Pathway to CLIL

–

A Proposed Sequence of Subjects in CLIL Education Based on Linguistic Requirements of Selected Subjects

Acknowledgments

At the end of a process of such nature, as a dissertation indeed is, it is time to look back and take the opportunity to thank everyone who was part of that process. Just like with raising children, it also takes a village to complete a dissertation.

Table of contents

Table of contents

Table of figures

Table of figures

Index of tables

Glossary of terms

Gymnasium	specific level of German secondary education, which is either bi- or tripartite; Gymnasium is the level aiming at preparing students to study at a university
Abitur	diploma from German secondary school qualifying for university admission or matriculation
Kultusministerkonferenz	Standing conference of all German secretaries of education, each federal state is represented individually
Institut für Qualitätsentwicklung für das Bildungswesen	institute for quality management and assurance for the German educational landscape
Anforderungsbereich	level of complexity along the lines of cognitive demands of tasks

List of abbreviations

BICS	Basic Interpersonal Communication Skills
CALP	Cognitive Academic Language Proficiency
CEFR	Common European Framework of Reference for Languages
CLIL	Content Language Integrated Learning, the bidirectional integration of content and language in teaching
STEM	group of subjects concerned with Sciences, Technology, Engineering, and Mathematics
KoA	*Keiner ohne Abschluss:* a special program in Rhineland-Palatinate to save youths to leave school without a qualifying degree
ECML	European Centre of Modern Languages
EVP	English Vocabulary Profile, an online service provided, i.a. by Cambridge University, assigning CEFR levels to single words, from their basic to abstract, rare meanings
BNC	British National Corpus
AWL	Academic Word Lists, a collection of eight lists, differentiated by academic genre in which the words listed are most likely used

1 Introduction

"Die Grenzen meine Sprache bedeuten die Grenzen meiner Welt."
(Wittgenstein 2016: v. 5.6)

The Austrian philosopher Ludwig Wittgenstein (1889 – 1951) could not have put the importance of language learning into better words than stating that the limits of one's language mean the limits of one's world. Especially in today's world, a globalized and interconnected world, in which speaking a foreign language is indispensable and a decisive factor in the individual's success, education should attempt to nullify such limits.

One approach to provide students with additional opportunities to language learning and speaking is the teaching of content subjects in a foreign language. Those programs not only target language competences but also cognitive skills and intercultural competences. Consequently, learning in a foreign language further expands the limits and has always been "at the core of the European school's ethos" (Leaton Gray, Scott & Mehisto 2018: 50). The *Kultusministerkonferenz* (KMK) has also recognized the benefits of teaching in a foreign language and encourages schools to implement programs focusing and intensifying combined language and content learning (cf. KMK 2013a). In Germany, such programs already existed before the recommendation by the KMK since after World War 2, when the Elysée Treaty decreed that France and Germany do everything in their power to reconcile the animosities between their peoples. Accordingly, subjects like History and Social Sciences were offered in the other country's language in schools along the French-German border (ibid.). The idea was to help to understand the other better and thus, expand the limits of peoples' views of the other. Over the course of time, however, the programs were offered more widely and changed from French to English (ibid.: 4). The subjects remained: especially History is still the most prominent subject to be taught in a foreign language and is well established in the educational landscape (see chapter 3). With the increasing popularity of such programs, further subjects joined the canon. Today, Geography and Biology are also often offered to be taught in a foreign language accounting for the fact that either the subjects contribute to fulfilling the mandate of education (see chapter 2) or a decisive factor in graduates' ability to compete in the international field of the subject (see chapter 3.2). With the growing number of schools implementing what is either called *bilingual teaching* or *Content Language Integrated Learning* (CLIL), publishers started to offer comprehensive textbooks for the well-established subjects in the target language, making teaching easier for teachers, especially since a targeted training to teach a subject in a foreign language is not yet required or standardized (cf. Fein 2021). Accordingly, teachers rely on the textbooks and trust that the language used in those is appropriate, meaning imposes an appropriate linguistic demand on students' foreign language skills within the targeted grade. The question whether this actually is the case, whether the language used is not too demanding for students, has not yet been raised. If, however, textbooks were too demanding, would that demand not set new limits and boundaries

to students' development and self-fulfillment? And would that not contradict the overall target of teaching and learning a subject in a foreign language?

Those questions are the focus of and motivation for this thesis. By identifying each subject's language requirements for receptive and productive skills, expressed in terms of the Common European Framework of Reference and its levels, this thesis sets out to systematically analyze teaching material – in specific textbooks – used in bilingual classes in Rhineland-Palatinate to unveil the linguistic demand the material imposes on students. The target is to identify a sequence of subjects that caters to the development of language skills students undergo throughout their school education. Currently, subjects are introduced into bilingual or CLIL teaching by means of teacher availability or popularity of the subject, adhering to the KMK's list of core subjects of bilingual programs (cf. KMK 2013a: 14; Ministerium für Bildung, Wissenschaft, Weiterbildung und Kultur 2011: sec. 1.1). However, if such a sequence was identified and in place, teachers could tap into the full potential of teaching in a foreign language by not only using the language to teach about a subject but also use the content subject to teach the language. This bidirectional approach is what makes teaching in a foreign language CLIL teaching – integrating content and language (see chapter 3.1). CLIL is considered the silver-bullet of language education (cf. Wolff & Sudoff 2015: 7) and although the KMK uses this term interchangeably with the term bilingual teaching, the latter is currently the most realized approach in Germany (see chapter 3.2).

The aspect of unknowingly teaching with potentially overtaxing material and the factors conditioning the introduction of subjects out of sequence may be caused by the current state of research in the field of bilingual teaching and CLIL, or rather the lack thereof. Although research on CLIL and bilingual teaching on other aspects is vast (see chapter 4.4), research dedicated to systematically analyze required language competences and proficiency of bilingual students is scarce (cf. Piske 2015: 110). Initially, research praised bilingual education for its successes in student achievement, in language competence but also especially with single, subject-related aspects such as *Fremdverstehen* (understanding of the other) in History (cf. Lamsfuß-Schenk 2008). Furthermore, the development of and student performance in selected skills was analyzed, e.g. reading skills (Doll 2002), the impact of bilingual education on motivation, especially in science subjects (cf. Piesche 2016; Rolletschek 2021), and the question whether students perform better because of bilingual education (Rumlich 2016). The aspect of imposed demands connected to bilingual education and discourse was researched by Dalton-Puffer (2007) but with a focus on cognitive demand. While Handro concluded that in History, students are already struggling with formulating their thoughts in the traditional language of schooling (German), bilingual History teaching further increases the problems for students (cf. 2013 : 332). Accordingly, a systematized multi-subject analysis of language requirements so far is a desideratum this thesis intends to address. Other studies concerned with the suitability of selected subjects for bilingual teaching (cf. Mentz 2010) did not look into the aspect of required language skills. A single study currently looked at language requirements in a field comparable to bilingual teaching: the study on language requirements

in Mathematics and History for migrant students conducted by the European Center of Modern Languages (see chapter 4.5).

This study served as a starting point for this thesis. In an attempt to cover multiple subjects and skills, two studies were conducted looking at the linguistic requirements in the subjects History, Geography, Biology, and Mathematics for receptive skills by analyzing the textbook material with the help of an automated tool (see chapters 5.2 and 7), and the linguistic requirements connected to classroom discourse (see chapters 5.2 and 8). The study was designed to be based on textbook material and what the researcher deemed as minimally required regarding student contribution to ensure educational success (see chapter 2.2). This way, the constraints of the global pandemic could be fully respected without delaying or disrupting the research. Consequently, the studies did not incorporate potential language support provided through teachers, the validation of the horizon of expectations with regards to student performance was secured through peer coding. Nevertheless, the results of both studies combined allowed for the proposal of a sequence of subjects pointing to a pathway to CLIL. Thus, this thesis combines findings for four selected subjects and their language requirements for three out of four skills in ten chapters.

To arrive at a necessary understanding of the importance of language learning and the curricular standards especially in connection with the educational success of the individual, the mandate of education and educational success have to be discussed first. Following this introduction, chapter 2 picks up on the opening quote to elaborate on the mandate of education and educational success in Germany, or specifically Rhineland-Palatinate. The chapter concludes elaborating on the *CEFR*, the Common European Framework of Reference for languages, which serves as the grading system and benchmark this thesis' studies also employed.

Chapter 3 elaborates on teaching in a foreign language and juxtaposes bilingual teaching and CLIL. Since this thesis is concerned with the requirements imposed by the subjects as they are taught in Rhineland-Palatinate, this chapter also sheds light on the realization of teaching in a foreign language in said federal state.

The next chapter discusses the role and character of text in teaching in general, but also in the subjects selected for this study. Furthermore, the productive skills speaking and writing are examined for their importance in teaching and their character in the various selected subject for this study. The chapter proceeds with a brief elaboration on how to identify language requirements through up-front assessment with the help of the descriptors provided by the CEFR companion volume (cf. Council of Europe 2018). An overview of the current status of research for bilingual teaching and CLIL completes chapter 4, including the study by the *European Center of Modern Languages* (ECML), in which language requirements for migrant students in History and Mathematics in a multi-step approach were determined, is dissected and critically discussed. Based on the findings and the criticism directed at the ECML study, the underlying idea is transferred onto the context of teaching in a foreign language.

