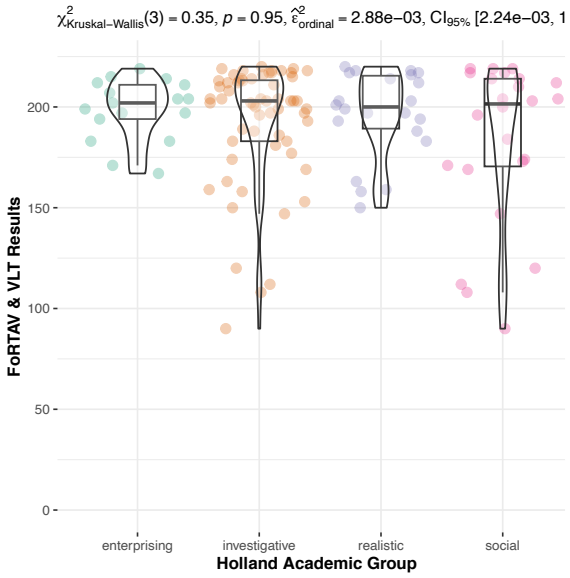


Figure 30. Boxplots for Holland's groups vs. combined VLT & FoRTAV results

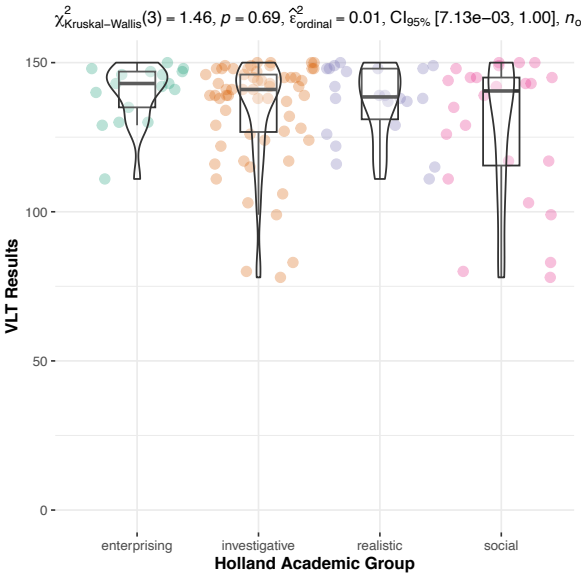


Figure 31. Boxplots for Holland's groups vs. VLT results

As was done previously with the Biglan groupings, the percentage of participants who passed the various sections will be compared between Holland's academic groupings. In Figure 32 (on the next page), the results are shown, with a surprisingly consistent pattering appearing in the results between the 3k and 5k test sections. In these sections, the enterprising group has the highest rate of passing participants, meanwhile the realistic group has the second highest passing rate, then the investigative grouping, and finally the social grouping. The 2k level and AWL section have similar results with the enterprising and the realistic groups showing a 100% pass rate, meaning all the participants in these groups passed these two sections. Then the investigative has the next highest rate, and again the social group has the lowest rate. The 10k results show the enterprising group having a much higher passing rate than the other groups at nearly 60%, with the other three groups close to each other at around 40% of the participants passing. Lastly, on the FoRTAV section, the enterprising group had the lowest passing rate, a marked change from the trend presented in the other sections. The investigative group has the highest passing rate at just above 50%, then the realistic group at exactly 50%, and in third is the social group with somewhere around 40% of participants passing the FoRTAV.

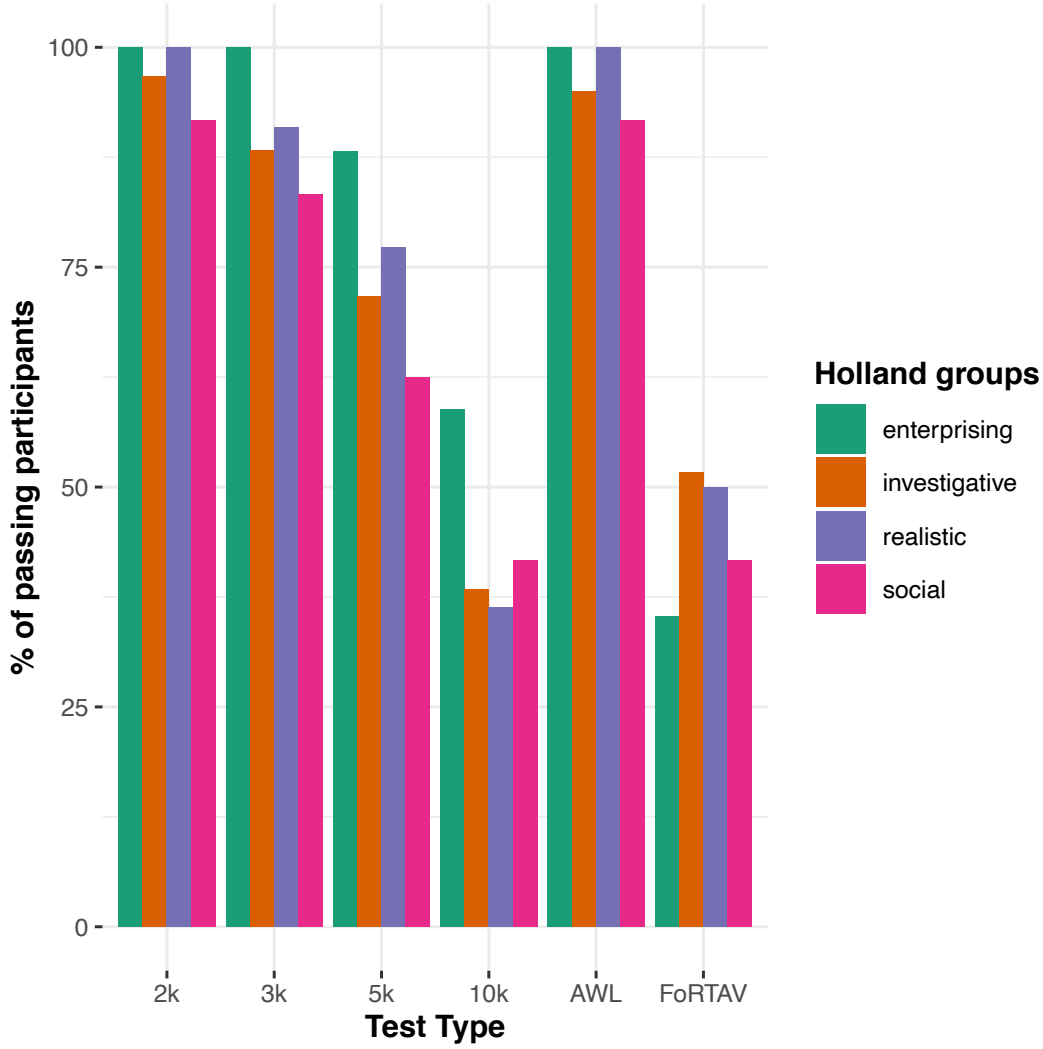


Figure 32. Passing rates as a percent of all participants on each test section split into Holland's groups

The last measurement that will be used to comment on differences present between the academic groupings according to Holland is by looking at the most frequently incorrectly answered items from the tests. As was done with the Biglan groupings, the top 15 most frequently incorrectly answered items were determined from all participants, and then the top 11 to 17 items were taken from the Holland groups. The number of items show for each group is again different due to shared ranks. The enterprising group has a lot of items due to this happening, but a comparison is still an interesting tool. The realistic group has two words on the list that did not appear on the general list with one of them being unique to the realistic list. The investigative list has one item that was not found on the general list. The social list has four items that were not found on the general list. Lastly, the enterprising list has 6 items that were not found on the general list, four of those being unique to the enterprising list.

Table 12. Most frequently incorrectly answered items for all participants and Holland's groups

All Participants	**Realistic**	**Investigative**	**Social**	**Enterprising**
Indolent (10k)	Indolent (10k)	Indolent (10k)	Derive	Induce
Wily (10k)	Wily (10k)	Wily (10k)	Indolent (10k)	Assume
Assume	*Trigger*	Assume	Wily (10k)	Allocate
Allocate	Brocade (10k)	Allocate	Assume	Indolent (10k)
Derive	Invoke	Assemble	Cringe (10k)	Wily (10k)
Invoke	Allocate	Invoke	*Attribute*	Eagerness (5k)
Induce	Assemble	Brocade (10k)	Mammoth (10k)	Conceive
Assemble	Assume	Derive	Squirm (10k)	***Comprise***
Conceive	Eagerness (5k)	*Trigger*	Assemble	*Infer*
Cringe (10k)	Cringe (10k)	Cringe (10k)	Invoke	***Throttle*** (10k)
Found	***Viscount*** (10k)	Mammoth (10k)	*Amend*	Squirm (10k)
Eagerness (5k)	Squirm (10k)		*Append*	***Occupy***
Brocade (10k)	Found		*Infer*	*Attribute*
Mammoth (10k)	Induce		Allocate	***Sustain***
Squirm (10k)	Derive			Invoke
				Found
				Derive

Note: words in *italics* are not featured in the top 15 words from all participants, words in **bold** are completely unique to the list when compared with the others

Again, if the four groups really were identical in their behavior, the same items should appear on the lists in a similar order. *Indolent* and *wily* appear at or near the top of all the lists, which is again an interesting observation. What is also interesting is that *eagerness*, an item from the 5k vocabulary test, has made it onto several of the lists, seemingly to be the most difficult 5k word for many of the participants.

4.4. English Medium Education & Vocabulary Scores

RQ3a: How does vocabulary knowledge compare between master's students with different proportions of English as part of their curriculum in Austria?

To determine the potential impact of the amount of English Medium Education exposure that the students experience in their master's programs the students were asked in the questionnaire to rank on a scale from never (0 classes) to always (5/5 classes) how often they have experienced teaching in English, reading in English, writing in English, and speaking/presenting over the course of their master's studies. These scores were taken on their own as well averaged to come up with an "average EME score". Additionally, the students were asked to report how many ECTS or credits they had completed so far. This was to reflect the logic and match the methodology of Busby's approach where her assumption is the farther along a student is in their students, the more English they will have been exposed to. She compares the results on the sections of the VLT to the number of semesters the students have been attending university (Busby 2021: 69). Therefore, the same procedure will be followed in this study.

To statistically determine the potential correlation between the EME exposure variables and the test results, a series of correlation tests were carried out using Spearman's ρ. The results from all the VLT sections and FoRTAV correlated to the teaching, writing, presenting, reading, combined scores, and number of semesters, are presented in Table 13 below. Any result that had a Spearman's ρ of ±0.1 or more is to be considered indicative of a small correlation (Baumann 2022a: slide 4). Spearman's ρ can only tell us the direction of the relationship (be it positive or negative) and the strength of the relationship, but it cannot tell us the type of relationship (i.e. linear, exponential, etc.).

Table 13. Spearman's ρ for all potential correlation combinations

	FoRTAV & VLT	FoRTAV	VLT	10k
Teaching	0.031	0.077	-0.046	-0.041
Writing	0.076	*0.123*	0.000	-0.004
Presenting	0.080	*0.120*	0.002	0.000
Reading	*0.140*	*0.177*	0.076	0.066
EME Combined	*0.114*	*0.167*	0.025	0.013
Semester Progress	0.092	0.070	*0.112*	*0.153*

Note: Items are italicized to denote a small correlation therefore warranting further investigation

From the results summarized in Table 13, the combined scores of the VLT & FoRTAV appear to have a small correlation with the Reading and Combined exposure scores. The FoRTAV scores alone have a small correlation to the writing, presenting, reading, and combined EME exposure scores. The VLT score has a small correlation with the semester progress. Lastly, the 10k scores have a small correlation with the semester progress as well. It is interesting that the combined scores and FoRTAV scores are correlated with a few EME

exposure factors, meanwhile the VLT score and 10k score are correlated with the duration of one's studies and seemingly nothing else. The fact these correlations do not conclude in remotely similar results would seem to hint at the fact that they are measures of different phenomena instead of contributing to one measurement of the same phenomenon.

Since some of the relationships were found to have a small correlation, the next step taken was determining if this relationship was perhaps a linear one. By performing correlation tests using Pearson's r instead of Spearman's ρ it can be determined if the relationship is better described with a linear correlation line. If the correlation statistic for Pearson's r is higher than the correlation statistic in the Spearman's ρ tests, it can be assumed that a linear model is a better fit and better representation of the data. In the following figures, Figure 33 – 40, the correlations determined to be of interest (or at least having a small correlation) in Table 13, have been plotted. In these plots, the test scores are on the y-axes and the variables found to be correlated are on the x-axes. A correlation line has also been added to all the plots to show the potential linear relationship as well as a margin of error where the "real" line could potentially actually exist. In the Pearson's r statistical test, there will be a 95% confidence interval given for the potential slope of this linear relationship. If the interval crosses 0, then the true slope of the relationship may lay anywhere within this range, thus the true correlation between the variables cannot be confirmed to be (with 95% confidence) positive or negative.

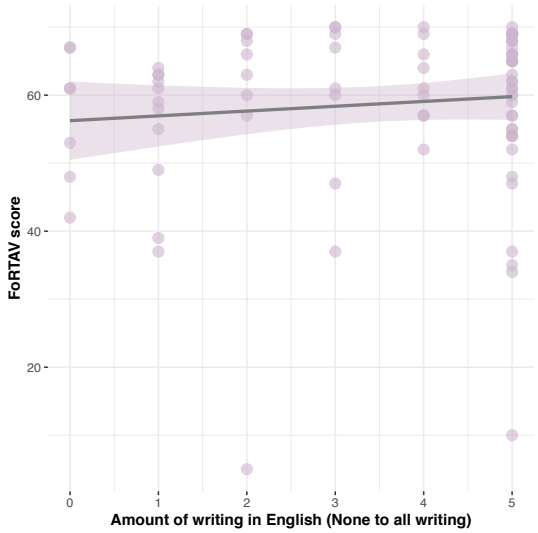


Figure 33. Amount of writing in English vs. FoRTAV scores

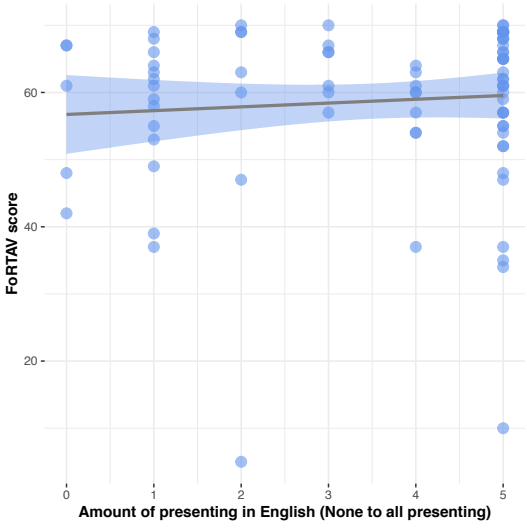


Figure 34. Amount of presenting in English vs. FoRTAV scores

Figures 33 and 34 (shown above) represent the correlations between the FoRTAV scores on the y-axes and the amount of writing in English score on the left chart and the amount of presenting in English on the right chart on the x-axes. After calculating Pearson's r for both relationships, the correlation coefficient for writing in English with the FoRTAV score was determined to be 0.105 with a confidence interval of -0.111 and 0.312. The correlation

coefficient here would suggest a slight positive relationship; however, as previously mentioned, the confidence interval crosses 0, therefore, the linear correlation is not significant. For the presenting in English and FoRTAV score correlation, the correlation coefficient was determined to be 0.083 with a confidence interval of -0.133 and 0.291 thus also crossing 0. This means both relationships are not significant with the current data set when looking at Pearson's r and the confidence intervals. A relationship may still exist between the two variables, as shown by Spearman's ρ but the linear model is most likely not the best fit to describe the nature of the relationships.