Introduction

The research questions discussed in chapter 5 set the direction in which this thesis' studies are headed and devises the design chosen to answer those questions. Two different approaches to determining linguistic requirements are presented. One approach is targeted at the textbook material and reading of which, while the other approach concentrates on the evaluation of the tasks to be carried out by students in the context of the textbook material. Research validity aspects are furthermore discussed and established.

After the theoretical foundation, the thesis moves on to the empirical work conducted to answer the research questions. Chapter 6 is concerned with the determination of language requirements of the textbook material for the four selected subjects. It illustrates the research design in more detail, explains the processes for material selection, the applied methodologies and key figures. Afterwards, the findings are described and discussed in chapter 7.

Chapter 8 features the same structure to illustrate, discuss, and explain the findings yielded from the task analysis, bar the part on material selection. Both chapters conclude with the comparison of the results obtained through the analyses with the curricular standards for language learning (cf. KMK 2003).

The following chapter outlines the potential sequential approach to introduce the selected and analyzed subjects by summarizing the results from chapters 7 and 8 and deriving the order from those. It further elaborates on the potential for each subject within the sequence to realize CLIL teaching with the help of teaching examples.

The final chapter, the résumé, summarizes all findings and results. Additionally, it critically analyses limitations and elaborates on future opportunities based on the thesis' findings. At this point, the relevance for and contribution to the research concerned with bilingual teaching and learning, and CLIL is reinforced as this thesis helps to fill in this gap in research.

2 Mandate of education and educational success

According to the *Universal Declaration of Human Rights*, agreed upon by the General Assembly of Nations in 1948, education is a human right. Nations are to ensure that "education shall be free, at least in the elementary and fundamental stages." (United Nations General Assembly 1948: art. 26). Furthermore, states are obliged to "promote understanding, tolerance and friendship among all nations, racial or religious groups, and shall further the activities of the United Nations for the maintenance of peace" (ibid.). This declared human right translates into the individual nations' mandates of education and the definition of educational success.

2.1 Mandate of education

In Germany, the state monitors the entire school system (cf. III, §27 (3), LV RLP; §7 (1), GG). The German *Grundgesetz*, the Basic Law, comparable to the idea of a democratic constitution, decrees that the federal states form schools in a way that children and youths are educated to become devout[2] citizens. The focus is on the education of citizens able to be charitable, attentive towards others, abide by the law and stand up for fairness, take over responsibility for themselves but also for nature and the environment. Furthermore, education is mandated to create citizens with a democratic mindset and good work ethics, contributing to the overall aim of reconciliation among the European peoples and international understanding (cf. III, §33, LV RLP). This mandate is based on the first version of the Basic Law of Rhineland-Palatinate of 1947, which is also to be found in the 1949 version of the German *Grundgesetz*. For today's understanding, the wording may sound a little outdated. Nevertheless, the underlying idea of schooling children and youths to become critical citizens capable of upholding a democracy has found its way into the current curricula for all subjects.

Through History teaching, e.g., students should become qualified to critically assess past events in order to develop an identification with a state ensuring the free democratic basic order (cf. Ministerium für Bildung, Wissenschaft, Weiterbildung und Kultur 2021: 2). History in specific is mandated with the education of "responsible and active citizens", to be a "decisive factor in reconciliation, recognition, understanding, and mutual trust between peoples", to promote "fundamental values such as tolerance, […], human rights and democracy", and lastly a mean to preserve "historical and cultural heritage in Europe" (Beacco 2007: 3). Mathematics contribute to the overarching aim of education by providing students solutions and approaches to solve problems in order to help them orient themselves in the present and future world (cf. Ministerium für Bildung, Wissenschaft, Weiterbildung und Kultur 2007: 3). Furthermore, the OECD sees Mathematics as a way to help individuals to "make well-founded judgments and […] to [live] as a constructive, concerned and reflective citizen" (OECD 2004: 24). Learning about natural sciences, e.g., in Biology, is thought to further provide students with knowledge that helps them become citizens with awareness of the importance of environmental questions as a mean to uphold a stable democracy (cf. Ministerium für Bildung, Wissenschaft,

---

[2] The term *devout* may appear outdated but is used on purpose since the Basic Law uses the German equivalent.

Weiterbildung und Kultur 2014a: 5). The goal of science teaching, including Biology, is not only "the mastery of the *basic structure* and of specific *items of knowledge*" (original emphasis) but also the development of a framework in which the sciences can help to understand the world (Vollmer 2010a: 6). Furthermore, science education should clarify the relevance of their discipline by putting topics, problems, and procedures into a socio-critical context (ibid.: 7). Lastly, the languages taught in school contribute to the aim of reconciliation and international understanding formulated in 1947 (cf. Ministerium für Bildung 2021: 2). Especially English is regarded as a window to the world serving as <u>the</u> *lingua franca*, and thus, is ascribed with increased significance (ibid.). Moreover, the development of this international understanding is only possible through the development of language proficiency (cf. Leaton Gray, Scott & Mehisto 2018: 54).

According to Mentz, schools and curricula are also increasingly influenced by economic interests and international studies comparing student competences (cf. Mentz 2010: 40). The European Council declared

> that Europe should become "the most competitive and dynamic knowledge-based economy in the world, capable of sustainable economic growth with more and better jobs and greater social cohesion
>
> (Education Council 2001: 4).

Curricula all over Europe adopt this target and seek to enable students or, much rather, graduates to succeed in their professional lives (cf. Mentz 2010: 40). In 2014, the European Commission enhanced this economic goal to "cultural unity and common values in a globalized world" (European Commission 2014: 6), adopting the post-war German mandate. German schools in Rhineland-Palatinate offer various programs for the heterogeneous student body to support this goal. On the one hand, the project *Keiner Ohne Abschluss*, a project attempting to prevent students from leaving school without graduating, offers weaker students a special program to achieve a certain level of education, allowing them to build a career (cf. KoA – Kompendium: 2013). On the other hand, stronger students are further supported by special educational tracks focusing on either STEM subjects (Sciences, Technology, Engineering, and Mathematics) or so-called *bilingual tracks*, offering various content subjects being taught in a foreign language (cf. KMK 2013a: 5). Such bilingual tracks are mandated, along with the above-described, with the realization of the European language policy seeking to enable Europeans to "have meaningful communicative competence in at least two other languages in addition to [their] mother tongue" (Commission of the European Communities 2003: 4). Language learning is a central aspect to the European Council to achieve the economic goal set and with that, realize "the development of the individual, who can thus realize [their] full potential and live a good life" (Education Council 2001: 4). This is also adopted by the *Standing Conference of the Ministers of Education and Cultural Affairs of the Länder in the Federal Republic of Germany* (KMK) and thus bilingual tracks are regarded as

education towards the potential to partake in the international working environment (cf. KMK 2013a: 6). In summary, the mandate of education to help students become critical citizens able to uphold democracy and develop into individuals realizing their full potential to build a career successfully in the European context is accounted for in the curricula. In their considerations on the role of language in the European schools, Leaton Gray et al. however state that "language classes could do more to help to prepare students" (2018: 61) for the challenges awaiting them in Europe.

2.2 Educational success

With an individual's educational success being a decisive factor in reaching the goals of both the European Union and the educational mandate in Germany, the results of the *Bildungsbericht* 2020 are sobering. In 2018, a total of 54.000 school students left the German educational system without having accomplished a qualification reversing the trend determined since 2006 (cf. DIPF & Leibniz-Institut für Bildungsforschung und Bildungsinformation 2020: 143). This observation also holds true for Rhineland-Palatinate, for which 6.7% of all students were recorded in 2017 to have left primary and secondary educational system without a qualification (cf. Landesamt für Statistik RLP 2018: 51). As outlined by the Education Council (see chapter 2.1), educational failure influences an individual's life just like educational success influences life. Basically, educational success has only a two-fold specification: failure or success (cf. Rüesch 2001: 11). The decisive factor for educational success is student performance, especially in Germany (cf. Deutsche Kinder- und Jugendstiftung 2014: 9). Student performance is the result of learning processes initiated and/ or guided by teaching and teachers (cf. Rüesch 2001: 9). Usually, summative assessments measure performance in schools, i.e., students are tested for knowledge and developed competences towards the end of a learning or thematic unit. Assessment throughout the thematic unit appreciating possible progress is still rarely done (ibid.: 11). The results of such summative assessments are then compared to standards set by various entities and institutions. In Rhineland-Palatinate, assessment of student performance follows a six-tiered grading system with tier 1 – *sehr gut* – certifying outstanding performance above the horizon of expectations. As of tier 5 – *mangelhaft* – certifying inadequate performance, a student or much rather their performance are no longer considered successful (cf. Rheinland-Pfalz 2009: sec. §53 (1 & 2) SchulO RP). On the one hand, either educational or learning standards partly pre-set the criteria and the horizon of expectation, e.g., the educational standards for the first foreign language by the KMK, and on the other hand, defined by each school and their subject-related conferences comprised of the teachers for the distinct subjects (cf. ibid.: sec. § 50 (1); §53 (4) SchulO RP).

Additionally, for language teaching, the *Common European Framework of Reference* (CEFR) serves as a standard to compare and assess the level of student performance. The CEFR "provides a common basis for […] curriculum guidelines, examinations, textbooks, etc. across Europe" (Council of Europe 2010: 1). It provides a detailed description of learning goals, differentiated in levels, to be achieved to have "meaningful communicative competences" (Commission of the European

Communities 2003: 4) subdivided into "six ascending levels of proficiency" (Harsch 2017: 251). The differentiated levels provide the basis to evaluate student performance and their progress, following the idea to set these levels as the European standard to measure language proficiency and performance independent of the context the measurement is taken in (cf. Council of Europe 2010: 1). Arguably, language and its proficiency come with a level of complexity that makes it difficult to set standards per level. The CEFR attempts to account for that complexity by its "taxonomic nature" (ibid.). It separates the four skills of languages into the productive skills of speaking and writing, as well as the receptive skills of listening and reading and attempts to itemize each into different levels of achievement (ibid.).

Basic users (levels A1 and A2) are mainly concerned with learning the language and its use in everyday situations to come by in those (cf. ibid.: 24). More advanced learners, independent users, are thought to be able to handle situations in, e.g., educational contexts, independently (cf. ibid.). Furthermore, the increased proficiency enables them to learn further, either the language or contents or both. Lastly, as of level B2, language use should be effortless and without significant problems, "[…] more advanced levels (B2-C2) are defined in terms that imply advanced levels of educational achievements […]" (Little 2010: 22). The distinct levels are described in much more detail in the *CEFR Companion Volume with new descriptors*, released in 2018. Those descriptors further differentiate the four skills and provide more detailed information on the targeted competences per level at a skill. Reading comprehension, e.g., is subdivided by the aspect of the nature or intention of the activity: *reading correspondence, reading for orientation, reading for information and argument, reading instructions, and reading as a leisure activity* (cf. Council of Europe 2018: 62f.). For each of those subdivisions of the skill, the descriptors per level indicate the characteristics of the skill to be accomplished at that stage. It is, however, essential to understand that "the scales do not describe knowledge of language [as such], but the ability to use language in different situations" (Moe et al. 2015: 13).

According to the educational standards for the first foreign language, i.e., in the majority of cases in Rhineland-Palatinate in English, students are expected to achieve level B1 and, in parts, B2 at the end of year 10 (cf. KMK 2003: 11f.). If English is continued through higher education, years 11 – 13, levels B2 and, in parts, C1 are targeted (cf. KMK 2014: 27f.). Those levels are also legally binding as they are set in the administrative directive issued by Rhineland-Palatinate in 2014 (cf. Ministerium für Bildung, Wissenschaft, Weiterbildung und Kultur 2014b: sec. 5). Consequently, the CEFR serves as the scale to measure student performance in the foreign language. The latest English curriculum, effective as of the school year 2022/2023, adopts the levels described for the skills in the CEFR descriptors implementing this standard into teaching to help achieve the targeted language proficiency levels (cf. Ministerium für Bildung 2021). Among other factors, language proficiency has been identified as a decisive factor in educational success. Moreover, the lack thereof was determined to be one of the main reasons students fail (cf. Deutsche Kinder- und Jugendstiftung 2014: 10). Hence, developing proficiency in a foreign language should be in focus when considering options to improve schools and their educational

portfolio. One aspect identified as being very beneficial for developing language proficiency within the school context was teaching content subjects in a foreign language which also ties in with the mandate of education as outlined above (see chapter 2.1).

3 Teaching in a foreign language

The practice of teaching content subjects or non-language subjects in a foreign language comes in various realizations worldwide. In North America, specifically Canada, teaching in a foreign language started in the form of *immersion programs* (cf. Wesche-Bingham 2002: 357) as an attempt to teach English-speaking students the other official language of the country, French. Further realizations focus on minority languages, e.g., in Spain, where Catalan and Basque are the regional languages of choice or Italy, where in Trentino Alto-Adige, German is the majority language and hence, the language of choice (cf. Eurydice Report 2017: 44). Regardless the approach, all share the use of a language different from the traditional language of schooling to teach content (cf. Wolff & Sudhoff 2015: 9).

3.1 Bilingual teaching and Content Language Integrated Learning

In Germany, the importance of proficiency in a foreign language was recognized as a decisive factor for the individual's success and, with that, the economic growth and power of Germany and Europe (cf. KMK 2013 a). As described above, teaching in a foreign language has made its way into the educational landscape to further language proficiency. Bilingual education helps students, i.a., to develop "advanced functional proficiency in L2 reading, writing,

Figure 3.1 - Bilingual teaching using the language as a mean to convey content subject knowledge

speaking, and listening" (Leaton Gray, Scott & Mehisto 2018: 51), to develop competences in intercultural communication, and academic achievement in the school content subjects (ibid.). The KMK uses the terms *bilingual teaching* and *Content Language Integrated Learning (*CLIL*)* synonymously (cf. KMK 2013a: 4). Thus, the KMK adds a level of imprecision when referring to the 2006 Eurydice report and defining their approach as teaching of non-language subjects in a foreign language (ibid.). This entails only one aspect, as shown in figure 3.1, i.e., the use of the language to teach the content instead of also using the content subject to teach language aspects. Historically, the KMK's approach dates back to the time after WWII, when in 1963, the Élysée Treaty was signed, agreeing on increased efforts to reconcile, especially the people of Germany and France after the war (cf. KMK 2013a: 4.). Consequently, schools along the French-German borders started to introduce History and Geography being taught in French in Germany, making this language the primary language of bilingual education (ibid.). With the developments in Europe, the formation of the EU, and the increasing combination of economic interests worldwide, the preferred and dominant language for bilingual education shifted from French to English in the mid-1990s (ibid.: 4). In compliance with the mandate of education as laid out in the Basic Law and the state order of Rhineland-Palatinate (see chapter 2.1), and later on the European Language Policy, bilingual teaching has constantly been confronted with high expectations declaring language proficiency at a near-native level at the end of bilingual education as the goal (cf. Mäsch 1994: 28). Bilingual education is tasked with not only

increasing the linguistic repertoire based on content but using the content to also allow for the "acquisition of new knowledge" (Cammarata 2016: 124) and a deepening of understanding and critical thinking (ibid.). Additionally, learning "non-linguistic content" should be the focus (ibid.). To account for this, the KMK recommended in 2011 to increasingly employ a foreign language as the working language to intensify subject- and language-related learning, quoting both the quality and efficiency as proven benefits of bilingual teaching (cf. KMK 2013a: 6). The content subject dominates the teaching approach and hence, didactic approaches of the distinct content subjects dictate the way classes are conducted (ibid.: 7). That means in turn, that content subject teachers use the language to teach about the content[3].

Although being used synonymously by the KMK, CLIL should be understood in a different way than bilingual teaching. Defining CLIL, however, has been a complex task ever since academia started to discuss the approach. Initially, Baetens Beardsmore stated that "there is no single blue-print of content and language integration that can be applied in the same way in different countries" (1993: 3), and this still is applicable especially in Germany, where the distinct federal states form individual educational programs. Additionally, bilingual education is coined by the content subject with no shared didactic approach across all content subjects (cf. KMK 2013a: 7). The attempt of a definition of CLIL by Coyle, Hood, and Marsh as a

> dual-focused educational approach in which an additional language is used for the learning and teaching of both content and language. That is, in the teaching and learning process, there is a focus not only on content, and not only on language. Each is interwoven, even if the emphasis is greater on one or the other at a given time
>
> (2010: 1)

clearly points at the difference between bilingual teaching and CLIL. In CLIL, content-subject teaching is done "not *in* a foreign language but *with* and *through* a foreign language" (European Commision 2006: 8, original emphasis), meaning that content and language are entities of equal importance. In contrast to bilingual teaching, CLIL fosters and addresses more than the development of competence in a foreign language by constructing knowledge through the use of the language (cf. Coyle, Hood & Marsh 2010: 36). CLIL employs not only language *for* learning, i.e., the "language needed to operate in a foreign language environment" (Coyle, Hood & Marsh 2010: 37), but also the language *of* learning, i.e., the "language needed to access for learners to access basic concepts and skills" (ibid.) as well as language *through* learning, i.e., the result of learning with "new meanings […] requir[ing] new language" (ibid.). Those three types of language are combined in the *Language Triptych* (ibid.: 36) and contribute to the development of language proficiency as well as content knowledge. Additionally, content knowledge, intercultural, and communicative competences are addressed (cf. Breidbach 2015: 210). All three are always embedded in a context; usually, that of the content subject, forming one of

---

[3] Further details on how bilingual teaching is realized, especially in Rhineland-Palatinate, are in chapter 3.2.

three core entities in CLIL teaching. However, CLIL works in four dimensions: content, communication, culture, and cognition. Along with content, the use of language is addressed within communication, and both equally evoke the area of cognition, which focuses on learning and thinking processes (cf. Coyle, Hood & Marsh 2010: 41). By shifting the focus onto content outside of strictly language learning, another dimension of learning receives attention: culture. Through incorporating authentic topics into teaching, CLIL fosters the development of intercultural understanding, contributing to the goal of students becoming critical citizens capable of upholding democracy. The four dimensions combined result in "learning to use a language appropriately whilst using language to learn effectively" (ibid.: 42).
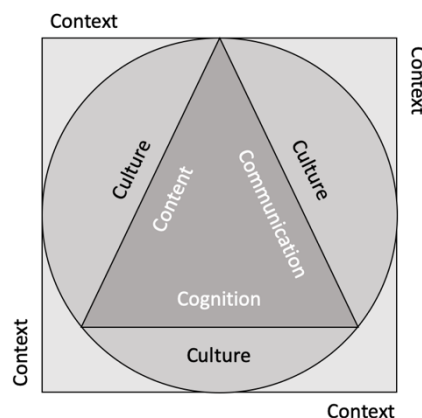
Figure 3.2 - The 4Cs Framework (Coyle, Hood & Marsh 2010: 41)

The four dimensions are subsumed in the concept of the 4Cs (see Figure 3.2) in CLIL (ibid.: 41).