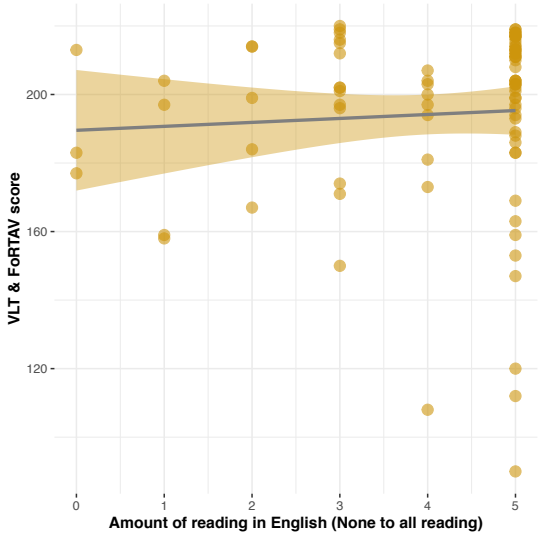


Figure 35. Amount of reading in English vs. combined VLT & FoRTAV score

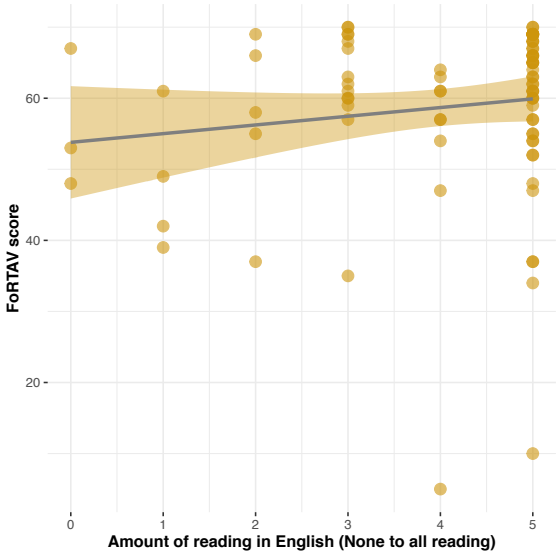


Figure 36. Amount of reading in English vs FoRTAV score

Figures 35 and 36 above show the relationships between the amount of reading present in the participant's master's programs on the x-axes and the results on the VLT & FoRTAV scores totaled for the y-axis on the chart on the left side and the FoRTAV score alone on the right side. Pearson's r for the combined scores is 0.061 suggesting a very slight positive relationship however, the confidence interval of -0.154 to 0.270 crosses 0 and therefore is not significant. For the FoRTAV score taken alone, the correlation coefficient is slightly higher at 0.143, but the confidence interval also crosses at 0 since it is from -0.073 to 0.345 so the relationship is not significant. Again, this does not mean that there is no relationship or correlation between the items, but if there is one, it is most likely not a linear relationship.

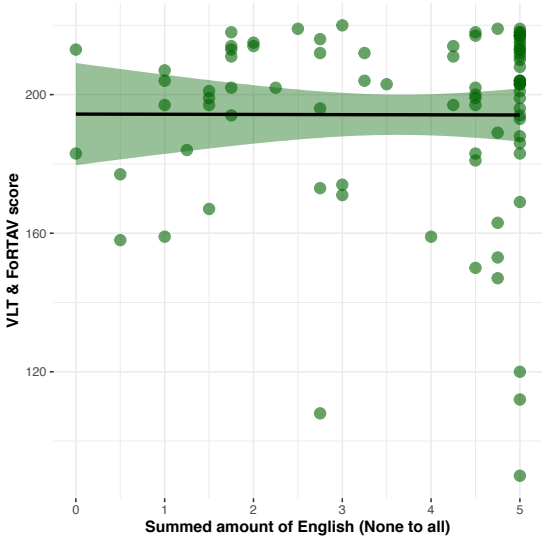


Figure 37. Summed EME vs. combined VLT & FoRTAV score

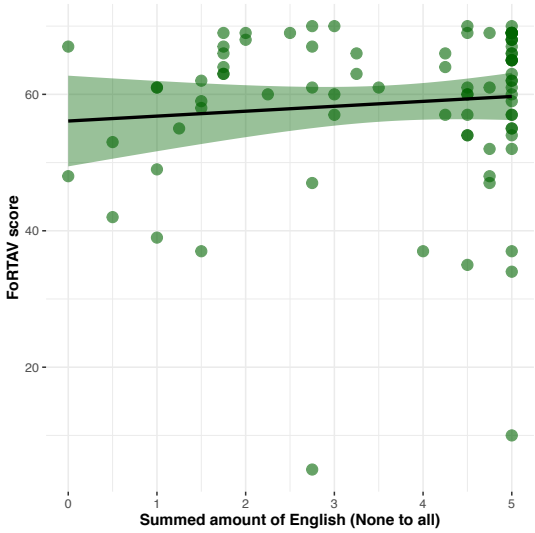


Figure 38. Summed EME vs. FoRTAV score

In Figures 37 and 38 above, the summed exposure to English scores are plotted on the x-axes of both charges along with the results from the VLT and FoRTAV summed for the y-axis on the left chart, and the FoRTAV score alone on the y-axis of the right chart. Pearson's r for the combined English exposure score and the combined VLT & FoRTAV test scores was determined to be -0.003, suggesting there is the slightest negative correlation between the two variables. This is a drastically different correlation coefficient to Spearman's ρ for the same relationship and thus why it is important to look at both values. Spearman's ρ for the same relationship was determined to be 0.114 which would suggest a small positive correlation between the variables. The confidence interval for Pearson's r crosses 0 as it is from -0.216 to 0.210. Combined with the fact that the Spearman's ρ value was much higher and a positive relationship, the relationship that does exist is not accurately represented by a linear model and a linear relationship is not the best fit for describing the relationship. For the relationship between the FoRTAV score and summed English exposure, Pearson's r was determined to be 0.093 but the confidence interval again crossed 0 going from -0.122 to 0.301. Neither of these relationships are best described by a linear relationship but the potential for a relationship to exist is still present.

The last factor included in this research question that will be investigated relates to the participants' progress through their studies in comparison to the results on the different sections of the test. As already mentioned, the results from Spearman's ρ suggest that the VLT scores and the results from the 10k level are correlated to the student's progress. Figures 39 and 40 below represent these relationships. Pearson's r for the 10k score and the progress through participant's studies (shown in Figure 39) was determined to be 0.154, which would be a

positive correlation. The confidence interval overlaps with 0 since it goes from -0.064 to 0.358, meaning it cannot be said that the relationship is not actually a slight negative relationship. Pearson' r for the VLT score and progress through studies (shown in Figure 40) is 0.159 but the confidence interval crosses 0 since it ranges from -0.059 to 0.362.

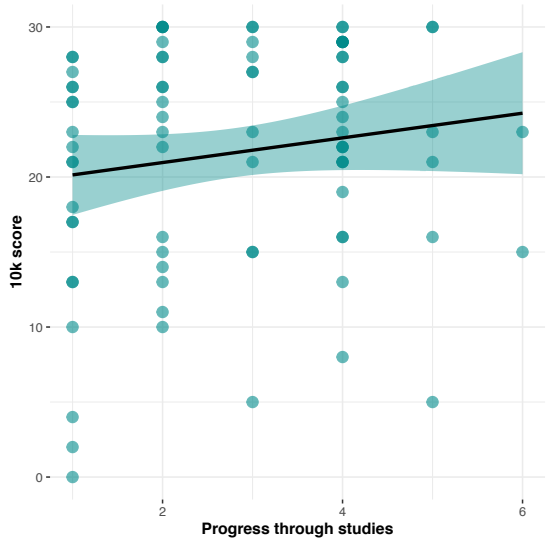


Figure 39. Number of semesters completed vs. 10k score

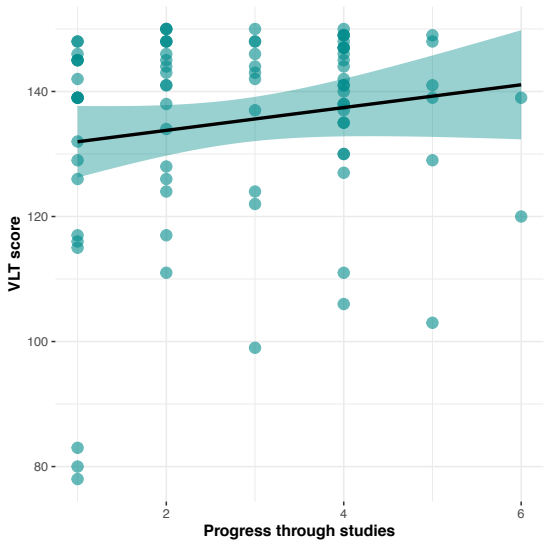


Figure 40. Number of semesters completed vs. VLT score

As already mentioned, comparing Spearman's ρ to Pearson's r can be an interesting tool to determine if the relationship between two variables appears to be linear in nature or would be better represented by a different type of relationship. In the above examples, none of them were shown to be significant using Pearson's r since all the confidence intervals crossed 0. However, due to Spearman's ρ it is clear some of the variables at least have a small relationship to the test results. In the next section, more variables will be investigated that did not relate to the amount of English present in the master's programs nor the duration of the time spent studying.

4.5. Other Factors & Vocabulary Scores

RQ3b: What other factors might impact the vocabulary knowledge of master's students in Austria?

In this section, further variables will be invested as they relate to the participant's results on the various vocabulary tests. Some of the variables are ordinal (i.e., age and number of languages regularly spoken), meanwhile, others are not (i.e., gender and certain yes/no questions the participants were required to answer). As already seen with other variables, when the variable is ordinal Spearman's ρ will be calculated, then if at least a small relationship is found (above 0.1), Pearson's r will be calculated as a correlation coefficient as well as the

confidence interval. Originally the plan was if the confidence interval does not include 0, a linear regression would be executed to measure the R^2 value, or effect size, and explained variance; however, since the test results are not normally distributed (as explained in Section 4.1) it does not make sense to execute simple liner regressions as it is a prerequisite for linear regressions (Baumann 2022b: slide 7). If a variable is not ordinal, the Kruskal Wallis test will be performed and if the p-value is determined to be below 0.05 further statistical testing will be carried out. Table 14 shows the initial results from the Spearman's ρ calculations and Kruskal Wallis p-values.

Table 14. Spearman's ρ or Kruskal Wallis p-value for all potential correlation combinations

	Variable	FoRTAV & VLT	FoRTAV	VLT	10k
Ordinal	Age	0.023	-0.090	*0.104*	*0.153*
	Number of Languages	0.069	0.056	0.059	0.052
	English Level	*0.353*	*0.319*	*0.334*	*0.335*
Non-ordinal	English at work	0.681	0.417	0.919	0.933
	Study Abroad	0.578	0.996	0.420	0.265
	Gender	0.304	0.228	0.613	0.416

Note: Items are italicized to denote at least a small correlation or significant p-value therefore warranting further investigation

None of the non-ordinal variables were found to have significant p-values including gender, the presence of English at work, nor study abroad experiences. On the other hand, a few of the ordinal values were determined to have at least a small correlation coefficient, as indicated by the italics, including age, number of languages used regularly, and the participants' self-ranked English levels. Further investigations into these relationships will be completed in the rest of this section.

The relationship between the participants' VLT scores and their self-ranked English levels is depicted on the next page in Figure 41. Spearman's ρ was 0.334 and Pearson's r is 0.450, which suggests a medium positive relationship when looking at either of these correlation coefficients. For Pearson's r, the confidence interval is from 0.262 to 0.605, suggesting the relationship is statistically robust and indeed indicative of a medium-strength linear relationship as Pearson's r is higher than Spearman's ρ.

In Figure 42, (below) the relationship between the combined VLT & FoRTAV scores and the participants' self-ranked English levels are compared. Spearman's ρ was 0.353 meanwhile Pearson's r was determined to be 0.452 with a confidence interval from 0.265 to 0.607. This again shows that there is a medium-strength relationship between the two variables and that the linear correlation coefficient predicts a stronger relationship between the two variables than a general correlation coefficient.

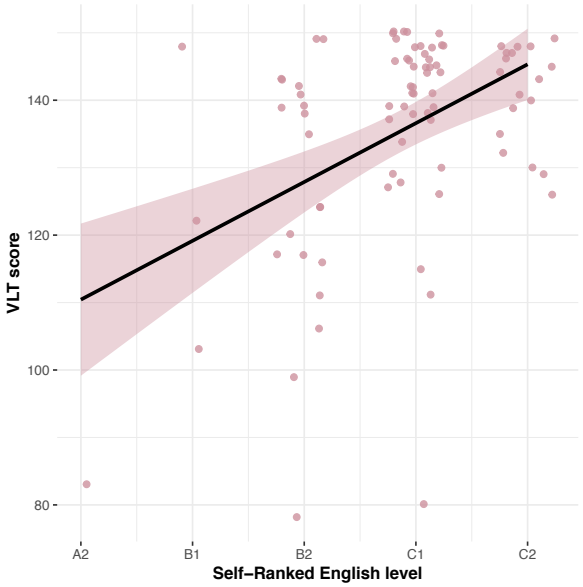


Figure 41. Self-ranked English level vs. VLT score

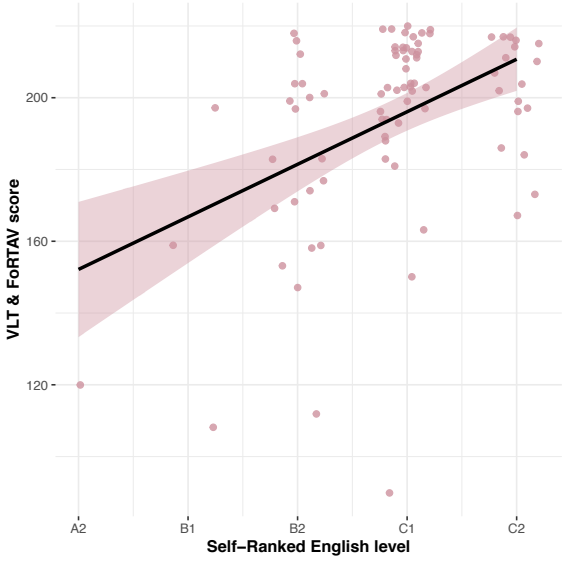


Figure 42. Self-ranked English level vs. combined VLT &FoRTAV scores

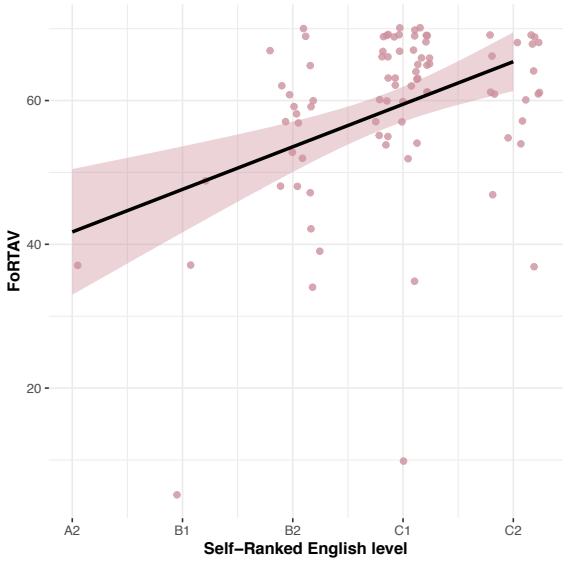


Figure 43. Self-ranked English level vs. FoRTAV score

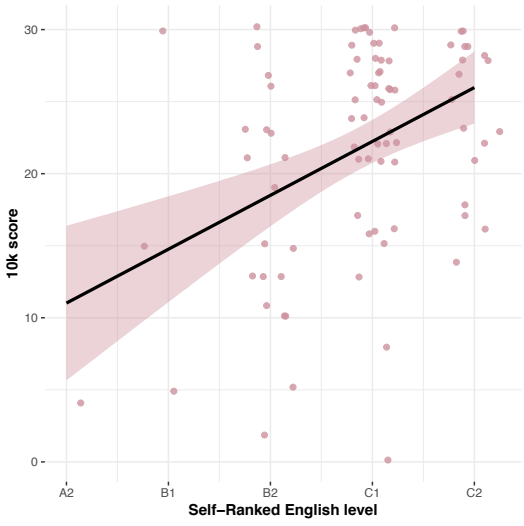


Figure 44. Self-ranked English level vs. 10k scores

The next figure, Figure 43, compares the results on the FoRTAV with the English levels of the participants. Spearman's ρ was 0.319 and Pearson's r was determined to be 0.404 with a confidence interval from 0.208 to 0.568. It is interesting to compare the correlation coefficients

between the VLT combined scores and FoRTAV scores as they are quite similar, showing a medium strength relationship in both cases. Furthermore, the relationship of the FoRTAV to the proficiency levels seem to be better described by Pearson's r leading to the conclusion that the relationship is probably closer to a linear relationship than another type of relationship.