Comparable to bilingual education, expectations toward CLIL as another approach to teaching in a foreign language are no less demanding than those stated above for bilingual teaching. Wolff and Sudhoff wrote that CLIL is regarded as "der Königsweg zur Vermittlung hoher fremdsprachlicher Kompetenzen[…]" (Wolff & Sudhoff 2015: 7) or the silver-bullet of language education. The European Commission even states that "CLIL has a major contribution to make to the Union's language learning goals" (Commission of the European Communities 2003: 8) and thus, attributes a mainly political aspect to the teaching approach. However, the intense combination of language, cognition, and communication in CLIL is not given in bilingual teaching; if at all, to a much lower extent. While CLIL uses language to teach the content and content to teach the language (among everything else featured in the 4Cs) (see Figure 3.3), bilingual teaching simply uses the language to follow the content subject's approaches
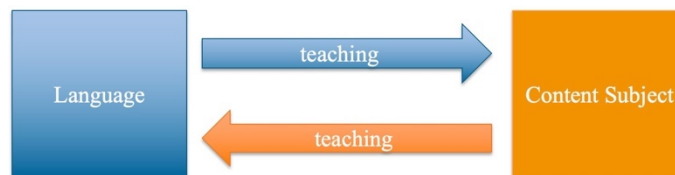
Figure 3.3 - CLIL schema for the use of language and content combined

and convey the knowledge of said content subjects (see Figure 3.1). Bilingual teaching arguably also features a language aspect, which is targeted through the content subject, but this is mostly limited to vocabulary and language chunks. CLIL*,* however, offers more than vocabulary work. CLIL offers the opportunity to use the content subject, e.g., History, to teach aspects such as tenses in the target language English, embedding something rather abstract and rule-based into an authentic context more likely to convey not only the application of the rule but also the actual meaning of the grammatical aspect. According to Albrecht and Böing, this approach was initially planned for in Germany. However, textbook editors deviated from this idea and published material that limited language work to vocabulary (cf. Albrecht & Böing 2010: 58). Hence, current textbooks cater to the requirements of bilingual teaching and do not support the dual-focused approach of CLIL.

3.2   Realization in Rhineland-Palatinate

Although the KMK uses the terms bilingual teaching and CLIL as synonyms, the term bilingual teaching will be used in this thesis when discussing the realization of teaching in a foreign language in Rhineland-Palatinate. Basically, Rhineland-Palatinate allows for three different ways to realize bilingual teaching in secondary schools. Schools either offer a complete bilingual track, teaching content subjects in a foreign language throughout the entire period of years 7 to 10, or 13. Alternately, schools may also offer single school years during which selected subjects are taught bilingually. Lastly, it is also possible to teach distinct modules throughout the school year as a bilingual module in various subjects (cf. KMK 2013a: 8). With the former approach, students can graduate with both the German *Abitur* and an equivalent international certificate (cf. Ministerium für Bildung, Wissenschaft, Weiterbildung und Kultur 2011: sec. 2.5). This is not possible with the latter approaches offering only temporal bilingual teaching or bilingual modules.

As of 2017, a total of 54 secondary schools offer bilingual teaching in Rhineland-Palatinate (cf. Bildungsserver RLP 2017) in the form of a bilingual track; a number that has very likely increased by 2022. English is the target language at 33 of those schools, while French is offered at 17 schools (ibid.; cf. KMK 2013a: 13). As defined by the KMK, Geography, History, and Social Sciences form the core of bilingual teaching; further subjects are possible but require permission issued by the supervisory school authority (cf. Ministerium für Bildung, Wissenschaft, Weiterbildung und Kultur 2011: sec. 1.1). In 2013, however, the KMK also listed Arts, Music, Biology, Physics, Chemistry, Mathematics, and Physical Education as subjects on offer bilingually (cf. KMK 2013a: 14). The KMK, however, does not differentiate as to whether this considerably increased offer includes all three types of realization or if it is realized through select modules rather than teaching those subjects continuously. Regardless, the teaching of those subjects begins in grade 7. During grades 5 and 6, preparatory courses are offered, featuring content of all bilingual subjects on offer at the respective school (cf. KMK 2013a: 8). These preparatory courses also serve for the selection process in place in Rhineland-Palatinate which allows schools to select students based on their overall performance in school, their performance in the target language, and in specific, their performance in those preparatory courses in addition to students actively applying (cf. Ministerium für Bildung, Wissenschaft, Weiterbildung und Kultur 2011: sec. 2.2.3). Once the actual bilingual teaching starts, schools are obliged to offer various subjects in their bilingual tracks (cf. KMK 2013a: 8). The core bilingual subjects are regularly taught for two lessons per week. Those two lessons are reserved for teaching the subjects in the target language. An additional lesson is held in German to ensure that all concepts and terms are also known in the traditional language of schooling (cf. Ministerium für Bildung, Wissenschaft, Weiterbildung und Kultur 2011: sec. 2.3.1). This is necessary to enable students to take their *Abitur*[4] in this traditional language of schooling since it is not possible in the foreign language. However, students within bilingual tracks can take an additional oral

---

[4] the equivalent to the British A-levels, a general qualification for university entrance

exam in the foreign language, which is certified in addition to the actual graduation record (ibid.: sec. 2.5.1).

Independent of the grade, assessment in bilingual tracks is only concerned with content knowledge and students' performance herein. The content subject and its content-related standards serve as the basis for assessment (cf. KMK 2013a: 15). Only if the quality of the language leads to a disruption in the quality of the content, the grade may reflect on those language problems. The extent to which this language problem influences the assessment of the content is defined for similar problems in the subject's standards for assessment and refers to the traditional language of schooling, German (cf. Ministerium für Bildung, Wissenschaft, Weiterbildung und Kultur 2011: sec. 2.3.2). Although language is basically not graded in bilingual tracks, the so-called DESI study, a study looking into students' performance in language awareness, reading, listening, and writing in grade 9 in both English and German, found out that students in bilingual tracks perform significantly better in the tested areas than their peers in non-bilingual tracks. Especially their listening skills were twice as developed as those of the peer group, leading to an advancement, expressed in school years, of one to two years (cf. DIPF & Leibniz-Institut für Bildungsforschung und Bildungsinformation 2020: 59). It has to be considered, though, that this study only measured short term effects of teaching in a foreign language and failed to integrate assumed effects through preparational courses students partake in in years 5 and 6 as well as the so-called "creaming of top"-effect described by Rumlich referring to the selection process of students prior to joining bilingual programs (cf. Rumlich 2016: 226). Hence, the diagnosed strengths have to be regarded as an effect of both increased contact time with the target language as well as improved or rather increased initial qualification among the selected students, i.e., 'a priori' differences (ibid.: 44, 448).

For teachers, additional training to teach their content subject in a foreign language is not mandatory; however training is offered in the post-exam phase of teacher training (e.g., Studienseminar Koblenz 2013). Although this program is an additional voluntary qualification, potential participants must apply and undergo a selection process (ibid.). If possible, schools are to mandate teachers who have such additional training; teachers without this training have to prove language proficiency equivalent to that of trained language teachers (ibid.). The requirement of teaching skills in the foreign language, however, is not explicitly mentioned (ibid.). Nevertheless, future teachers and teachers take up the opportunity for additional training. On average, approximately six teachers in training participate in the abovementioned voluntary program in Rhineland-Palatinate (cf. Schröder 2022). In 2020, a combined total of 186 teachers and teachers in training completed an additional training to teach bilingually (ibid.). In relation to the 330 teachers in training for English in 2020 in total, of which 175 went for *Gymnasium*, the average of six teachers in training across the various school types receiving this training is soberingly low (cf. Statistisches Landesamt Rheinland-Pfalz 2020a: 12). Likewise, the 186 teachers who were recorded to have an additional training for bilingual teaching are the minority among the 5047 English teachers employed in Rhineland-Palatinate during the schoolyear 2019/2020

(cf. Statistisches Landesamt Rheinland-Pfalz 2020b: 25). Although in accordance to the statistics agency, the numbers of participants of the teacher training offer are constantly increasing (cf. Schröder 2022), it may be assumed that often the availability of teachers to teach the content subject in a foreign language influences the selection of subjects offered at schools.

Additionally, the subjects mentioned above as core subjects are those for which comprehensive teaching material, i.e., textbooks, is available. That aspect may be further influencing the choice of content subjects. Research on Rhineland-Palatinate secondary schools, more precisely their web pages and textbook lists, revealed that despite the extensive list of subjects published by the KMK, the subjects offered most often are History, Geography, Social Sciences, and Biology. Out of a compiled list of schools offering bilingual tracks, provided by the *Bildungsserver* Rhineland-Palatinate of 2016/2017 (cf. Bildungsserver RLP 2016), only 28 schools actually offered the information on subject choice and textbooks used online (see chapter 12.2). This figure cannot be seen as entirely up-to-date or reliable, much rather as an indication regarding the choice of subjects available. This selection is further supported by the abovementioned aspect of available, ready-made teaching material. Comprehensive textbooks for Biology, Geography, and History are available and designed to comply with the language skills at the given grades, however, the degree to which this is realized is also the subject of research in this thesis.

4   Teaching, text, and language

As described, bilingual teaching in Rhineland-Palatinate or the choice of subjects on offer in bilingual strands is, i.a., influenced by the available teaching material (cf. chapter 3.2). In general, textbooks are the main source of teaching in Germany (cf. Michalak, Lemke & Goeke 2015: 86; Oleschko & Moraitis 2013: 13). Accordingly, publishers develop the textbooks to match the level of skill and slightly challenge it per grade, following the educational standards per subject and in specific, for languages, the administrative directive issued by Rhineland-Palatinate in 2014 (cf. Ministerium für Bildung, Wissenschaft, Weiterbildung und Kultur 2014b: sec. 5). Although the educational standards set strict guidelines for publishers and schools in their selection of textbooks, there is no guideline that requires the selection committee in schools to look at the (linguistic) demand the textbooks impose. If teaching is centered around these textbooks, looking at texts and their language in teaching should be a mandatory step to understand the language requirements those texts actually do and should impose on students.

4.1   Text in teaching

"Teaching and learning in the classroom are mainly carried out through talk" (Llinares & Morton 2017: 244) supported and fed by, i.a., texts. Accordingly, textbooks or moreover the "[…] 'importance of the textbook cannot be overestimated.' […]" (Sudhoff 2015: 271). All discourse, be it student contributions, additional material, or tests, is based on text (cf. Zydatiß 2013: 319). In school, subject-related knowledge is mainly presented and transmitted through textbooks and written material (cf. Michalak, Lemke & Goeke 2015: 86; Oleschko & Moraitis 2013: 13). Since texts and their input are the *sine qua non* in (bilingual) teaching (cf. Gass & Mackey 2020: 194) and serve as a medium to convey essential information on the content subjects rather than being the subject of text analysis only (cf. Krechel 2010: 159), they are designed to provide information to retrieve and process during class (ibid.). In bilingual teaching, texts carry content-related information to be extracted and processed while in language teaching, text is often interpreted or analyzed for the employed grammatical phenomena (ibid.). Other than spoken discourse, written text primarily employs formal, formatted, and rule-based language (cf. Zydatiß 2013: 159). Furthermore, texts differ with regard to their type or intention, fulfilling different functions which depend on the subject and the teaching approach (cf. Llinares, Morton & Whittaker 2012: 110).

Among the four subjects discussed in this thesis, texts share the importance to fulfill the mandate each subject is tasked with and to convey subject-related vocabulary. Beyond the mandate, texts have different tasks or functions for each subject. In science subjects, such as Mathematics or Biology, various types of text can be found, which are also identified in the framework of *Systemic Functional Linguistics* (cf. Llinares & Whittaker 2010: 126). They differ regarding their intention, describing either *procedures*, i.e., giving instructions on what to do and how to do it, or *reports* to structure or organize content knowledge (cf. Llinares, Morton & Whittaker 2012: 112). A third text type explains selected scientific processes (ibid.). In Mathematics, e.g., a text comes in "a form of language that has been invented […]

to discuss abstract concepts" (Metsisto 2005: 11), which leads to high information density, with "more concepts per sentence and paragraph than any other type of text" (ibid.). This type of text is classified as a *procedure* in Mathematics (cf. Llinares, Morton & Whittaker 2012: 112), featuring primarily instructions in the imperative form. Usually, texts in Mathematics represent word problems, which are supposed to teach students to decode text and transfer the information onto mathematical concepts, corresponding procedures, and calculations – a core skill in the subject (ibid.: 9). Additionally to this decoding, texts in Mathematics can prove especially challenging as vocabulary can adopt a completely different meaning when used as subject-specific terms than in everyday language (cf. Kremer 2017: 39). Another aspect of a text in Mathematics is that often visuals, e.g., graphical representation of a problem or symbolic depiction of mathematical problems, accompany the text to support understanding.

In Biology however, visuals and text combined represent the information required for learning (cf. Ziepprecht et al. 2017: 116). Content knowledge is conveyed through the combination of continuous and discontinuous texts, which have to be processed based on prior knowledge and reading skills to extract, decode, and relate the information (cf. Olthoff 2018: 4). A unique aspect of texts in, i.a., Biology, is that described aspects can often be reproduced in experiments, making the content tangible for students (cf. Bohn & Doff 2010: 78). While in Mathematics, most texts fall into the procedure category, texts in Biology most often give reports*, describing what things are (cf. Llinares, Morton & Whittaker 2012: 120), or *explanations* giving reasons for and the manner of select phenomena in the subject (ibid.)*.

In Geography, texts primarily contribute to the ability to understand subject-specific language and vocabulary and, with that, the ability to partake in subject-related discourse (cf. Vollmer 2010: 245, 251). All Geography texts inform on subjects, events, circumstances, or societal challenges to enable students to evaluate the given information (cf. Kuckuck & Röder 2017: 232). Text types in Geography differ from those in science subjects, although sharing the basic idea of organizing, describing, and explaining subject-related aspects. A frequent type of text used in textbooks in Geography is *descriptive reports* (cf. Llinares, Morton & Whittaker 2012: 126)*. The general idea of such descriptive reports is to "organize relevant information" (ibid.: 127) through structured description of, e.g., actual situations. Texts in Geography are rich in subject-specific vocabulary, specifically incorporating definitions, and thus, are considered challenging when employed in teaching (cf. Kuckuck & Röder 2017: 232). Furthermore, they display a rather descriptive character to convey information accompanied by maps, charts, images, and graphs. In part, this visual material is to support the understanding of the text, and in part, the text helps to understand the visual. While both, text and visuals, oftentimes complement one another, presenting students with only either of the two would most likely be overtaxing. In combination with visual support, text can also be of explanatory character, giving either *sequential* or *causal* reasons for events, e.g., natural disasters like volcano eruptions (cf. Llinares, Morton & Whittaker 2012: 129). In cases dealing with the effects of "natural or human activity" (ibid.: 130), explanations are rather *consequential.* Similar to Biology, the content conveyed through text can be experienced personally,

e.g., the seasons and numerous topics are present in students' living environment. This aspect, however, is not given for texts in History.

Texts in History hold a unique position as the subject matter is only accessible through text or, as Goertz puts it, language and history are inseparable (cf. Goertz 1995: 147f.) "with so much of its [history's] evidence embodied in texts" (Llinares, Morton & Whittaker 2012: 132). Textbooks in History feature usually four different types of text: the *period study, recounts in history, historical accounts,* and *historical explanations* which can be presented in, i.a., descriptions written by textbook authors, secondary literature, or source texts. It is important to understand that textbook material is primarily based on sources and provides an interpretation of those as sources are the remainder of records of the past still present today (cf. Pandel 2012: 10f.). These sources construe a reality that no longer can be retraced or reproduced (cf. Maset 2015: 86) but evaluated based on the knowledge obtained through reading both primary and secondary literature. And although these sources are essential to modern History teaching, they are frequently thought as an add-on to secondary literature (cf. Maset 2015: 80). With the help of period studies, students learn about the characteristics of a historical period (cf. Llinares, Morton & Whittaker 2012: 133) as texts provide a description of the era given. Recounts in History aim at helping students to understand the intricacies of a distinct period or societal aspect during this period (ibid.:135). History textbooks most often employ *biographical recounts* to convey information on significant events. Additionally, third-person recounts allow for a sequential structure making knowledge easily accessible for students (ibid.). Historical accounts develop biographical recounts further as they move to a more abstract level. In these historical recounts, an ascribed sequence no longer provides a text with a structure to the same extent as it did with biographical recounts*.* Moreover, the aspect of *cause* gives reasons for events and the order in which they happened (cf. Llinares, Morton & Whittaker 2012: 138). Having dealt with causes for events, texts on the next level are historical explanations. Events in this genre are no longer presented in a chronological manner but are explained along causes and consequences, allowing for "a complex view of events" (cf. 2012: 139). Despite seemingly increasing complexity along the types of text, the period study, recounts in history, historical accounts, and historical explanations, all four are already featured in early history teaching textbooks (ibid.).

Following the *Systemic Functional Linguistics* framework as applied by Llinares, and Whittaker (cf. 2010: 126f.), the text type also influences how easily accessible and extractable information is. In comparison, the texts in the four different subjects and how those texts and the language contain information can be differentiated between *indirect* and *interpretative* information to *explanatory* and *direct* information and lastly to *instructive* and *imperative* texts. In History, especially with source texts as recounts or historical accounts, information is usually embedded in interpretations (cf. Pandel 2012: 10f.). Thus, source texts in History textbooks fall into the category of providing rather indirect and interpretative information (cf. Llinares & Whittaker 2010: 127f.). The texts concerned with the topic at hand, the informative texts provided by the editors, present information rather directly. Geography

textbooks provide information directly through informative texts, while Biology and its texts employ information access both through informative texts with direct information and instructive and imperative texts. Mathematics material clearly provides information in the form the latter type of text (ibid.).

All textbooks usually feature graphical presentations to accompany or supplement the written parts. While in History textbooks, illustrations, caricatures, or graphs mainly serve the purpose of deepening thematic inquiries and discussions with a more interpretative nature to them, in Geography, maps, graphs, and illustrations have a rather explanatory character. In Biology, illustrations serve the purpose of both conveying new information and explaining aspects discussed in the text. For Mathematics, however, illustrations in the form of graphs, geometric figures, angles, i.a., fulfill an explanatory and supportive function to clarify the requirements outlined in the tasks. These different types of illustrations and their intention indicate how, with the additional non-textual input, information from the text and their processing can be differentiated in accordance with the Systemic Functional Language framework between *thought-provoking, investigative and interpretative, descriptive,* and *explanatory and supportive* (ibid.)*,* with illustrations in History to be placed in the first category, Geography between the first and second category, Biology between the second and the third category, and Mathematics in the third category.