The last chart including the self-ranked English levels of the participants is featured in Figure 44. The levels are compared with the results from the 10k section of the VLT as it had the widest score distribution of all the levels and includes the least frequent words. Spearman's ρ was 0.335 and Pearson's r was determined to be 0.414, both figures indicating a medium strength positive correlation. The confidence interval for Person's r is from 0.22 to 0.576 which does not cross 0 indicating the correlation coefficient is statistically robust and positive.

It should not be shocking that a relationship was consistently found between the student's self-ranked levels and their performance on the tests at multiple levels. The correlation between these variables and the results is most likely a reflection of the ability of the VLT and FoRTAV to truly measure one's vocabulary abilities in conjunction with the abilities of the students to report their English level somewhat accurately. This topic will be further elaborated on in the discussion and potential implications of the results. It is also important to mention that most participants considered themselves to have a B2, C1 or C2 level in English therefore the results concerning the lower proficiency levels are not fully validated.

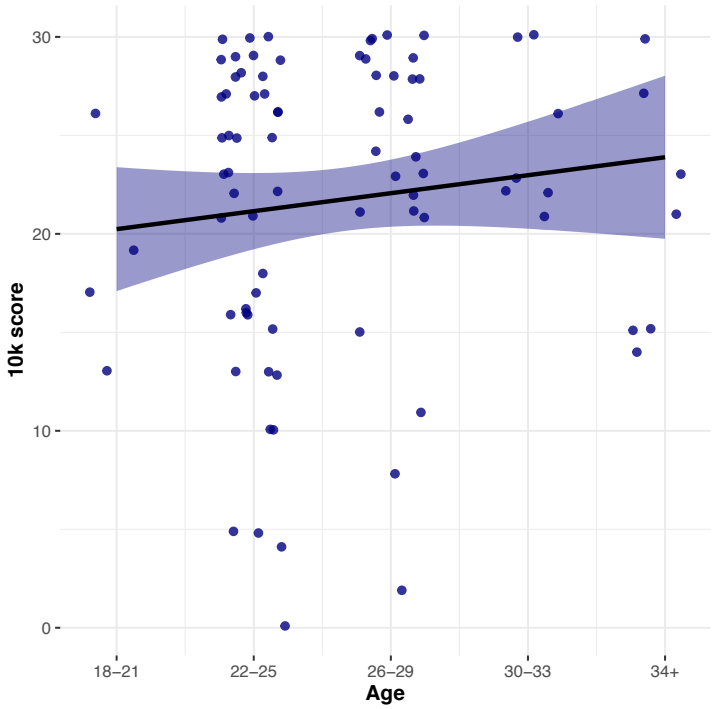


Figure 45. Participant age vs. 10k score

Table 45 (above) and Table 46 (below) show the relationships between age and test results on the 10k level test and the VLT when the total scores of all sections are summed. At

the 10k level, Spearman's ρ is 0.153 suggesting a small positive correlation and Pearson's r was determined to be 0.121 with a confidence interval from -0.094 to 0.326 meaning it is not statically robust as the slope of the linear relationship may be negative. For the VLT scores when correlated with age, Spearman's ρ is 0.104 which indicated a small correlation and Pearson's r was determined to be 0.098 which does not qualify as a significant relationship. Even more so since the confidence interval is from -0.117 to 0.305 and crosses zero.

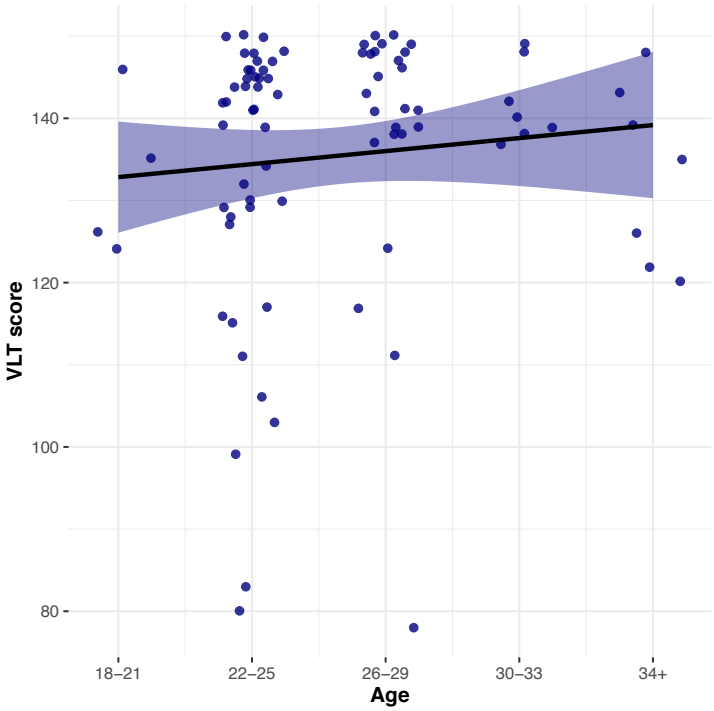


Figure 46. Participant age vs. VLT score

Of all the additional variables that were investigated in relation to the participants' vocabulary results only a few showed statistical symptoms of a significant correlation and potential further investigation. These included the reported proficiency levels of the students as well as their ages. Items that were determined to not have a statically significant correlation include study abroad experience, the number of languages spoken on a regular basis, the presence to English at work, and gender. Implications of these results will be discussed further in Section 4.7.

4.6. AWL vs. FoRTAV

RQ4: How do the results from the AWL section of the VLT compare with the results from the FoRTAV?

To address the last research question, the results from the AWL and FoRTAV will be statistically compared to each other with the help of Spearman's ρ and Pearson's r.

Additionally, a closer look will be taken at the frequency at which items were incorrectly answered for both tests. To see if there was a correlation between the scores on the AWL section of the VLT as well as the FoRTAV correlation coefficients were calculated. One would expect students to perform in a similar fashion on both tests or at least for the scores to be correlated, since both tests are made from items on Coxhead's Academic Word List (AWL). Additionally, since the participants are all students currently enrolled in higher education, high scores and high proficiencies should be expected as well. The AWL tests all parts of speech for the word families, meanwhile the FoRTAV focuses on academic verbs. The resulting chart from their comparison can be found in Figure 47.

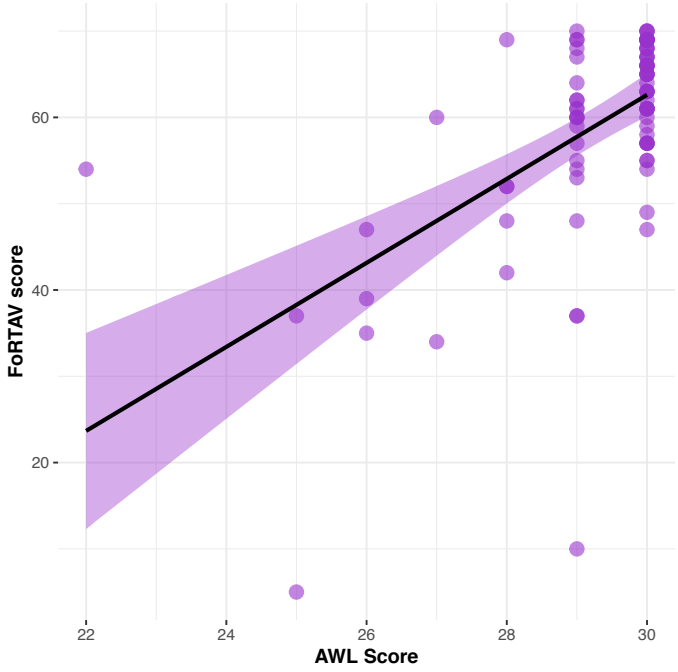


Figure 47. AWL score vs. FoRTAV score scatter plot

Spearman's ρ was determined to be 0.51, meaning there is a strong positive relationship between the results of the two tests. Furthermore, Pearson's r was determined to be 0.566, again indicating a strong positive correlation between the two tests. The confidence interval ranges from 0.402 to 0.695. This range does not include 0 meaning that the resulting correlation coefficients are statistically robust and provide even more evidence that the relationship exists and is a strong relationship.

In addition to a statistical analysis, the individual tested words from both the AWL subsection of the VLT as well as the tested items from the FoRTAV have been charted in Figure 48 on the next page. The items tested on the AWL subsection are indicated by a teal dot and the items tested on the FoRTAV are indicated by an orange triangle. Their position on the x-axis is related to how frequently they were correctly answered by all the participants. The closer the item is to the left side of the chart means it was frequently incorrectly answered by

participants. If the word is positioned on the far-right side of the chart everyone, or nearly everyone was able to answer the item correctly. The position on the y-axis is not very relevant to the interpretation of the chart as vertical spacing was added to make the chart more readable; however, the general upward trend of the slope shows that most test items were correctly answered. Only a few items had a low correct response rate. The most frequently incorrectly answered item from the FoRTAV was *allocate*, and the most frequently incorrectly answered item from the AWL subsection was *correspond*.

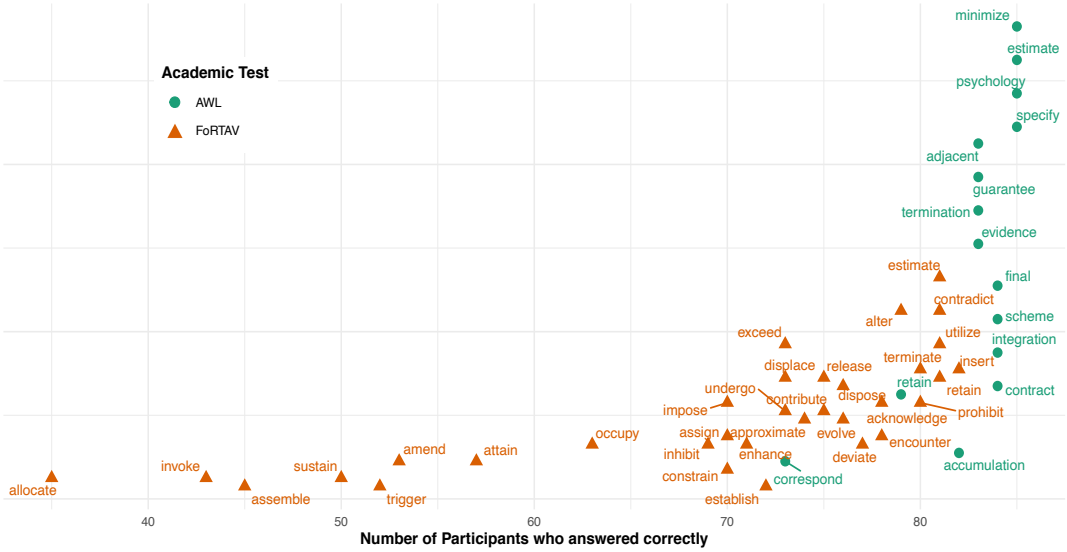


Figure 48. Tested items from the AWL and FoRTAV tests plotted to represent a scale from frequently incorrectly answered to infrequently incorrectly answered

It is also interesting to point out that participants performed extremely well in general on the AWL, hence many of the items are clustered on the right side of the chart. The spread of the frequency rates of FoRTAV items is much larger with some items being found on the far-left side of the chart, such as *invoke* and *assemble*, and others on the far right, such as *retain* and *utilize*. One concerning finding is that an item that was tested on both the AWL and FoRTAV does not appear to have the same number of participants who answered it correctly. This means that more participants correctly answered on the AWL subsection but immediately afterward when taking the FoRTAV, they were not able to repeat their results. This can be seen with the results of *estimate*, as it was tested on both the AWL subsection and the FoRTAV. From the results, approximately 3 more participants were able to answer it correctly on the AWL than could on the FoRTAV. A short discussion and potential reasons for this difference will be suggested in the next section.

A complete version of this chart showing all tested items from every section of the vocabulary tests and the number of participants who answered them correctly can be found in Appendix IV.

4.7. Analysis and Interpretation

Ultimately in the analysis and discussion of the results, a few key features of the data collected have impacted how the results can be discussed and the claims that can be made. It is important to preface the discussion with a short summary of these data features and limitations, so they are apparent. The first feature of the data collected is the fact that the test scores were non-parametric. This means that many types of statistical claims cannot be made, and certain statistical tests and statistical explorations could not be conducted as one prerequisite for their use is a normal distribution. The non-normal distributions of the results on all levels of the vocabulary tests can be seen when comparing the mean scores with the median scores on each section of the test, as there are considerable gaps between the two values showing a skewed data set. Additionally, the number of participants was only 85, a number that is not sufficient for large sweeping claims. Furthermore, of the 85 participants, the data was heavily skewed towards students enrolled at the University of Vienna, who identified themselves as female, and whose study program belonged to Holland's Investigative grouping. Lastly, there was a ceiling effect for many levels of the Vocabulary Levels Test, meaning many of the participants performed very well on the test.

The following subchapter will feature discussions surrounding the chosen passing rate, a short comparison between the AWL results and the FoRTAV results, the Biglan and Holland sorting schemes, the potential subject specific bias of the AWL tests, variable meaning, the participants abilities to self-rank their English level, as well as a comparison to Busby's (2021) results will be featured.

Passing Rate –

When determining what is to be considered passing in the current study, two figures needed to be considered. First and foremost, what is the target vocabulary knowledge for students undertaking academic studies in English, and secondly, what score on the VLT and FoRTAV corresponds with this vocabulary knowledge.

To address the first figure, the closest estimate that seemed appropriate for academic vocabulary knowledge was highlighted in Section 2.4.2. It is a quote from Hazenberg and Hulstun (1996: 158) that reads, "individuals with a vocabulary of fewer than 10,000 base words run a serious risk of not attaining the reading comprehension level required for entering university studies". Furthermore, as the quote claims, 10,000 base words is a minimum for

reaching reading comprehension. This claim matches best with the receptive type tests that the VLT and FoRTAV are, and for these reasons this figure was accepted for the present study as an acceptable cutoff. Higher vocabulary estimates were proposed as well by other researchers but due to the lack of support for these higher numbers and the inability of the VLT to measure levels above 10k, they were not considered for the present study. A lower value was not considered since Hazenberg and Hulstun (1996: 158) further emphasize that vocabulary coverage is potentially even more important when confronting academic texts when compared to everyday life, as the content of the texts is more challenging than the content of texts that one may come across in typical everyday life scenarios.

The second figure considered when setting the passing rate for the levels of the VLT and the FoRTAV was the number of items that had to be answered correctly to be seen as level mastery. The score of 26/30 was taken for the VLT levels since it was the level that Busby (2021) took for her participants in a similar study. To compare the results, it was important to remain consistent in what was deemed mastery. Furthermore, there is extreme variation in the proposed mastery levels by the test creators with scores ranging from 24/30 to 30/30 for each level. With this passing rate determined, it was applied to the FoRTAV as well since there is no proposed passing rate from the test creator.

As it was already hinted at in the results section, the ceiling effect is potentially partially due to the nature of the VLT where the level that is required for a "passing" score or mastery of a section is so close to a perfect score. This means that most of the results when aiming to measure mastery in a group that the test is not too hard for, will be centered around the higher range of scores and this will inevitably lead to a ceiling effect for many data sets. This means that if this figure is correct, to be considered to have adequate receptive vocabulary knowledge to take on university level reading in English, a participant would need to pass every single section of the VLT, including the 10k level. In the current study, only a minority of participants passed the 10k level, which would suggest that most of the participants do not have the vocabulary knowledge deemed necessary by some to even begin university studies in English, nevertheless at the master's level.

When 26/30 or more is needed on each section of the VLT to be considered mastery, 95.3% of participants passed the 2k level, 90.6% of participants passed the 3k level, 76.5% of participants passed the 5k level, 41.2% of participants passed the 10k level, and 96.5% of participants passed the AWL level. Essentially, participants could only answer 20 items incorrectly out of a possible 150 items and still be considered passing. When the same passing level was applied to the FoRTAV results, only 49.41% of the participants passed it.

Some additional considerations to take before reaching any conclusions is that the passing levels were not and are not fully tested, especially on the FoRTAV. The participants represent a broad group of students who encounter English at varying levels in their current master's programs. While the results relating to the differing EME involvement factors did not show any statistically significant trends between the amount of English exposure in the master's programs and the results on any test sections, these were all self-reported values from the students, meaning they may not be completely accurate. From the data alone, it appears that students' performance on the vocabulary tests was not significantly impacted at all by the amount of English present in their current master's programs. However, the students were asked questions concerning what percentage of reading, writing, speaking, and listening in their classes occurred in English, but there was no consideration for total reading, writing, listening, or reading present in their program of study. Two students may have reported that all their speaking occurs in English, but for one they only attend lectures and therefore do not have to speak a lot in class yet for another student they have to speak all the time since they could be in small group classes based around discussions. These differences were not accounted for in the variable development and may have impacted the EME results and ultimately the passing rates.

A potential workaround for this issue would have involved explicitly splitting the students into two groups, EME students and non-EME students. Undertaking this difficult task was only briefly considered, but due to the diverse nature of EMEMUS, it proved nearly impossible to determine how much English would have to be present in a program for it to be considered EME or non-EME as many of the programs where the students were enrolled offer a sort of hybrid or mixed option.

AWL vs. FoRTAV Results -

When the passing rate on the AWL section of the VLT is compared with the passing rates from the FoRTAV, 96.5% of participants passed the AWL section meanwhile only 49.41% of participants passed the FoRTAV. As was already hinted at, this is mostly likely further evidence that the passing rate for the VLT sections is not the correct passing rate for the FoRTAV.

It is interesting to note as well that the academic discipline passing rates were impacted by the FoRTAV results. When the passing rates (not the scores) between all the sections of the VLT are individually compared with the passing rates with the FoRTAV, the order of the disciplines always changes. These are different results from those obtained when looking at the scores. When only the scores were considered, there were not any statistical differences

between the groupings in Biglan or Holland's schemes. The enterprising grouping consistently performed the best or tied with another grouping on nearly all the sections of the VLT. However, when looking at the passing rate on the FoRTAV, the enterprising grouping seemingly struggled considerably when compared in relative terms to the other groupings and the enterprising group's previous performance on the VLT sections. The jump in passing rate and the fact the enterprising group went from the highest passing rate to the lowest, is noteworthy. Furthermore, in Biglan's grouping, the applied grouping performs at a nearly identical level regarding the passing rates as the soft grouping for the 2k, 3k, 5k, and AWL levels of the VLT. However, on the contrary at the 10k level of the VLT and the FoRTAV, the performance of the applied grouping drops off in comparison to the soft grouping. An exact explanation for this phenomenon is not known, but again it is worth noting.

The difference in passing rates between the AWL and FoRTAV for all the participants could also be because of the test formats. The AWL section tests the participant's ability to match 3 definitions when given 5-6 words. Meanwhile, the FoRTAV flips the test and focuses on the ability to match 3 words when given 5-6 definitions. The VLT has never been tested in this other type of format, so it is impossible to predict if this could be a more difficult task and could have an impact on the scores.

Another possible difference between the VLT and FoRTAV that could be contributing to the difference in passing rates has to do with the wording of the definitions. This is especially apparent when the results on the AWL section of the verb *estimate* are compared with the results of the verb *estimate* on the FoRTAV. On the AWL section, students should have matched the item with the definition "to guess the number or size of something". On the FoRTAV, students should have matched the definition of "to roughly judge size, amount, or value" to the word *estimate*. Both tests test the verb *to estimate* so word class cannot explain the difference in the number of participants who answered one correctly versus those who did not. It appears that 4 more students knew that the verb *to guess* is a synonym for *to estimate* than knew *to judge* can also be a synonym for *to estimate*. In this case, the FoRTAV tested participants' knowledge of the verbs *to judge* and *to estimate* meanwhile the AWL section tested the knowledge of *to estimate* and *to guess* which was more well-known to the participants of this study. While this is not a huge difference, it should be considered and could have potential implications for other items on the FoRTAV. There are no further direct comparisons that are possible as that was the only item of overlap between the two tests. A common concern with test items on multiple matching vocabulary tests is that the test is

focused on more than just the target word, and this phenomenon could potentially be at play here, thus impacting the final scores.

Biglan and Holland Academic Disciplines –

When simply looking at the numbers, the claim that differences exist between different academic groups (when separated by Biglan's and Holland's schemes) in terms of vocabulary knowledge was not substantiated. In the present study, none of the resulting p-values from statistical testing were low enough to claim any statistically significant difference between the academic discipline groupings. It is important to state that not all the groups had sufficient sample sizes, especially when considering Holland's groupings; however, since the p-values were so high, it is unlikely that larger groupings would result in a different outcome.

When considering the passing rates of the student groupings in comparison to each other on the different levels of the VLT along with the FoRTAV, certain groups consistently performed better than others. Similar patterns were also observed in each of the grouping schemes. In Biglan's scheme, applied and soft fields consistently did the best in most of the categories. In Holland's groupings, enterprising and realistic fields generally did the best as well. Given Durrant's (2014: 535) claim that "philosophy students used specialized adjectives, engineers used specialized nouns, and science students use specialized verbs" it could be expected that the science members of the hard and pure disciplines would perform the best on the FoRTAV. That is the exact opposite of what happened in the present study, as the applied and soft disciplines had the highest passing rates on the FoRTAV, and this seems to reflect their knowledge of verbs since the test only includes verbs and all four disciplines had similar passing rates on the AWL section of the VLT.

Another interesting observation comes from the most frequently incorrectly answered word lists for each academic grouping. Generally speaking, the soft grouping and the enterprising group had seemingly unique lists in comparison with the other groups. Both groupings also performed relatively poorly on the FoRTAV, so the list of frequently incorrectly answered items may reflect their performance on the FoRTAV. In this case, more of the verbs found on the FoRTAV are found at the tops of their lists, meanwhile the other difficult items from the 10k and 5k lists are lower on the ranking for the soft and enterprising groups.

Lau and Gardner (2019: 265) claimed that the "soft-pure field typically demands much higher linguistic repertoires from its learners in order to negotiate meaning and convey academic thoughts". Lau and Gardner separate soft-pure into one field with the others being

soft-applied, hard-pure, and hard-applied. Meanwhile in the present study, soft was a grouping, pure was a separate grouping, applied was another, and hard was the last grouping. As already mentioned, the soft and applied groups did the best consistently, thus not fully supporting the claim by Lau and Gardner. Perhaps the unique grouping strategies explain some of the differences, but the full idea that soft-pure fields would do the best is not supported.

Ultimately, the method of separating students into their academic discipline groupings based on Holland's and Biglan's schemes did not lead to the discovery of many substantial differences. There are several potential reasons why this happened. First, it could be that vocabulary knowledge is not at all impacted by academic discipline in a way that is measurable by vocabulary testing. As stated in the literature review, Dafouz and Smit (2019: 52) highlight the fact that students need to "acculturate in a second or foreign language" while they also "develop disciplinary discourse conventions". One possibility is that the students may be developing non-vocabulary-related conventions, or perhaps the vocabulary-specific conventions have not yet been developed to the point where significant differences appear on the vocabulary tests. Furthermore, as some of the students in the present study were able to fully pass the vocabulary tests, the argument could be made that more results could have been found if a harder or more detailed vocabulary test was used, or a test of a more involved vocabulary construct was used such as collocation testing. By using a harder test, deeper knowledge of the words could be examined, since many of the participants may be familiar with a lot of the tested word families, but only a more difficult or more detailed test would be able to determine how well the words are known. Furthermore, a text-based genre analysis may have been more revealing to determine which words are not only known by students in academic disciplines but instead focusing on which word families are used by students in different academic disciplines.

Another potential cause for the lack of results relating to the Biglan or Holland groupings could be that perhaps the groupings from Biglan and Holland do not hold up when used in the present-day Austrian university system. Academic discipline is not the only potential factor that could impact students' vocabulary knowledge, and when the Holland and Biglan schemes are used outside of their traditional settings (US-based universities) the patterns may not emerge with the same significance. Knowledge construction has been claimed to be "specific to particular cultures, settings, periods and the natural languages used to construct it" (Trowler in Smit & Dafouz 2020: 51), so any of these additional factors may have an impact that is more significant than academic discipline. Only one study was found to use the Biglan sorting method outside of anglophone universities and it dealt with administrator behaviors

instead of student behaviors so there was not significant evidence in the first place to support the idea that differences between these groups of students would emerge.

Lastly, another critical consideration that may have led to the vocabulary test results not significantly differing between the academic discipline groupings may relate to the interdisciplinary nature of many of the programs that the students are enrolled in. One of the hardest tasks in the present study was sorting the students into their academic disciplines using the Biglan and Holland sorting methods. Many of the EME programs that are offered are relatively new and cover topics that are not covered in Biglan and Holland's initial sorting criteria. This led to many of the programs being sorted into multiple categories, thus larger and more distinct groupings may have resulted in a different outcome of the study. However, many of the programs that were represented in the present study dealt with interdisciplinary or multidisciplinary topics such as Environmental System Sciences, Science Technology and Society, and East Asian Economy and Society. As these programs seemed disproportionally overrepresented in the EME program listings when compared to the general program offerings, it may be too artificial to limit further research to subject that fit neatly into the perhaps outdated methods from Biglan and Holland.

For the reasons mentioned above, as well as a potentially limitless number of additional reasons, the present study did not find any trends when the students were groups by academic discipline according to Biglan or Holland.

Potential biases of the VLT –

A common claim against the AWL as an academic word list is that it is seen as being biased towards commerce and law studies. This is due to the fact, as Durrant (2014: 331) explains, "commerce and law account for 50% of the content of the corpus, whereas diverse disciplines grouped under the headings of Arts and Science share the remaining 50%". By this logic, it should be expected that the academic disciplines including commerce and law would perform the best as the students would be more familiar with a higher percentage of the words present in the vocabulary tests with items taken from the AWL (the AWL section of the VLT and the FoRTAV). Law and commerce would both be defined as belonging to Biglan's soft and applied categories and Holland's enterprising category.

As was already mentioned, when statistically looking at the groups, none of the differences were significant. When looking at the passing rates between the groups, it is true that the enterprising group from Holland's classification performed the best on all sections

except for the FoRTAV. However, this does not support the bias idea as the enterprising students did not outperform the others only on the AWL and FoRTAV sections, but instead outperformed the other groups on the 2k, 3k, 5k, 10k, and AWL sections. It would be surprising if the entire VLT was biased towards one academic discipline grouping as the items from the general levels are not taken from academic texts but the GSL. It is also true that when looking at Biglan's classification scheme results, the soft category consistently performed the best in all the tests and the applied category performed the best except for the 10k and FoRTAV level. If the AWL was truly biased, the results on the AWL section of the VLT and the results on the FoRTAV for these groups would probably also be significantly higher than the other groupings. The present study then provides some supporting evidence for the claim that the AWL and the vocabulary tests based on the AWL are not biased towards a select few academic disciplines.

EME exposure vs. Semesters studied -

When comparing the participant's results to their questionnaire answers, only a few items were found to be correlated. Two variables were developed to use as methods to measure the amount of English medium education exposure the students experienced in their master's programs. The first was a simple ranking with the students scoring from 0 (none) to 5 (5 out of 5 classes) how often they experienced teaching in English, writing in English, reading in English, and speaking or presenting in English. These scores were taken as individual variables in addition to being averaged for a combined EME exposure score. The second method to measure English exposure followed Busby's (2021: 69) reasoning that the more semesters a student had completed at university, inevitably the more English they would have been exposed to. Thus, the students were asked to report how far along in their studies they were in terms of ECTS.

Using the first method, the EME exposure scores were found to have a small and positive non-linear relationship to multiple tests; however, none of the results pertaining to the EME scores were found to be statistically significant. The writing scores and FoRTAV scores, the presenting scores and FoRTAV scores, the reading scores and the combined VLT & FoRTAV scores, the reading scores and the FoRTAV scores, the combined EME exposure score and the combined VLT & FoRTAV scores, and finally the combined EME scores and the FoRTAV scores, all showed this small positive relationship that were not statistically robust.

Using the second method, the semester progress was found to have a small and positive non-linear relationship with the VLT scores. The semester progress was also found to have a small and non-linear relationship with the 10k level scores from the VLT when taken on their

own. This is the same results as the relationships found between participant age and results, again age was found to have a small and positive non-linear relationship with the VLT scores and the 10k scores when taken on their own. As the relationships are different between the two tested methods of measuring EME exposure, it is likely evidence for the claim that the semester progress is a different way to measure age and is not really a reflection of the EME exposure.

Student ability to estimate English knowledge

The relationships that were statistically determined to be the strongest of the variables that were tested, were the relationships between the test results and the self-reported English levels of the students. The self-reported English levels were found to have a medium and a more linear than non-linear relationship with the combined VLT & FoRTAV scores, the FoRTAV scores, the VLT scores, and the 10k level scores. This seems to reflect the student's abilities to self-report their level of English as it is to be expected that higher English levels perform better on the tests. This validates their self-reporting abilities and shows that they are typically aware of their level in English when given a few sorting criteria. This shows that the participants were self-aware of their abilities from the beginning and when given a few can do statements, they were able to accurately separate themselves into leveled groups. It is not a perfect demonstration of their ability to separate themselves as there is no direct score correspondence between FoRTAV results, VLT results, and their CEFR level, but in general students who performed better on the vocabulary tests ranked themselves higher on the CEFR scale.

Busby (2021) comparison –

When comparing the results of the present study to the results from Busby's 2021 study, it is interesting to look at the passing rates from the different academic disciplines. In her study, Busby looked at student performance on the VLT. She had participants from the English department, IT department, and Sociology department in a Norwegian university. She chose these three departments as she knew students from the English department would have the highest amount of English exposure, then the IT students with a medium amount, and finally the Sociology students with minimal to no English exposure in their degree programs. In her study, the English students performed the best and were most likely to pass each level of the VLT. After that, the IT students did the next best and finally, the Sociology students did the worst consistently across the 5 levels of the VLT.