In summary, in Mathematics, texts are central to developing the analytical skills needed. Graphical presentations are meant as support. In Biology, text and visuals are equally important to describe (mostly) tangible aspects. Descriptions of tangible aspects are also the main focus of text in Geography. Additionally, texts are thought to serve as the basis for evaluations on the side of students. Other than the former three subjects, in History, the text is the subject matter. Texts fulfill a dual function in History as on the one hand they are the medium through which information is transported and on the other hand are the most important access to history and historical events. Described aspects can no longer be relived or reproduced, and everything depends on what was handed down from the past. The mandate to encourage critical thinking is more prominent in History and the texts provided than in other subjects. Another shared aspect, however, is that all knowledge conveyed in texts is only accessible if language competences are sufficient to process the text (cf. Piaget 2010; Heine 2010: 203; Lötscher 2006: 19; Beck & Klieme 2007: 11).

Language competence is a construct of multiple single competences. Linguistically, language is a system combining phonology, morphology, and syntax along with pragmatics and sociolinguistic aspects (cf. Schwarz 1996: 40f.). Knowledge about these components of language and the ability to apply the knowledge in combination is what language teaching should convey to increase language competence. Accordingly, language competence on the one hand describes the level of knowledge of a language, e.g., in terms of grammar, lexis, and register; on the other hand, though, it is the ability to apply this knowledge in either receptive or productive activities (cf. Beck & Klieme 2007: 11). When it comes to reading texts, the required language competence is on the receptive side focused on processing and understanding information (cf. Nold & Rossa 2007: 197). The process of reading consists of the

decoding of the written signs, the application of prior knowledge, and lastly the integration of newly obtained information into the existing concepts and repositories of knowledge (ibid.), or as Ellis and Robinson put it, "cognition and language create each other" (2008: 3). Reading in the foreign language can thus be limited by way of limited lexis or limited prior knowledge on the language as such and its application (cf. Nold & Rossa 2007: 198). In cases of disruption in the reading process, Nold and Rosse cite Alderson, who concludes that most often a problem with the language as such is the root (ibid.). Understanding a text depends not only on the reader as such but also on the text, whereas the reader is the most decisive factor (cf. Demarmels: 105). Now, if in bilingual teaching content-subject knowledge is to be obtained through the reading of subject-related expert texts to both foster language and content learning (cf. Zydatiß 2013: 316), texts have to be at a level of complexity that allows for both. Clearly, the teacher's task is to mediate between the level of the text and the level of student's language competence (cf. Leisen 2015: 225) but teachers would have to be aware of the level of complexity a text holds.

Text complexity can be measured based on multiple key figures. Starting with obvious properties, e.g., the length of a text and its nature give a first impression on what the text demands. Furthermore, the employed vocabulary, grammatical features, the accessibility of information (direct/indirect), the length and structure of sentences, the number of words per sentences, and the distribution of information are features that have to be analyzed to determine the complexity of text (cf. Green, Khalifa & Weir 2013: 25). Those correspond to the criteria of *readability* and *intelligibility* (cf. Demarmels: 108) with readability describing a style in writing that allows for easy comprehension and that encourages the reader (cf. DuBay 2004: 3). Readability can be measured. A formula still considered as accurate is the so-called *Flesch-Reading-Ease Formula* (cf. Ziafar & Namaziandost 2019: 816) which calculates an index value with the help of the average sentence length and the average number of syllables per word (Flesch 1948) (see 6.3.4). Additionally to reading ease, the vocabulary is a "strong predictor for text difficulty" (Chall & Dale 1995: 86). In general, vocabulary metrics provide the most consistent distinction when it comes to text complexity (cf. Leaton Gray, Scott & Mehisto 2018: 26). Vocabulary or much rather the complexity of vocabulary can be measured in various ways. Texts can be analyzed for *lexical sophistication*, that is the ratio of common vs. uncommon words used in a text and giving information on the volume of a learner's active vocabulary and the ability to use it (cf. Malvern & Richards 2002: 85). To evaluate, different corpora can be consulted to determine the frequency and connected level complexity per word item (see 6.3.1). Furthermore, *lexical diversity*, the grade of variation of vocabulary used in a text (cf. Johansson 2008: 62), also indicates the complexity of a text in regard of vocabulary. Again, various key figures can be calculated to arrive at a measurement and evaluation of this aspect (see 6.3.2). Another indicator in connection to the vocabulary used in a text and the resulting complexity is the relation of content vs. function words. This key figure is called *lexical density*, which indicates the way information is packaged (cf. Johansson 2008: 65) within a text, meaning that a text with a high number of content words transports significantly more information than a text

with a high number of function words (ibid.). All those key figures are computed on a word or sentence basis and require meticulous analysis of the text which can be accomplished through automated web-based analytical tools. In teaching, however, working with a text does not stop at reading it. Working with texts becomes beneficial when information and contents are discussed in class.

4.2 Speaking and writing – the productive use of language in teaching

Interaction in the classroom, most often spoken, is a decisive and essential part – a condition even – to facilitating learning (cf. Dalton-Puffer 2007: 276). As Moe et al. put it

> all students need to master the varieties of academic language used in different subjects matters, and to be able to interact fluently with peers. Therefore, students need to develop a good command of different language skills in order to learn and be successful in all school subjects
>
> (2015: 15).

Through classroom discourse, a process is set in motion during which students internalize the information presented to them through social interaction. This process requires text as input and language to talk about this input. Consequently, spoken language is used as a tool to verbalize thoughts concerned with the content or information presented (cf. Zydatiß 2013: 317) or as Butzkamm put it "*Sprache im Unterricht ist das Werkzeug, das man gebraucht, während man es noch schmiedet*" (Butzkamm 2002: 110), meaning that language is being used as tool while students are still developing their proficiency. In content-subjects, language furthermore serves as a mediating factor to convey subject knowledge presented in expert language on a level students can understand (cf. Leisen 2015: 225). In addition, expert language does not focus on discourse (ibid.), while in teaching, discourse is vital. Generally, teaching has to treat subject knowledge and language as equal entities to foster communication and discourse, leading to actual knowledge acquisition rather than learning things by heart (ibid.: 226). Of the two, the acquisition is considered more valuable (ibid.). But the role of language in education is not only a means to transmit knowledge but also to create "structure and assist the training [...] of social actors" (Beacco 2007: 3). In CLIL but also in bilingual teaching, the role of language is even more central as it is regarded as "language learning for language usage" (Coyle 2011: 55). Therefore, language use in the classroom has to be differentiated for its function. Either it is used for social interaction or to transmit knowledge (cf. Leisen 2015: 225).

Knowledge, however, is not solely scientific or academic. According to Bernstein, knowledge is subdivided into *horizontal* and *vertical* knowledge leading to *horizontal* or *vertical* discourse. While horizontal knowledge and discourse refers to "every day or 'common-sense' knowledge" (Bernstein 1999: 159) and is accessible to a broad public, the term vertical knowledge describes a "principled, hierarchically organized" structure of knowledge or "specialized in criteria for the production and the

circulation of text" (ibid.: 160). Horizontal knowledge is believed to more likely be found in the humanities while vertical knowledge is said to be more prevalent in the natural sciences (ibid.). Although horizontal knowledge and discourse seem to be closely related to social relationships and interpersonal discourse, it cannot be regarded as commonly accessible. The distinction between those two kinds of knowledge is essential for bilingual teaching because students have to learn to employ both in discourse in a controlled and leveled manner (cf. Llinares, Morton & Whittaker 2012: 39).

Bernstein's differentiation of knowledge and discourse can also be connected to Cummins categorization of language into *Basic Interpersonal Communicative Skills* (BICS) and *Cognitive Academic Language Proficiency* (CALP). To Cummins, CALP is mainly comprised of literacy and vocabulary knowledge; BICS, on the other hand, can be compared to "the ability to tell jokes effectively" (Cummins 1999: 3). While Cummins stresses that both categories are acquired through conversational interaction, he also states that in L2 learners, CALP proficiency can precede BICS proficiency from the lack of using the language in non-academic or non-educational social interactions (ibid.: 4). Connecting BICS and CALP to horizontal and vertical discourse, vertical discourse can be seen as building concepts and constructs on one another to arrive at overarching subject-related theory (cf. Llinares, Morton & Whittaker 2012: 39), which primarily engages language displaying CALP-quality. The aspect of social interaction within a culture ties BICS to Bernstein's horizontal discourse which is acquired through participation in local practices. Other than vertical discourse, horizontal discourse does not differentiate between right and wrong (ibid.: 41). The observation that with increasing class levels, the ratio of vertical discourse increases as well ties in with Cummins' take on CALP proficiency being a result of lifelong learning (or at least throughout schooling) (cf. Cummins 1999: 3). BICS, on the other hand, or much rather the language acquired through social interaction outside of the academic field, "reach at a plateau in the first six or so years" (ibid.).

Transferring Bernstein's differentiation of discourse onto the various school subjects, vertical discourse can mostly be found in science subjects like Mathematics, Physics, or Biology (cf. Bernstein 1999: 39). Horizontal discourse employs primarily social interactions like discussions, the reiteration of personal experience, or arguments for or against a viewpoint and can mostly be found in subjects of the humanities, such as History or Sociology (ibid.: 41). Although Bonnet argues that in Biology interaction is a sine qua non, he also declares the overall mandate of language use in this subject to foster *scientific literacy* (cf. Bonnet 2015: 166f.). This literacy can be achieved in four categories:

(1)   Knowledge of scientific concepts and principles
(2)   Knowledge of scientific methodologies
(3)   Knowledge of the nature of natural sciences
(4)   Knowledge of the relation between natural sciences, technology, and society

(ibid.).