In the present study, no English students were included in the results, as it was known from pilot testing and personal experience from belonging to the department that their results would skew the final scores. IT students were classified as belonging to Biglan's hard and applied categories and Holland's realistic and investigative groupings. There were not any sociology students in the present study, but they would have belonged to Biglan's soft and pure categories and Holland's social groupings. The results from the present study suggest different findings from Busby's but it is important to note that the students in her study were from the bachelor's level and the students in the present study were from the master's level. The difference in the results could also be attributed to the fact that perhaps sociology students in Norway receive much less instruction in English than sociology students in Austria, as suggested by Busby (2021: 63) in her discussion of her choice of participants. This could inevitably change the results, even with the results from the present study suggesting that EME exposure only has a small impact on vocabulary scores.

5. Conclusion

Ultimately, the four guiding research questions related to the present study revealed interesting insights into the vocabulary knowledge of students enrolled in master's programs in Austria. After conducting a small research project with 85 qualifying participants current enrolled in a master's program at a university or FH in Austria, it was shown that while many of the participants have sufficient vocabulary knowledge, there is a large group of students who do not seem adequately prepared for the vocabulary demands of advanced studies in English, if they were interested in pursuing these types of study programs. However, not all the participants were studying only in English and represented a large range of English involvement in their current programs of study.

The participants were asked to complete two vocabulary tests, the VLT and the FoRTAV. The VLT is split into 5 levels including the 2k, 3k, 5k, 10k, and AWL section out of 30 points each. The FoRTAV is one larger test out of 72 total points. The median scores were as follows: 2k section = 30/30, 3k section = 29/30, 5k section = 29/30, 10k section = 23/30, AWL section = 30/30, and FoRTAV = 61/72. When a passing rate of 86% was taken for the VLT test section, 95.3% of participants passed the 2k section, 90.6% of participants passed the 3k section, 76.5% of participants passed the 5k section, 41.2% of participants passed the 10k section, and 96.5% of participants passed the AWL section.

In addition to the tests, a questionnaire was included to create several factors that were examined for any correlation to the vocabulary test results. These factors included: the students

self-ranked English level; their semester progress in their current study program; study abroad experience; an individual English Medium Exposure score relating to reading, writing, listening, and speaking/presenting; a composite EME score; commonly used languages at university, at home, and in their workplace; the number of languages used on a frequent basis; and the Biglan and Holland grouping based on their program of study.

From these factors, only a few were found to have any correlation to the vocabulary test results. The groupings proposed by Biglan and Holland did not explain any of the differences between the performance of the students but led to an investigating into the passing rates of the various subjects as well as the order of words that were most frequently incorrectly answered. The role of EME in vocabulary knowledge was also examined and it seemed to only have a small correlation in a few specific cases, but the results did not hold any statistical significance. The students were reliably able to self-report their English level as it was found to have a medium strength positive correlation to the total VLT scores and the combined VLT and FoRTAV scores. Participant age was found to have a medium strength positive correlation to the 10k section results from the VLT. Lastly, the AWL section of the VLT was compared to the newer FoRTAV test and some of the differences between the results and implications of the two tests were discussed since a strong positive correlation was found between the results.

With the growing popularity of EMEMUS, it is interesting to see what the future holds for HEIs and students as they embark on their chosen paths. Further investigations into what implications this has for student learning and preparedness to tackle studies in English at this high level with the specific needs to academic vocabulary require further study.

Word Count: 38,054 words

6. References

Airey, John. 2011. "The disciplinary literacy discussion matrix: a heuristic tool for initiating collaboration in higher education". *Across the Disciplines* 8(3), 1–9.

Alise, Mark A. 2008. *Disciplinary differences in preferred research methods: a comparison of groups in the Biglan classification scheme*. Louisiana State University and Agricultural & Mechanical College.

Anderson, Richard C.; Freebody, Peter. 1981. "Vocabulary knowledge". In Guthrie, John (ed). *Comprehension and teaching: Research reviews*, Newark, DE: International Reading Association, 77–117.

Ballantyne, Roy; Bain, John D.; Packer, Jan. 1999. "Researching university teaching in Australia: themes and issues in academics' reflections". *Studies in Higher Education*. Taylor & Francis 24(2), 237–257.

Barnes, Laura LB; Agago, Menna O.; Coombs, William T. 1998. "Effects of job-related stress on faculty intention to leave academia". *Research in Higher Education* 39, 457–469.

Barnes, Laura LB; Bull, Kay S.; Campbell, N. Jo; Perry, Katye M. 2001. "Effects of academic discipline and teaching goals in predicting grading beliefs among undergraduate teaching faculty". *Research in Higher Education* 42, 455–467.

Bauer, Laurie; Nation, Paul. 1993. "Word families". *International Journal of Lexicography* 6(4), 253–279.

Baumann, Andreas. 2022a. *Effect sizes* [PowerPoint presentation]. Moodle. https://moodle.univie.ac.at/pluginfile.php/26090661/mod_resource/content/0/Slides%20session%206.pdf.

Baumann, Andreas. 2022b. *Normal distribution* [PowerPoint presentation]. Moodle. https://moodle.univie.ac.at/pluginfile.php/26027025/mod_resource/content/0/Slides%20session%204.pdf.

Becher, Tony. 1994. "The significance of disciplinary differences". *Studies in Higher Education* 19(2), 151–161.

Becher, Tony; Trowler, Paul. 1991. *Academic tribes and territories: Intellectual enquiry and the cultures of disciplines*. 77(3), 60.

Biglan, Anthony. 1973a. "The characteristics of subject matter in different academic areas". *Journal of Applied Psychology* 57(3), 195–203.

Biglan, Anthony. 1973b. "Relationships between subject matter characteristics and the structure and output of university departments". *Journal of Applied Psychology* 57(3), 204–213.

Bracewell, Robert J.; Frederiksen, Carl H. Frederiksen; Donin Frederiksen, Janet. 1982. "Cognitive processes in composing and comprehending discourse". *Educational Psychologist* 17(3), 146–164.

Braxton, John M.; Bayer, Alan E. 1999. *Faculty misconduct in collegiate teaching*. Baltimore, MD: Johns Hopkins University Press.

Braxton, John M.; Hargens, Lowell L. 1996. "Variation among academic disciplines: Analytical frameworks and research". *Higher Education-New York-Agathon Press Incorporated* 11, 1–46.

Braxton, John M.; Olsen, Deborah; Simmons, Ada. 1998. "Affinity disciplines and the use of principles of good practice for undergraduate education". *Research in Higher Education* 39, 299–318.

Browne, Charles; Culligan, Brent; Phillips, Joseph. 2013. *The new academic word list*.

Bryan, Jennifer. 2023. *gapminder: Data from Gapminder*. https://CRAN.R-project.org/package=gapminder.

Busby, Nicole Louise. 2021. "Words from where?: predictors of L2 English vocabulary among Norwegian university students". *ITL – International Journal of Applied Linguistics* 172(1), 58–84.

Campion, Mary E.; Elley, Warwick B. 1971. *The academic vocabulary list*. Wellington: New Zealand Council for Educational Research.
Chen, Qi; Ge, Guang-chun. 2007. "A corpus-based lexical study on frequency and distribution of Coxhead's AWL word families in medical research articles (RAs)". *English for Specific Purposes*. Elsevier 26(4), 502–514.
Chung, Teresa Mihwa; Nation, Paul. 2004. "Identifying technical vocabulary". *System* 32(2), 251–263.
Colbeck, Carol L. 1998. "Merging in a seamless blend: how faculty integrate teaching and research". *The Journal of Higher Education*. Taylor & Francis 69(6), 647–671.
Corson, David. 1997. "The learning and use of academic English words". *Language Learning* 47(4), 671–718.
Costa, Francesca; Coleman, James A. 2013. "A survey of English-medium instruction in Italian higher education". *International Journal of Bilingual Education and Bilingualism* 16(1), 3–19.
Coxhead, Averil. 2000. "A new academic word list". *TESOL Quarterly* 34(2), 213–238.
Coxhead, Averil. 2011. "The academic word list 10 years on: Research and teaching implications". *TESOL Quarterly* 45(2), 355–362.
Coxhead, Averil. 2017. *Vocabulary and English for specific purposes research: quantitative and qualitative perspectives*. Abingdon: Routledge.
Coxhead, Averil; Dang, Thi Ngoc Yen. 2019. "Vocabulary in university tutorials and laboratories". In Hyland, Ken; Wong, Lillian L.C. (eds). *Specialised English: new directions in ESP and EAP research and practice*. Abingdon: Routledge, 120–134.
Crossley, Scott; Salsburg, Tom; McNamara, Danielle. 2010. "The development of polysemy and frequency use in English second language speakers: polysemy and frequency use in English L2 speakers". *Language Learning* 60(3), 573–605.
Dafouz, Emma; Smit, Ute. 2016. "Towards a dynamic conceptual framework for English-medium education in multilingual university settings". *Applied Linguistics* 37(3), 397–415.
Dafouz, Emma; Smit, Ute. 2020. *Road-mapping English medium education in the internationalised university*. Cham: Palgrave Macmillan.
Dafouz, Emma; Smit, Ute. 2023. *Researching English-medium higher education: diverse applications and critical evaluation of the ROADMAPPING framework*. New York: Routledge.
Daller, Helmut; Milton, James Milton; Treffers-Daller, Jeanine (eds.). 2007. *Modelling and assessing vocabulary knowledge* (The Cambridge Applied Linguistics Series). Cambridge: Cambridge University Press.
Daller, Michael; Phelan, David. 2013. "Predicting international student study success". *Applied Linguistics Review* 4(1), 173–193.
Dang, Thi Ngoc Yen. 2018a. "A hard science spoken word list". *ITL - International Journal of Applied Linguistics* 169(1), 44–71.
Dang, Thi Ngoc Yen. 2018b. "The nature of vocabulary in academic speech of hard and soft-sciences". *English for Specific Purposes* 51, 69–83.
Dang, Thi Ngoc Yen; Coxhead, Averil; Webb, Stuart. 2017. "The academic spoken word list". *Language Learning* 67(4), 959–997.
Dang, Thi Ngoc Yen; Webb, Stuart. 2014. "The lexical profile of academic spoken English". *English for Specific Purposes* 33, 66–76.
Dannerer, Monika; Gaisch, Martina; Smit, Ute. 2021. "Englishization 'under the Radar'." In Wilkinson, Robert; Gabriels, René (eds.). *The englishization of higher education in Europe*. Amsterdam: Amsterdam University Press, 281-306.
Doberneck, Diane M.; Schweitzer, John H. 2017. "Disciplinary variations in publicly engaged scholarship: An analysis using the Biglan classification of academic disciplines". *Journal of Higher Education Outreach and Engagement* 21, 78–103.

Durrant, Philip. 2014. “Discipline and level specificity in university students’ written vocabulary”. *Applied Linguistics* 35(3), 328–356.
Farrell, Paul. 1990. “Vocabulary in ESP: a lexical analysis of the English of electronics and a study of semi-technical vocabulary”. CLCS Occasional Paper No. 25.
Favero, Marietta Del. 2006. “An examination of the relationship between academic discipline and cognitive complexity in academic deans’ administrative behavior”. *Research in Higher Education* 47, 281–315.
Fox, Craig. 2014. “Wittgenstein on family resemblance”. In *Wittgenstein: key concepts*. London: Routledge, 51-62.
Galloway, Nicola. 2017. *Global Englishes and change in English language teaching: attitudes and impact*. London: Routledge.
Gardner, Dee; Davies, Mark. 2014. “A new academic vocabulary list”. *Applied Linguistics* 35(3), 305-327.
Geisler, C. 1994. *Academic literacy and the nature of expertise: reading, writing, and knowing in academic philosophy*. New York: Routledge.
Ghadessy, Mohsen. 1979. “Frequency counts, word lists and materials preparation: a new approach”. *English Teaching Forum* 17(1), 24–27.
Ghamarian-Krenn, Katharina. 2023. “Extramural English and academic verb knowledge: a longitudinal study of Viennese students majoring in English”. PhD Thesis, University of Vienna.
Gottfredson, Gary D.; Holland, John L. 1996. *Dictionary of Holland occupational codes*. Psychological Assessment Resources Incorporated.
Gottfredson, G.D.; Holland, John L. 1991. *Position classification inventory (PCI): professional manual*. Psychological Assessment Resources.
Goulden, Robin; Nation, Paul; Read, John. 1990. “How large can a receptive vocabulary be”. *Applied Linguistics* 11(4), 341–363.
Grabe, William. 1988. “Reassessing the term ‘interactive’.” In Carrell, Patricia L.; Devine, Joanne; Eskey, David E. (eds). *Interactive approaches to second language reading*. Cambridge: Cambridge University Press, 56–70.
Grabe, William. 1991. “Current developments in second language reading research”. *TESOL Quarterly*. Wiley Online Library 25(3), 375–406.
Hazenberg, Suzanne; Hulstun, Jan H. 1996. “Defining a minimal receptive second-language vocabulary for non-native university students: an empirical investigation”. *Applied Linguistics* 17(2), 145–163.
Hearn, James C.; Anderson, Melissa S. 2002. “Conflict in academic departments: an analysis of disputes over faculty promotion and tenure”. *Research in Higher Education* 43, 503–529.
Henriksen, Birgit; Albrechtsen, Dorte; Haastrup, Kirsten. 2004. “The relationship between vocabulary size and reading comprehension in the L2”. In Henriksen, Birgit; Albrechtsen, Dorte; Haastrup, Kirsten. *Writing and vocabulary in foreign language acquisition*, Copenhagen: Museum Tusculanum, 129–140.
Hiller, Stephen. 2021. “Disciplinary differences in faculty emphasis on deep approaches to learning: comparing conceptualizations of academic discipline”. Paper presented at the *American Educational Research Association Annual Meeting* (AERA 21), Virtual, Apr. 2021.
Holland, John. L.; Whitney, D. R.; Cole, N. S.; Richards J. M. 1969. “An empirical occupational classification derived from a theory of personality and intended for practice and research”. ACT Research Report No. 29. *Iowa City: American College Testing Program*.
Holland, John L. 1997. *Making vocational choices: a theory of vocational personalities and work environments*. Psychological Assessment Resources.

Holland, John L.; Gottfredson, Gary D. 1976. "Using a typology of persons and environments to explain careers: some extensions and clarifications". *The Counseling Psychologist* 6(3), 20–29.
Holland, John Lewis; Holmberg, Kay; Rosen, Donald. 1989. *SDS-the self directed search: the college majors finder*. Psychological Assessment Resources, Incorporated.
Hsu, Wenhua. 2011. "The vocabulary thresholds of business textbooks and business research articles for EFL learners". *English for Specific Purposes* 30(4), 247–257.
Hsu, Wenhua. 2014. "Measuring the vocabulary load of engineering textbooks for EFL undergraduates". *English for Specific Purposes* 33, 54–65.
Hsueh-Chao, Marcella Hu; Nation, Paul. 2000. *Unknown vocabulary density and reading comprehension*. University of Hawaii National Foreign Language Resource Center.
Hutchings, Catherine; Garraway, James (eds). 2010. "Beyond the university gates: provision of extended curriculum programmes in South Africa". *Proceedings of the January 2009 Rhodes University Foundation Seminar*.
Hvitfeldt, Emil. 2021. *paletteer: Comprehensive Collection of Color Palettes*. https://github.com/EmilHvitfeldt/paletteer.
Hyland, Ken. 2002. "Authority and invisibility: authorial identity in academic writing". *Journal of pragmatics*, 34(8). 1091–1112.
Hyland, Ken; Test, Polly. 2007. "Is there an 'academic vocabulary'?" *TESOL Quarterly* 41(2), 235–253.
Jones, Willis A. 2012. "Variation among academic disciplines: an update on analytical frameworks and research". *The Journal of the Professorate* 6(1), 9-27.
Kamimoto, Tadamitsu. 2008. "Nation's vocabulary levels test and its successors: a re-appraisal". PhD thesis, Swansea University.
Kassambara, Alboukadel. 2023. *ggpubr: "ggplot2" Based Publication Ready Plots*. https://CRAN.R-project.org/package=ggpubr.
Kekäle, Jouni. 1999. "'Preferred' patterns of academic leadership in different disciplinary (sub) cultures". *Higher education*, 37(3) 217–238.
Kidwell, Linda A.; Kidwell, Roland E. 2008. "Do the numbers add up to different views? Perceptions of ethical faculty behavior among faculty in quantitative versus qualitative disciplines". *Journal of Business Ethics*, 78 141–151.
Kolb, David A. 1981. "Learning styles and disciplinary differences". *The Modern American College* 1, 232–235.
Konstantakis, Nikolaos. 2007. "Creating a business word list for teaching business English". *ELIA* 7, 79-102.
Kremmel, Benjamin; Schmitt, Norbert. 2016. "Interpreting vocabulary test scores: what do various item formats tell ss about learners' ability to employ words?". *Language Assessment Quarterly* 13(4), 377–392.
Kremmel, Benjamin; Schmitt, Norbert. 2017. "Vocabulary levels test". In Liontas, John I.; DelliCarpini, Margo (eds). *The TESOL encyclopedia of English language teaching*. Hoboken, NJ: John Wiley & Sons, 1–7.
Kuteeva, Maria; Airey, John. 2014. "Disciplinary differences in the use of English in higher education: reflections on recent language policy developments". *Higher Education* 67(5), 533–549.
Lam, Jacqueline Kam-mei. 2001. "A study of semi-technical vocabulary in computer science texts". PhD thesis, University of Exeter.
Lau, Ken; Gardner, David. 2019. "Disciplinary variations in learning styles and preferences: implications for the provision of academic English". *System* 80, 257–268.
Laufer, Batia. 1992. "How much lexis is necessary for reading comprehension". In Arnaud, P. J. L.; Bejoing, H. (eds). *Vocabulary and applied linguistics*. London: Macmillan, 126–132.

Laufer, Batia; Nation, Paul. 1999. “A vocabulary-size test of controlled productive ability”. *Language Testing* 16(1), 33–51.
Laufer, Batia; Ravenhorst-Kalovski, Geke C. 2010. “Lexical threshold revisited: lexical text coverage, learners’ vocabulary size and reading comprehension”. *Reading in a Foreign Language* 22(1), 15–30.
Li, Yongyan; Qian, David D. 2010. “Profiling the academic word list (AWL) in a financial corpus”. *System* 38(3), 402–411.
Lynn, Robert W. 1973. “Preparing word-lists: a suggested method”. *RELC Journal* 4(1), 25–28.
Macalister, John; Nation, I.S.P. (eds.). 2011. *Case studies in language curriculum design: concepts and approaches in action around the world* (ESL & Applied Linguistics Professional Series). New York: Routledge.
Martínez, Iliana A.; Beck, Silvia C.; Panza, Carolina B. 2009. “Academic vocabulary in agriculture research articles: a corpus-based study”. *English for Specific Purposes* 28(3), 183–198.
Maxwell, Lesli A. 2013. “Common core ratchets up language demands for English learners”. *Education Week* 33(10), 14–16.
Milem, Jeffrey F.; Umbach, Paul D. 2003. “The influence of precollege factors on students’ predispositions regarding diversity activities in college”. *Journal of College Student Development* 44(5), 611–624.
Milton, James; Fitzpatrick, Tess (eds). 2013. *Dimensions of vocabulary knowledge*. London: Palgrave Macmillan.
Ming-Tzu, Karen Wang; Nation, Paul. 2004. “Word meaning in academic English: homography in the academic word list”. *Applied Linguistics* 25(3), 291–314.
Moreno, Roxana Ed; Park, Babette. 2010. *Cognitive load theory: historical development and relation to other theories*. Cambridge: Cambridge University Press.
Morstain, Barry R.; Smart, John C. 1976. “Educational orientations of faculty: assessing a personality model of the academic professions”. *Psychological Reports* 39(3), 1199–1211.
Mudraya, Olga. 2006. “Engineering English: a lexical frequency instructional model”. *English for Specific Purposes* 25(2), 235–256.
Nagy, William E.; Anderson, Richard C. 1984. “How many words are there in printed school English”. *Reading Research Quarterly* 19(3), 304–330.
Nation, I.S.P. 2001. *Learning vocabulary in another language: the goals of vocabulary learning*. Cambridge: Cambridge University Press.
Nation, I.S.P. 2006. “How large a vocabulary is needed for reading and listening”. *Canadian Modern Language Review-revue Canadienne Des Langues Vivantes* 63(1), 59–82.
Nation, I.S.P. 2008. *Teaching vocabulary: strategies and techniques*. Boston, MA: Heinle.
Nation, I.S.P. 2011. “Research into practice: vocabulary”. *Language Teaching* 44(4), 529–539.
Nation, I.S.P. 2016. *Making and using word lists for language learning and testing*. Philadelipha, PA: John Benjamins Publishing Company.
Nation, I.S.P. 2018. *Learning vocabulary in another language*. (2nd edition). Cambridge: Cambridge University Press.
Nation, I.S.P.; Coxhead, Averil. 2013. “Special purposes vocabulary”. In Chapelle, Carol A. (ed.). *The encyclopedia of applied linguistics*. Chichester: Wiley-Blackwell.
Nation, I.S.P.; Webb, Stuart. 2010. *Researching and analyzing vocabulary*. Boston, MA: Heinle.
Nelson Laird, T. F.; Schwarz, Michael J.; Kuh, G. D. Kuh; Shoup, R. 2006. “Disciplinary differences in faculty members’ emphasis on deep approaches to learning”. Paper presented at the *Association for Institutional Research Annual Forum* (Air Forum 46), Chicago, 17 May 2006.

Nelson Laird, Thomas F.; Shoup, Rick; Kuh, George D.; Schwarz, Michael J. 2008. "The effects of discipline on deep approaches to student learning and college outcomes". *Research in Higher Education* 49(6), 469–494.
Nesi, Hilary. 2002. "An English spoken academic word list". In *Proceedings of the tenth EURALEX International Congress*, Copenhagen, Denmark, 351–358.
Neuwirth, Erich. 2022. *RColorBrewer: ColorBrewer Palettes*. https://CRAN.R-project.org/package=RColorBrewer.
Nizonkiza, Déogratias. 2016. "First-year university students' receptive and productive use of academic vocabulary". *Stellenbosch Papers in Linguistics* 45(1), 169–187.
Norton, Lin; Richardson, T.E.; Hartley, James; Newstead, Stephen; Mayes, Jenny. 2005. "Teachers' beliefs and intentions concerning teaching in higher education". *Higher Education* 50, 537–571.
Palfreyman, David M.; Walt, Christa Van Der. 2017. "Introduction: biliteracies in higher education". In Palfreyman, David M.; Walt, Christa Van Der (eds.). *Academic Biliteracies*. Bristol: Multilingualism Matters, 1–18.
Patil, Indrajeet. 2021. Visualizations with statistical details: the "ggstatsplot" approach. *Journal of Open Source Software*. The Open Journal 6(61). 3167. https://doi.org/10.21105/joss.03167.
Pecorari, Diane; Shaw, Philip; Malström, Hans. 2019. "Developing a new academic vocabulary test". *Journal of English for Academic Purposes* 39, 59–71.
Peters, Dianne S. 1974. "The link is equitability". *Research in Higher Education* 2(1), 57–64.
Peters, Pam; Fernández, Trinidad. 2013. "The lexical needs of ESP students in a professional field". *English for Specific Purposes* 32(4), 236–247.
Pike, Gary R.; Killian, Timothy S. 2001. "Reported gains in student learning: do academic disciplines make a difference?". *Research in Higher Education* 42(4), 429–454.
Praninskas, Jean. 1972. *American university word list*. London: Longman.
Qian, David D. 1999. "Assessing the roles of depth and breadth of vocabulary knowledge in reading comprehension". *Canadian Modern Language Review-revue Canadienne Des Langues Vivantes* 56(2), 282–307.
Qian, David D. 2002. "Investigating the relationship between vocabulary knowledge and academic reading performance: an assessment perspective". *Language Learning* 52(3), 513–536.
R Core Team. 2020. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.
Read, John. 2000. *Assessing vocabulary*. 1st edn. Cambridge: Cambridge University Press.
Richards, Jack C.; Pun, Jack. 2021. "A typology of English-medium instruction." *RELC Journal* 54(1), 216-240.
Rose, Heath; Macaro, Ernesto; Sahan, Kari; Aizawa, Ikuya; Zhou, Sihan; Wei, Minhui. 2021. "Defining English medium instruction: striving for comparative equivalence". *Language Teaching* 56(4), 1–12.
Rosen, Donald; Holmberg, Kay; Holland, John L. 1994. *The educational opportunities finder: for use with the self-directed search and the vocational preference inventory*. Psychological Assessment Resources.
Santos, M. G. 2010. "Depth of academic vocabulary knowledge: investigating depth of academic vocabulary knowledge among language-minority community college students". *Reflections on English Language Teaching* 9(1), 19–42.
Schmitt, Norbert. 2010. *Researching vocabulary: a vocabulary research manual*. New York: Palgrave Macmillan.
Schmitt, Norbert; Jiang, Xiangying; Grabe, William. 2011. "The percentage of words known in a text and reading comprehension". *The Modern Language Journal* 95(1), 26–43.
Schmitt, Norbert; Schmitt, Diane. 2020. *Vocabulary in language teaching*. 2nd edn. Cambridge: Cambridge University Press.

Schmitt, Norbert; Schmitt, Diane; Clapham, Caroline. 2001. "Developing and exploring the behaviour of two new versions of the vocabulary levels test". *Language Testing* 18(1), 55–88.
Scott, Peter. 2011. "The university as a global institution". In King, Roger; Marginson, Simon; Naidoo, Rajani (eds.). *Handbook on globalization and higher education*. Edward Elgar Publishing, 59-75.
Simpson, Adrian. 2017. "The surprising persistence of Biglan's classification scheme". *Studies in Higher Education* 42(8), 1520–1531.
Slowikowski, Kamil. 2023. *ggrepel: Automatically Position Non-Overlapping Text Labels with "ggplot2."* https://CRAN.R-project.org/package=ggrepel.
Smart, John C. 2006. "Holland's theory and patterns of college student success" in the *Commissioned report for the national symposium on postsecondary student success: spearheading a dialog on student success*.
Smart, John C.; Ethington, Corinna A.; Umbach, Paul D.; Rocconi, Louis M. 2009. "Faculty emphases on alternative course-specific learning outcomes in Holland's model environments: the role of environmental consistency". *Research in Higher Education* 50, 483–501.
Smart, John C.; Feldman, Kenneth A.; Ethington, Corinna A. 2000. *Academic disciplines: Holland's theory and the study of college students and faculty*. Vanderbilt University Press.
Smart, John C; Umbach, Paul D. 2007. "Faculty and academic environments: using Holland's theory to explore differences in how faculty structure undergraduate courses". *Journal of College Student Development* 48(2), 183–195.
Stæhr, Lars Stenius. 2008. "Vocabulary size and the skills of listening, reading and writing". *Language Learning Journal* 36(2), 139–152.
Stewart, Jeffrey; White, David A. 2011. "Estimating guessing effects on the vocabulary levels test for differing degrees of word knowledge". *TESOL Quarterly* 45(2), 370–380.
Strevens, Peter. 1973. "Technical, technological, and scientific English (TTSE)". *ELT Journal* 27(3), 223–234.
Tatzl, Dietmar. 2011. "English-medium masters' programmes at an austrian university of applied sciences: attitudes, experiences and challenges". *Journal of English for Academic Purposes* 10(4), 252–270.
Thompson, Michael. 2003. "Disparate academic environments: an emergent framework of socialization". *The Qualitative Report* 8(3), 408-434.
Todd, Richard Watson. 2017. "An opaque engineering word list: which words should a teacher focus on?". *English for Specific Purposes* 45, 31–39.
Townsend, Dianna; Collins, Penny. 2009. "Academic vocabulary and middle school English learners: An intervention study". *Reading and Writing* 22, 993–1019.
Trowler, Paul. 2014a. "Depicting and researching disciplines: strong and moderate essentialist approaches". *Studies in Higher Education* 39(10), 1720–1731.
Trowler, Paul. 2014b. "Academic tribes and territories: the theoretical trajectory". *Austrian Journal of Historical Studies* 25(3), 17-26.
Umbach, Paul D. 2006. "The contribution of faculty of color to undergraduate education". *Research in higher education* 47(3), 317–345.
Vongpumivitch, Viphavee; Huang, Ju-yu; Chang, Yu-Chia. 2009. "Frequency analysis of the words in the academic word list (AWL) and non-AWL content words in applied linguistics research papers". *English for Specific Purposes* 28(1), 33–41.
Wang, Jing; Liang, Shao-lan; Ge, Guang-chun. 2008. "Establishment of a medical academic word list". *English for Specific Purposes* 27(4), 442–458.
Ward, Jeremy. 2009. "A basic engineering English word list for less proficient foundation engineering undergraduates". *English for Specific Purposes* 28(3), 170–182.

Webb, Stuart. 2008. “The effects of context on incidental vocabulary learning”. *Reading in a Foreign Language* 20(2), 232–245.
Webb, Stuart; Sasao, Yosuke; Ballance, Oliver. 2017. “The updated vocabulary levels test: developing and validating two new forms of the VLT”. *ITL-International Journal of Applied Linguistics* 168(1), 33–69.
Wickham, Hadley. 2007. “Reshaping data with the reshape package”. *Journal of Statistical Software* 21(12), 1–20.
Wickham, Hadley; Averick, Mara; Bryan, Jennifer; Chang, Winston; D’Agostino McGowan, Lucy; François, Romain; Grolemund, Garrett, et al. 2019. Welcome to the tidyverse. *Journal of Open Source Software* 4(43). 1686. https://doi.org/10.21105/joss.01686.
Wickham, Hadley; François, Romain; Henry, Lionel; Müller, Kirill; Vaughan; Davis. 2023. *dplyr: A Grammar of Data Manipulation*. https://CRAN.R-project.org/package=dplyr.
Wickham, Hadley; Hester, Jim; Bryan, Jennifer. 2023. *readr: Read Rectangular Text Data*. https://CRAN.R-project.org/package=readr.
Wickham, Hadley; Vaughan, Davis; Girlich, Maximilian. 2023. *tidyr: Tidy Messy Data*. https://CRAN.R-project.org/package=tidyr.
Xing, Peiling; Fulcher, Glenn. 2007. “Reliability assessment for two versions of vocabulary levels tests”. *System* 35(2), 182–191.
Xue, Guoyi; Nation, Ian SP. 1984. “A university word list”. *Language Learning and Communication* 3(2), 215–229.
Ylijoki, Oili-Helena. 2000. “Disciplinary cultures and the moral order of studying: a case-study of four Finnish university departments”. *Higher Education* 39, 339–362.
Zechmeister, Eugene B.; Chronis, Andrea M.; Cull, William L.; D’Anna, Catherine A.; Healy, Noreen A. 1995. “Growth of a functionally important lexicon”. *Journal of Literacy Research* 27(2), 201–212.
Zeeland, Hilde van; Schmitt, Norbert. 2013. “Lexical coverage in L1 and L2 listening comprehension: the same or different from reading comprehension?”. *Applied Linguistics* 34(4), 457–479.

7. Appendix I – Questionnaire

Dear Participants,

Thank you for agreeing to participate in my study about English vocabulary knowledge of students in the higher education system. My name is Jennie Atwell and I currently am a masters student at the University of Vienna in the English department. My hope is to help you understand your vocabulary abilities in English as well as compile a database that will help universities, administrators, professors, and other students understand the vocabulary needs of students in Austria. With this knowledge, they will be able to make more informed decisions relevant to teaching in English as well as set individual learning goals for students in the Austrian high education system.