All four contribute to the ability of "talking science" (ibid.) which Bonnet regards as a stage between everyday language and scientific language. Talking science, consequently, engages rather in vertical discourse, focused on CALP and the addressing of so-called *higher order thinking skills* (ibid.: 170). Social interaction is limited in this kind of discourse and constrained by the overarching target to arrive at said scientific literacy. Mathematics, another science subject discussed later in this thesis, is characterized as equally vocabulary prone (cf. Hancewicz 2005: 83). The discourse here is focused on giving and understanding definitions to the technical terms and, thus, can be characterized as vertical discourse (ibid.). In theory, Mathematics allows for three types of discourse:

(1)   the traditional teacher-dominated discourse

(2)   the probing discourse, which involves students but is still teacher-centered

(3)   the rich discourse, which leaves the negotiation of meaning mainly to the students

(ibid.: 72).

Although Mathematics seems to allow for discourse when employing rich discourse, which could be seen as horizontal and allowing for BICS, most instructions follow the traditional or probing approaches and thus, use primarily vertical discourse structures. When it comes to rich discourse, Mathematics at most involves the reconciliation of the meaning of technical terms among students (cf. Baker 2011: 170). Prototypical situations for discourse can thus be summarized as a presentation by the teacher, related teacher-learner interaction, and learners reading and studying the text and having related teacher-learner interaction. Accordingly, spoken language competences are located with descriptions, explanations, and summarizing (cf. Vollmer 2010a: 14). According to Bohn and Doff, discourse in Biology mostly comprises precise descriptions and logical expressions employing a rather impersonal language leaving little to no room for emotional comments, humor, or irony (cf. Bohn & Doff 2010: 79). The same holds true for Geography, which in higher grades incorporates subject-related discussions but mainly focuses on the development of a cognitive-academic register regarding language competence as subject competence (cf. Vollmer 2010: 243). Accordingly, language in the sciences is more part of a vertical discourse with language that resembles CALP rather than BICS.

The humanities derive knowledge not primarily from vocabulary and definitions but from discourse on the negotiation of meaning and interpretation of a text (cf. Wannagat 2013: 203). The contents humanities deal with also differ from those in the sciences. Humanities frequently discuss, i.a., political agendas (past or present), public discussions, eyewitness reports or personal accounts, and heritage. Although the content is either "intrinsic to social life" or based on "media use, access to knowledge and the formation of knowledge" (Beacco 2007: 5), discourse leans towards being horizontal, in which language proficiency is required and acquired through the participation in local practices (cf. Llinares, Morton & Whittaker 2012: 41). Accordingly, discourse in the humanities is often concerned with, i.a., descriptions, explanations, but also with arguments for personal or third-person

points of view, general narrations, and conflictive discussions on content matters (cf. Beacco 2007: 8f.). Contrary to the natural sciences, the humanities actually require the use of emotional comments, e.g., and especially in History, metaphors are an important aspect. One prominent example for knowledge gain and creation of awareness of varying perception of historic events (called *Standort* in German history science (cf. Hasberg 2004: 229)) is the different metaphors used for the ethnogenesis from the northern parts of Germania to the South in the 4th and 5th century AD. In German history, this movement of tribes and peoples into the territory of the Roman Empire is described in a metaphor called *Völkerwanderung*. This term conveys the idea of a peaceful movement of a large group of people. In Anglo-American history, the same movement, which eventually led to the downfall of the Roman Empire, is called *Babaric Invasion*, conveying a different perception of the events. To decode and understand the differences, students would most likely need more than just the mere definition of the terms used. Moreover, students would negotiate the meaning and intent of both terms through horizontal discourse, probably discussing the differences with a change of perspective and the emotions pegged to each viewpoint. This negotiation would also tie in with Bärbel Diehr's *Integrated Dynamic Model* (IDM), which argues that in classroom discourse the aspect of differing key concepts of subjects has to be respected and addressed when teaching a content subject in a foreign language (cf. Diehr 2012: 72–74).

Consequently, subjects of the humanities require a comparably higher share of BICS for classroom discourse. This higher share is not limited to spoken interaction in class but includes also text, especially sources. Lamsfuß-Schenk, among others, points out that bilingual History classes require students to process and elaborate on information much more thoroughly than they would in the actual language of schooling (cf. Lamsfuß-Schenk 2008: 247ff.). This processing and elaboration are done through discourse, horizontal discourse, and negotiating meaning with the help of interpersonal language.

To further a structured understanding of discourse in the classroom, its quality, and the category of language employed, the types of communicative approaches by Mortimer and Scott can be used to systematize communication in the classroom. It is differentiated into two dimensions, both of which have two opposites: *dialogic* vs. *authoritative* and *interactive* vs. *non-interactive* (cf. Mortimer & Scott 2003: 35). Dialogic communication describes classroom discourse during which students can develop and discuss their own ideas and theories on a matter; on the contrary, authoritative communication entails the explanation and elaboration on commonly endorsed theories, scientific principles, or natural laws (cf. 36f.). In interactive settings, discourse engages numerous participants, while in non-interactive settings, communication is limited to a single person intervening (ibid.: 38f.). Bringing Bernstein's differentiation of knowledge and discourse into this structure on communication, the four aspects can be displayed in the form of a grid (see Table 4.1). In this grid, the distinct categories of dialogic and interactive discourse settings are most likely to employ a high share of horizontal discourse to transmit horizontal knowledge; thus, a high proportion of BICS is involved. The share of BICS and horizontal

discourse decreases when communication in the classroom is located in the dialogic non-interactive quadrant, yet still, students' own ideas dominate the discourse. As soon as the quality of ideas discussed moves from own ideas to commonly endorsed knowledge and theories, the communication turns authoritative and vertical knowledge prone to CALP is involved. In an interactive setting, in which a group of students is involved in the communication, the share of CALP and vertical knowledge is most likely smaller than in a non-interactive setting. Based on the characterization of language in the content subjects above, science subjects fit into the authoritative category while the humanities seem to fit into the dialogic category.

Additionally, Mentz's structures for spoken discourse in class, along the lines of addressed or required language competence and the methodological focus of the topic or content subject, add another aspect of classifying classroom discourse. For the methodological focus, he differentiates between practical, experimental, and descriptive or problem-oriented focus and allocates the required language competence to the three foci. For subjects with a practical focus, mainly receptive language skills are required if students are to react only physically. Experimentally oriented subjects, such as science subjects, require controlled or limited productive skills, while problem-oriented, descriptive subjects require sufficient proficiency to engage in freely productive discourse (cf. Mentz 2015: 259).

The categorizations of knowledge, spoken discourse, and BICS or CALP, can also be transferred onto Mortimer and Scott's schema of types of communication so that a comprehensive overview of classroom discourse evolves. Depending on the quadrant and the thus implied combination of communication types, the afore-discussed distinct features of discourse are prevalent to varying degrees. In the table, bold print implies prominence or degree of aspect discussed per category. e.g., in dialogic and interactive discourse, BICS is most prominent, discourse usually evolves around problem-oriented aspects to which students contribute in a freely productive manner and horizontal knowledge based on "every day or 'common sense'" (Bernstein 1999: 159).

	interactive	non-interactive
dialogic	**horizontal knowledge** **BICS** **problem-oriented /** **freely productive**	horizontal knowledge BICS descriptive/ freely productive
authoritative	vertical knowledge CALP experimental / limited productive	**vertical knowledge** **CALP** **experimental /** **limited productive**

Table 4.1 - Mortimer and Scott's four types of communication, adapted and amended (2003: 35)

If the two distinct disciplines, the humanities and natural sciences, were to be allocated with this grid and its characteristic attributes per quadrant, the humanities are most likely allocated with either

interactive or non-interactive discourse in dialogic communication while the natural sciences rather employ an authoritative communication. Granted, communication and classroom discourse in both disciplines can also feature elements of the quadrants identified to be less prominent (cf. Llinares & Morton 2017), yet still, according to Mortimer and Scott, a rather clear allocation of the humanities and natural sciences along with the various content subjects derived from the two is possible (cf. 2003: 35f.). In fact, the content subjects discussed in this thesis, show clear tendencies toward the distinct categories in Mortimer and Scott's structure on communication (see Table 4.1 and chapter 8.4).

The above-described discourse in class is initiated by so-called discourse functions, activity verbs in their imperative form. Every discourse function initiates oral production to contribute to communication in the classroom. The level and extent of such contributions differ not only in dependence on the discourse function as such but also on the subject in which said discourse function is used. Discourse functions can occur in both tasks and exercises. While tasks typically require a certain level of planning, effort, and completion within a specific timeframe, exercises are generally designed to practice or enhance a distinct skill set. This thesis, however, only uses the term "task" as especially in teaching and discourse in a foreign language, an exercise is not just practice of a single skill but also requires planning on the language part, e.g., when results are to be discussed. In other words, the aspect of working in a foreign language adds an additional layer to exercises which qualify these as tasks in a broader sense.

Tasks in Biology are designed to contribute to basic knowledge of the subject and to help to interconnect information and concepts across the subjects of natural sciences (cf. Ministerium für Bildung, Wissenschaft, Weiterbildung und Kultur 2014a: 8). Furthermore, subject-related competences such as experiments, hypothesizing, and subject-appropriate communication are addressed through the tasks. Communicative skills to be developed are concerned with subject-appropriate language application and vocabulary along with a certain extent of systematicity when it comes to communicating on routine procedures (ibid.). Accordingly, tasks in Biology and the language used to carry out these tasks can be allocated with the authoritative type of communication, with mostly interactive character. In class, discourse mostly deals with vertical knowledge and CALP vocabulary aiming at scientific literacy.