The following survey is split into a few sections. First some background information will be asked for about you, your studies, and your previous experiences studying English. Then the vocabulary test will begin. This includes 6 separate sections and they will get harder as you progress through them. Since this is a passive vocabulary test, please answer the questions quickly and do not spend much time thinking about the answers. If you do not know the answer you can skip the item or click "I don't know".

If you have any questions regarding the study or concerns, you can always reach me at:
a11922041@unet.univie.ac.at

Thank you again,
Jennie Atwell
* Indicates required question

Data Processing

If you would like me to score and process your results please fill in your email below. This will give you a rough idea of the number of word families you know in English as well as areas you can focus your attention.

1. Did your teacher/professor give you a code? If so, write it below. (ex. WU01, TU08, UW11) If you have multiple codes write them all here. _____

2. Your email address will not be distributed or used for any other purpose than* besides returning vocabulary scores if you so desire and to prevent participants from taking the test twice. For all data processing and publishing purposes, the data will be completely anonymized. If you select "no" you cannot participate in this study.

   To confirm you understand and accept this, please check "yes".
   Mark only one oval.
   - ◯ yes
   - ◯ no

3. If you want your results emailed to you, enter an email address below where the results can be sent. If no, ignore this question. _____

Personal Information

Information about you for statistical purposes only

4. Where are you from? (Nationality or birthplace) *
Mark only one oval.
- ○ Austria
- ○ Germany
- ○ Italy
- ○ Hungay
- ○ Switzerland
- ○ Other:

5. How old are you? *
Mark only one oval.
- ○ 18-21
- ○ 22-25
- ○ 26-29
- ○ 30-33
- ○ 34+

6. What is your gender? *
Mark only one oval.
- ○ Female
- ○ Male
- ○ Non-binary
- ○ Prefer not to say
- ○ Other:

Language Skills

Languages you use regularly

7. What is/are your first language(s)? *
Check all that apply.
- ☐ German
- ☐ English
- ☐ Italian
- ☐ Hungarian
- ☐ Turkish
- ☐ Bosnian/Croatian/Serbian
- ☐ Other:

8. What language(s) do you use at **work**? (select all that apply) *

Check all that apply.

- [ ] German
- [ ] English
- [ ] Italian
- [ ] Hungarian
- [ ] Turkish
- [ ] Bosnian/Croatian/Serbian
- [ ] Other:

9. What language(s) do you use **with your family**? (select all that apply) *

Check all that apply.

- [ ] German
- [ ] English
- [ ] Italian
- [ ] Hungarian
- [ ] Turkish
- [ ] Bosnian/Croatian/Serbian
- [ ] Other:

10. What language(s) do you use **with your friends**? (select all that apply) *

Check all that apply.

- [ ] German
- [ ] English
- [ ] Italian
- [ ] Hungarian
- [ ] Turkish
- [ ] Bosnian/Croatian/Serbian
- [ ] Other:

11. How would you rate your English level? *

Mark only one oval.

- ( ) A1 Beginner - You can interact simply as long as the other person speaks slowly and clearly
- ( ) A2 Elementary - You can understand and commincate routine information
- ( ) B1 Intermediate - You can communicate in situations and use simple language to communcate feelings, opinions, plans and experiences.
- ( ) B2 Upper Intermediate - You can communicate easily and understand and express some complex ideas and topics
- ( ) C1 Advanced - You can use English flexibly and effectively without much effort for social and academic purposes.
- ( ) C2 Proficiency - You can communicate very fluently and precisely in complex situations as well as summarize information effectivel
- ( ) English is my first language or I am a native English speaker

12. Have you studied abroad or lived abroad before? *
Mark only one oval.
- ▢ yes
- ▢ no

13. If yes, where and for how long?

_____

University Information

About your current study program

14. Which university do you attend? *
Mark only one
- ○ Akademie der bildendenden Künste Wien
- ○ Anton Bruckner Privatuniversität
- ○ Bertha von Suttner Privatuniversität
- ○ CAMPUS 02 Fachhochschule der Wirtschaft GmbH
- ○ Central European University Private University
- ○ Charlotte Fresenius Privatuniversität
- ○ Danube Private University
- ○ Fachhochschule Burgenland GmbH
- ○ Fachhochschule Campus Wien
- ○ Fachhochschule des bfi Wien GmbH
- ○ Fachhochschule für angewandte Militärwissenschaften
- ○ Fachhochschule Joanneum GmbH
- ○ Fachhochschule Kärnten
- ○ Fachhochschule Kufstein Tirol Bildungs GmbH
- ○ Fachhochschule Salzburg GmbH
- ○ Fachhochschule Sankt Pölten GmbH
- ○ Fachhochschule Technikum Wien
- ○ Fachhochschule Vorarlberg GmbH
- ○ Fachhochschule Wiener Neustadt GmbH
- ○ Ferdinand Porsche Fernfachhochschule GmbH
- ○ FH Gesundheitsberufe OÖ GmbH
- ○ FH OÖ Studienbetriebs GmbH
- ○ FHG – Zentrum für Gesundheitsberufe Tirol GmbH
- ○ FHW Fachhochschul-Studiengänge Betriebs- und Forschungseinrichtungen der Wiener Wirtschaft GmbH (FHW GmbH)
- ○ Gustav Mahler Privatuniversität für Musik
- ○ IMC Fachhochschule Krems GmbH

- JAM MUSIC LAB Private University for Jazz and Popular Music Vienna
- Johannes Kepler Universität Linz
- Karl Landsteiner Privatuniversität für Gesundheitswissenschaften
- Katholische Privatuniversität Linz
- Lauder Business School
- MCI Management Center Innsbruck – Internationale Hochschule GmbH
- Medizinische Universität Graz
- Medizinische Universität Innsbruck
- Medizinische Universität Wien
- MODUL University Vienna Privatuniversität
- Montanuniversität Leoben
- Musik und Kunst Privatuniversität der Stadt Wien
- New Design University Privatuniversität St.Pölten
- Paracelsus Medizinische Privatuniversität
- Privatuniversität Schloss Seeburg
- Sigmund Freud Privatuniversität
- Stella Vorarlberg Privathochschule für Musik
- Technische Universität Graz
- Technische Universität Wien
- UMIT TIROL – Privatuniversität für Gesundheitswissenschaften und technologie
- Universität für angewandte Kunst Wien
- Universität für Bodenkultur Wien
- Universität für künstlerische und industrielle Gestaltung Linz
- Universität für Musik und darstellende Kunst Graz
- Universität für Musik und darstellende Kunst Wien
- Universität für Weiterbildung Krems
- Universität Graz
- Universität Innsbruck
- Universität Klagenfurt
- Universität Mozarteum Salzburg
- Universität Salzburg
- Universität Wien
- Veterinärmedizinische Universität Wien
- Webster Vienna Private University
- Wirtschaftsuniversität Wien
- Other

15. In your current program, what languages do you use with **classmates in class**? *

Check all that apply.

- [ ] German
- [ ] English
- [ ] Italian
- [ ] Hungarian
- [ ] Turkish
- [ ] Bosnian/Croatian/Serbian
- [ ] Other:

16. In your current program, what languages do you use with **classmates outside of class**? *

Check all that apply.

- [ ] German
- [ ] English
- [ ] Italian
- [ ] Hungarian
- [ ] Turkish
- [ ] Bosnian/Croatian/Serbian
- [ ] Other:

17. In your current program, what languages do you use with **professors**? *

Check all that apply.

- [ ] German
- [ ] English
- [ ] Italian
- [ ] Hungarian
- [ ] Turkish
- [ ] Bosnian/Croatian/Serbian
- [ ] Other:

18. In your current program, what languages do you use with **administrators**? *

Check all that apply.

- [ ] German
- [ ] English
- [ ] Italian
- [ ] Hungarian
- [ ] Turkish
- [ ] Bosnian/Croatian/Serbian
- [ ] Other:

19. What do you study currently? *

Write the full name of your study program (ex. Supply Chain Management, Steuern und Rechnungslegung, Biotechnology, Digital Humanities, etc.)

_____

20. How many ECTS have you completed in your master's program so far? (an approximation)
Mark only one oval.
- 0-29 (this is my first semester)
- 30-59 (~1 full-time semester completed)
- 60-89 (~2 full-time semesters completed)
- 90-119 (~3 full-time semesters completed)
- 120-159 (~4 full-time semesters completed)
- 160+
- not applicable (see next question)

21. If you study law, medicine, or another non-bachelor/master program: how many ECTS have you completed so far?
Mark only one oval.
- 0-29 (this is my first semester)
- 30-59 (~1 full-time semester completed)
- 60-89 (~2 full-time semesters completed)
- 90-119 (~3 full-time semesters completed)
- 120-159 (~4 full-time semesters completed)
- 160-189 (~5 full-time semesters completed)
- 190-219 (~6 full-time semesters completed)
- 220-259 (~7 full-time semesters completed)
  260-289 (~8 full-time semesters completed)
- 290+

22. So far in your current program, what percentage of your classes have been **taught** in English? (approximately)
Mark only one oval.
- 0% - none
- 1- 20% ~ 1 in 5
- 21- 40% ~ 2 in 5
- 41 - 60% ~ 3 in 5
- 61 - 80% ~ 4 in 5
- 81 - 100% ~ 5 in 5

23. So far in your current program, how much of your **reading** for university has * been in English?
Mark only one oval.
- 0% - none
- 1- 20%
- 21- 40%
- 41 - 60%
- 61 - 80%
- 81 - 100%

24. So far in your current program, how much of your **writing** for university has * been in English?
Mark only one oval.
- o 0% - none
- o 1- 20%
- o 21- 40%
- o 41 - 60%
- o 61 - 80%
- o 81 - 100%

25. So far in your current program, how much of your **speaking or presentations** for university have been in English?
Mark only one oval.
- o 0% - none
- o 1- 20%
- o 21- 40%
- o 41 - 60%
- o 61 - 80%
- o 81 - 100%

26. In your current program, is there an English for specific purposes class *
(Academic writing, English for law, English for medicine, English for business, English for presentations, etc.)
Mark only one oval.
- o Yes, it is mandatory and I have taken it already
- o Yes, it is mandatory but I have not taken it yet
- o Yes, it is optional and I have taken it already
- o Yes, it is optional but I have not taken it yet
- o Yes, but it is optional and I will not take it
- o No, it is not offered
- o Other:

Bachelor's degree

or first 3 years of a non-master/bachelor program

27. What was the name of the program your bachelors degree was in? (Write the full name of your program ex. Supply Chain Management, Steuern und Rechnungslegung, Biotechnology, Digital Humanities, Medicine, etc.)

_____

28. What language was your bachelors in? *
Check all that apply.
- German
- English
- Other:

29. In your bachelors program, what percentage of your classes were **taught** in * English? (approximately)
Mark only one oval.
- 0% - none
- 1- 20% ~ 1 in 5
- 21- 40% ~ 2 in 5
- 41 - 60% ~ 3 in 5
- 61 - 80% ~ 4 in 5
- 81 - 100% ~ 5 in 5

30. In your bachelors program, how much of your **reading** for university was in English?
Mark only one oval.
- 0% - none
- 1- 20%
- 21- 40%
- 41 - 60%
- 61 - 80%
- 81 - 100%

31. In your bachelors program, how much of your **writing** for university was in * English?
Mark only one oval.
- 0% - none
- 1- 20%
- 41 - 60%
- 61 - 80%
- 81 - 100%

32. In your bachelors program, how much of your **speaking or presentations** for * university were in English?
Mark only one oval.

- ○ 0% - none
- ○ 1 - 20%
- ○ 21- 40%
- ○ 41 - 60%
- ○ 61 - 80%
- ○ 81 - 100%

33. Have you studied anything else? *
Mark only one oval.

- ○ yes
- ○ no

34. If yes, write the program name and how many ECTS you completed (approximately)

_____

8. Appendix II – Vocabulary Tests

The FoRTAV test was obtained through personal correspondence with Dr. Katharina Ghamarian-Krenn. The relevant PhD thesis relating to the vocabulary test is cited below.

Ghamarian-Krenn, Katharina. 2023. "Extramural English and academic verb knowledge: a longitudinal study of Viennese students majoring in English". PhD Thesis, University of Vienna.

The version of the VLT that was used in the current master's thesis was found in Appendix 2 of Schmitt, Schmitt, and Clapham's 2001 article about developing and exploring the behavior of two new versions of the VLT. The citation for this article is also found below.

Schmitt, Norbert; Schmitt, Diane; Clapham, Caroline. 2001. "Developing and exploring the behaviour of two new versions of the Vocabulary Levels Test". *Language Testing* 18(1), 55–88.

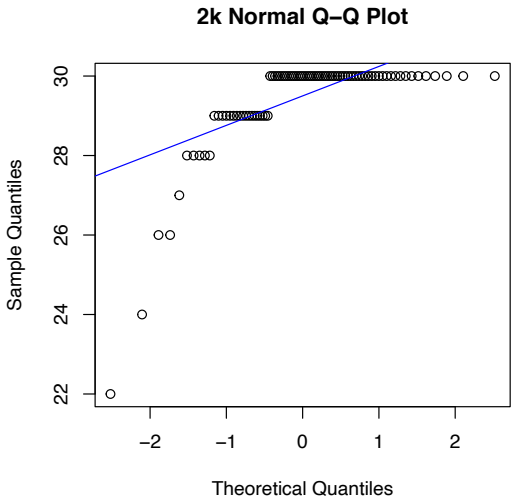
9. Appendix III – Data set and R-code

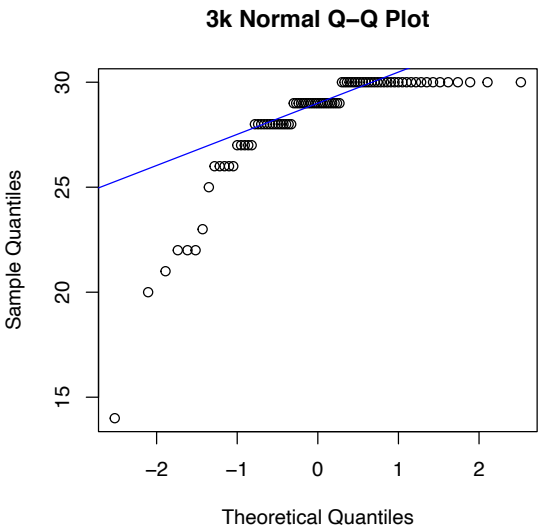
The full data sets that were created and used to create the visualizations and statistical measures can be found in the linked PHAIDRA collection below. In addition to the data sets, in this collection the R-code that was created for this thesis can also be found.

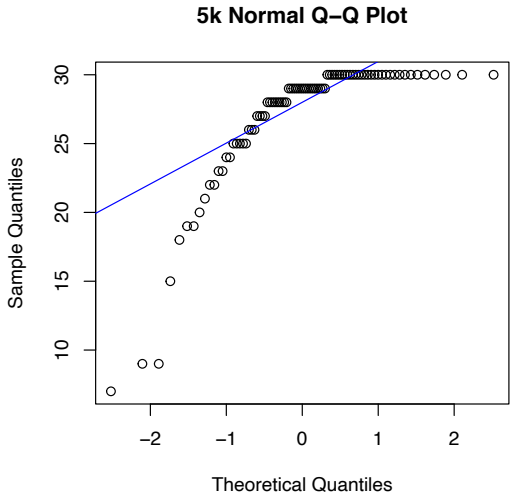
Collection on PHAIDRA - https://phaidra.univie.ac.at/o:1795045

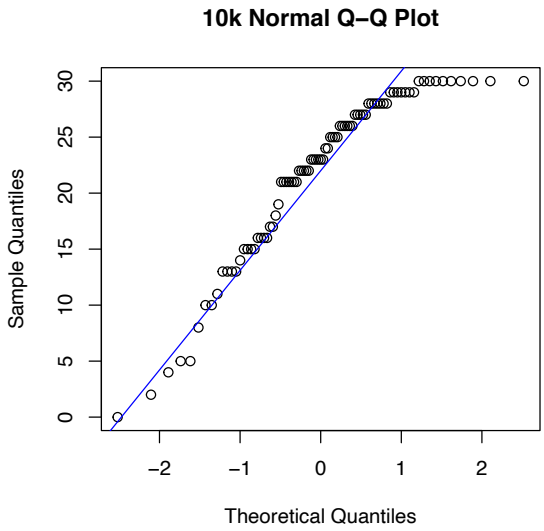
10. Appendix IV – Extra Figures

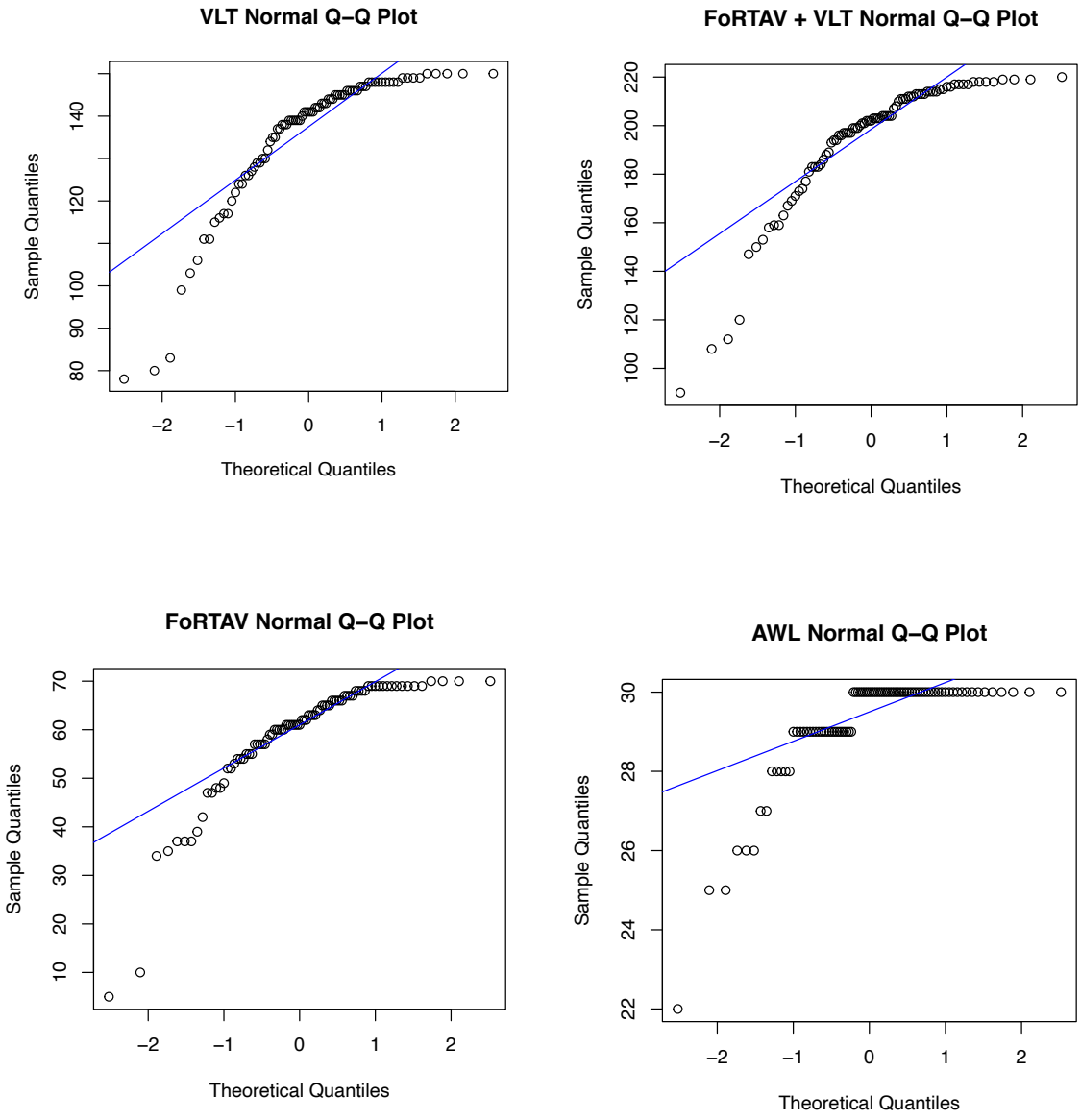
These are all the Q-Q plots that were created to show the non-parametric distributions of the results from the vocabulary test. The residuals of the test scores are the small dots on each plot and the line represents the normal residual line. For all the plots a significant number of points are found off the line therefore the data is general is not normally distributed.



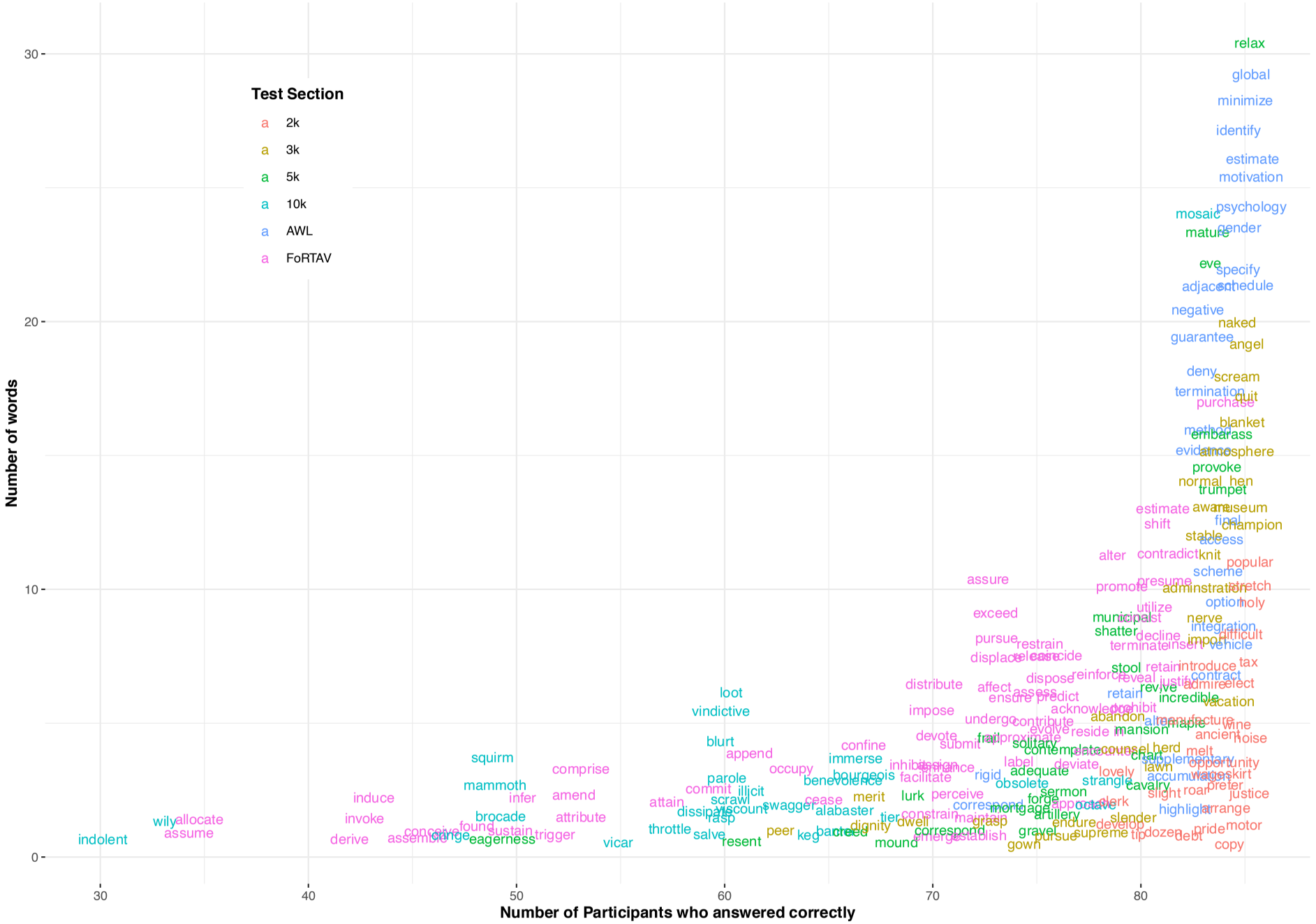








On the next page is the full version of Figure 48. In this chart, all the items that were tested are plotted on the x-axis depending on the frequency at which the item was incorrectly answered. Items incorrectly answered the most are found on the left side of the plot while items correctly answered the most are found on the right side of the plot. The y-axis is a rough estimate of the number of words that were at the same frequency level, therefore only few items were frequently incorrectly answered while many items were frequently correctly answered. This plot is best viewed in color as the label colors correspond to the matching test section.


Test Section
a 2k
a 3k
a 5k
a 10k
a AWL
a FoRTAV
Number of words
Number of Participants who answered correctly
0
10
20
30
30
40
50
60
70
80
relax
global
minimize
identify
estimate
motivation
psychology
mosaic
gender
mature
eve
specify
adjacent
schedule
negative
naked
guarantee
angel
deny
scream
termination
quit
purchase
blanket
embarrass
atmosphere
provoke
normal
hen
trumpet
estimate
awareness
museum
final
champion
shift
stable
access
alter
contradict
knit
popular
scheme
promote
presume
administration
stretch
option
holy
utilize
municipal
nerve
integration
shatter
decline
import
difficult
terminate
insert
vehicle
assure
exceed
pursue
restrain
displace
release
tax
stool
retain
introduce
contract
dispose
reveal
admire
elect
distribute
affect
assess
revive
incredible
vacation
ensure
predict
impose
undergo
contribute
abandon
manufacture
wine
evolve
reside
mansion
maple
ancient
noise
devote
submit
frail
proximate
solitary
contemplate
counsel
herd
melt
loot
vindictive
blurt
append
confine
immerse
occupy
squirm
comprise
parole
mammoth
induce
invoke
brocade
attribute
infer
amend
commit
illicit
scrawl
attain
dissipate
swagger
cease
merit
lurk
perceive
alabaster
throttle
salve
resent
peer
keg
mound
correspond
gravel
pursue
supreme
gown
debt
dozen
tip
endure
develop
slender
artillery
sermon
forge
clerk
strangle
cavalry
lawn
chart
supplementary
opportunity
accumulate
skirt
prefer
roar
slight
highlight
arrange
justice
motor
pride
copy
indolent
wily
allocate
assume
derive
assemble
change
eagerness
conceive
found
sustain
trigger
vicar
bourgeois
benevolence
facilitate
rigid
adequate
obsolete
label
deviate
lovely
inhibition
assign
tenance
inhabitant
correspondent
mortgage
tier
dwell
grasp
maintain
constrain
dignity
bankrupt
creed
merge
establish
lease
eviction
homicide
reinforce
reveal
justify
retain
acknowledge
exhibit
contribute
evolve
reside
frail
proximate
solitary
contemplate
counsel
herd
chart
supplementary
lawn
accumulate
skirt
prefer
roar
slight
highlight
arrange
justice
motor
pride
copy
125

11. Appendix V – German Abstract

Die Zahl der Studierenden, die während ihrer Hochschulausbildung in irgendeiner Form Englisch verwendet, nimmt weltweit zu. Darüber hinaus spielt Englisch im akademischen Bereich eine unbestreitbare Rolle als Hauptsprache für die Veröffentlichung und Vermittlung wissenschaftlicher Erkenntnisse. Eine wichtige Debatte dreht sich derzeit um die potenziellen Auswirkungen, die zusätzlicher Kontakt mit Englisch auf die Sprachkenntnisse von Studierenden haben kann.

In der folgenden Studie wurden die Vokabelergebnisse Masterstudierender in Österreich in zwei englischen Vokabeltests verglichen, um festzustellen, ob bestimmte Faktoren einen statistisch signifikanten Unterschied zwischen Gruppen von Studierenden nahelegen. Bei den verwendeten Vokabeltests handelte es sich einerseits um den etablierten Vocabulary Levels Test (VLT) von Schmitt, Schmitt und Clapham (2001), der zur Messung des allgemeinen Wortschatzwissens diente. Zusätzlich wurde der neu entwickelte Form Recognition Test of Academic Verbs (FoRTAV) von Ghamarian-Krenn (2023) verwendet, der sich auf akademisches Englisch konzentriert. Für die vorliegende Studie wurden mehrere Variablen verwendet, um die Teilnehmer zu gruppieren und die Auswirkungen von zusätzlichem Kontakt mit Englisch zu messen: die akademische disziplinäre Gruppierung der Studierenden nach zwei verschiedenen Sortierschemata; der Umfang von Englisch in den Masterprogrammen; das von den Studierenden selbst eingeschätzte Niveau in Englisch; das Alter; der Semesterfortschritt; die Studienerfahrung im Ausland und die häufig verwendeten Sprachen innerhalb und außerhalb des universitären Umfelds.

Die Variablen wurden mit dem Ziel entwickelt, vier Hauptforschungsfragen zu beantworten. Erstens: Über welche Wortschatzkenntnisse verfügen Masterstudierende an österreichischen Universitäten gemäß den VLT- und FoRTAV-Ergebnissen auf den verschiedenen Frequenzniveaus und in den verschiedenen Fachgebieten? Zweitens: Unterscheidet sich das Vokabularwissen, wenn die Studierenden nach Biglans oder Hollands Schema in akademischen Disziplinen gruppiert werden? Drittens: Korrelieren die Verwendung von Englisch im aktuellen Studiengang oder andere biografische Faktoren mit den Vokabelergebnissen der Studierenden, und wenn ja, auf welche Weise? Und schließlich: Wie verhält sich die Leistung der Studierenden im akademischen Teil des VLT (AWL) im Vergleich zu ihrer Leistung im FoRTAV?

Bei einer Stichprobengröße von 85 Masterstudierenden an österreichischen Universitäten zeigte sich eine statistisch signifikante positive, mittelstarke Korrelation zwischen den von den Studenten angegebenen Englischkenntnissen und den VLT-Ergebnissen sowie den kombinierten VLT- und FoRTAV-Ergebnissen. Das Alter wies eine statistisch signifikante, mittelstarke positive Korrelation mit dem 10k-Niveau auf, und die Ergebnisse des AWL-Teils des VLT und des FoRTAV wiesen eine statistisch signifikante, stark positive Korrelation auf. Die Gruppierungsschemata der akademischen Disziplinen wiesen keine Korrelation mit den Ergebnissen des Vokabeltests auf. Im Durchschnitt schnitten die Teilnehmer gut ab, wobei die Medianwerte in den VLT-Abschnitten bei 2k = 30/30, 3k = 29/30, 5k = 29/30, 10k = 23/30, AWL = 30/30 und FoRTAV = 61/72 lagen. Bei einer Mindestanforderung von 86% richtig beantworteter Fragen bestanden 95,3% der Teilnehmer den 2k-Abschnitt, 90,6% der Teilnehmer den 3k-Abschnitt, 76,5% der Teilnehmer den 5k-Abschnitt, 41,2% der Teilnehmer den 10k-Abschnitt und 96,5% der Teilnehmer den AWL-Abschnitt. Die Ergebnisse der Studierenden im AWL-Teil des VLT korrelierten stark und positiv mit den Ergebnissen des FoRTAV. Die möglichen Auswirkungen dieser Ergebnisse und die Vorbehalte gegenüber diesen Ergebnissen werden ebenfalls erörtert.