For Mathematics, tasks have different functions. Not only form the tasks the text available for the subject, i.e., not an addition but the main aspect, but also are tasks aimed at practicing routine procedures, applying a systematized process to solving problems (cf. Ministerium für Bildung, Wissenschaft, Weiterbildung und Kultur 2007: 6). Moreover, the mandate of developing communicative skills in Mathematics also adheres to systematized or pre-set patterns of communication, a fixed set of rules to be followed and the use of mainly subject-related language and vocabulary (ibid.: 7f.). Given those characteristics, Mathematics, its tasks, and the appendant discourse can be placed with the

authoritative, non-interactive communication type, engaging the negotiation of vertical knowledge using CALP in a limited and controlled manner.

The tasks in Geography aim at supporting the subjects mandate to foster the development of awareness of sustainability, the future- and value-oriented behavior, which is supposed to be accomplished through problem-oriented tasks and methodological competence (cf. Ministerium für Bildung, Wissenschaft, Weiterbildung und Kultur 2021: 27ff.). Tasks, therefore, are focused on relevant (basic) content with a clear directive to be cognitively activating (ibid.: 28). Geography cannot be clearly allocated with the types of communication outlined above (see Table 4.1), mainly focusing on the ability to understand subject-specific language and vocabulary and, with that, the ability to partake in subject-related discourse. Hence, tasks in Geography are mostly about vertical knowledge conveyed through authoritative communication with a CALP focus. At the same time, tasks are also partly designed for students to discuss interactively, apply dialogic communication, and develop BICS. Language production thus varies between freely productive and limited and controlled.

While Geography tasks can require all four types of communication, i.e., dialogic and authoritative in combination with interactive and non-interactive aspects, tasks in History mostly require dialogic and interactive production. The curriculum clearly states the mandate to foster historical awareness and knowledge, to help connect the past to the present, and to promote the development of a set of competences through the work with texts and tasks (cf. Ministerium für Bildung, Wissenschaft, Weiterbildung und Kultur 2021: 81f.). As was stated above (see chapter 4.1), language in History has a special role as the information on the subject matter is mainly available through text only. Those texts, especially sources of oral history being transcripts of narration, are often comprised of a language that is more likely to be used in a social, interactive context. The competences to be developed are the ability to narrate, interpret, assess and judge, and develop awareness on historical culture (cf. Gautschi 2015: 48–65). Accordingly, tasks designed to address these competences and the development thereof mostly require horizontal knowledge to negotiate meaning and "challenge[…] the student to interpret, analyze, or manipulate information" (Newmann 1990: 44) in a freely productive manner. At the same time, tasks needing reiteration of information and/or the application of processes with high similarity to already learned and practiced ones (ibid.) require rather limited or controlled production. With the characteristic features of communication and classroom discourse discussed above and displayed in Table 4.1 in mind, a further amended grid helps to allocate the four content subjects with the various composition of these features. Accordingly, the categories of discourse of Mortimer and Scott, Bernstein, and Mentz, combined, form a grid on which the four subjects selected for this thesis can be placed. The manner of

communication and oral production forms one dimension whereas the kinds of knowledge, the quality of classroom discourse and vocabulary used form the other dimension:
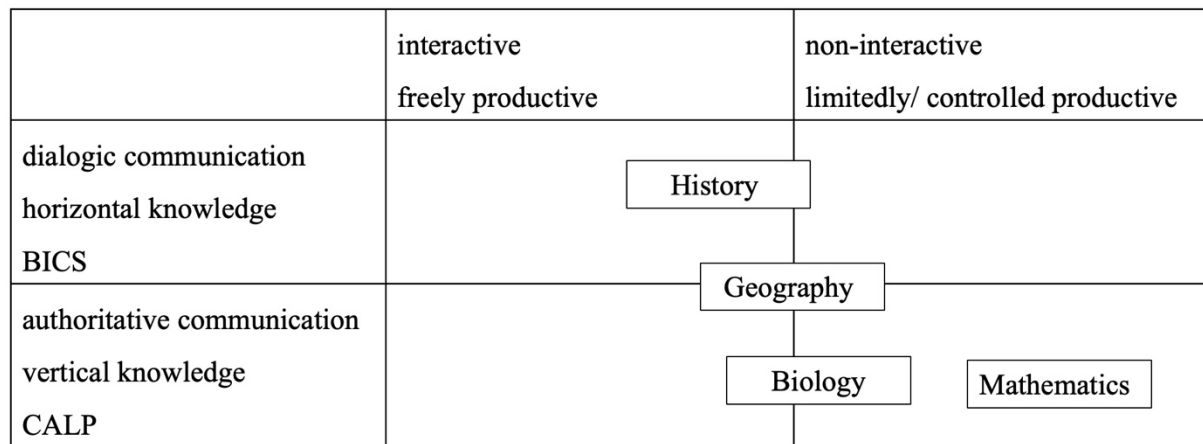
	interactive freely productive	non-interactive limitedly/ controlled productive
dialogic communication horizontal knowledge BICS	History	
	Geography	
authoritative communication vertical knowledge CALP	Biology	Mathematics

Figure 4.1 - The four subjects in focus allocated with the types of communication, kinds of knowledge and quality of language used (BICS, CALP)

For History, Geography, and Biology, another categorization is possible for discourse functions, reflecting the level of cognitive demand connected to each discourse function. This cognitive demand is subdivided into three categories called *Anforderungsbereiche* (AFB). The first Anforderungsbereich, AFB 1, focuses on the repetition of limited subject-related aspects and employs strictly reproduction of knowledge (cf. Kultusministerkonferenz 2005: 6). AFB 2 requires progressed skills of delivering explanations, ordered lists, and elaboration independently within the constraints of pre-known aspects (ibid.). AFB 3 requires students to reflect upon unfamiliar problems, methods, and results to arrive at independent conclusions, judgments, and interpretations of the subject matter at hand (ibid.). The progression through the three AFB describes a development similar to *Bloom's Taxonomy*, i.e., the hierarchical levels of learning objectives (cf. Coyle, Hood & Marsh 2010: 31). Each AFB subsumes a list of discourse functions that reflect the required activity and level of complexity, e.g., the discourse function DESCRIBE requires students to apply their prior knowledge and practiced methodology to "tell what someone or something is like" (Kröger & Lohmann 2007: half-title). In German didactic approaches for Mathematics, those AFB are neither defined nor applied.

Although speaking plays an essential role in classroom discourse, another productive skill is equally important to facilitate learning: writing. Although the role of writing in the learning context is largely unrecognized (cf. Llinares, Morton & Whittaker 2012: 244), it takes over an essential part in teaching as a means of support during teaching, the way to secure learning outcomes, and to allow students to practice newly learned content and structure their thoughts. Likewise, students learn to follow the requirements different genres impose on a written text as well as the requirements of the intended reader (cf. Llinares, Morton & Whittaker 2012: 268). Accordingly, students learn to place basic but indispensable information at the beginning of a text and then build their elaboration, arguments, and conclusion on that (ibid.). This manner of language production requires different choices from the linguistic system than speaking does. Written text is usually significantly denser than spoken language

featuring more lexical and nominal items (ibid.: 247). Through writing, among the abovementioned intentions of support and practice, students are given the opportunity to demonstrate their learning progress in both content and language use. Students should demonstrate their understanding and ability to apply concepts and terminology (ibid.: 257).

Writing in bilingual classes adds another layer to this skill as students can show their knowledge of grammatical and lexical aspects of the foreign language. Based on this idea of writing as a part of classroom discourse, the language used in writing tends to be more CALP-oriented and displays more vertical knowledge. Dependent on the content subject and task, however, writing can also entail larger shares of BICS-oriented language, e.g., in cases in which students are asked to elaborate on their personal point of view of a given topic in writing. Although writing displaying horizontal knowledge can occur, most in-classroom writing tends toward recording vertical knowledge in an authoritative manner (ibid.: 262). To summarize, "text production is both a means of learning the content subject and evidence that the content [and language have] been learnt" (Llinares & Whittaker 2010: 126). Both modes are equally important for classroom discourse (ibid.), although speaking takes the majority share in it.

Based on the elaborations above, the centrality and importance of discourse for learning are evident. Therefore, it appears almost self-explanatory that "students need opportunities to produce language" to facilitate learning which "stretches their current level of competences" (Llinares, Morton & Whittaker 2012: 53). If opportunities to produce language, and with that discourse, are of such importance for knowledge acquisition and student success, Wannagat's and Dalton-Puffer's findings that teachers speak most during classes seem to be juxtaposed (2010; 2003). A possible explanation could be that teachers want to ensure students do not fall behind on the content side because the language overly challenges them. This assumed intention leads to the question of which language requirements and proficiency levels are required by bilingual teaching and its content subjects.

4.3 Language requirements

As stated above, classroom language has to consider students' proficiency and ability (cf. Leisen 2015: 232). These two aspects cannot only be determined based on written assessment, especially not with the importance of spoken discourse. A teacher's activities encompass assessing and evaluating the two aspects to prepare content for bilingual teaching. But given the observation that teachers in bilingual settings are most often subject specialists rather than language or language teaching specialists and thus, have limited expertise in the intricacies of language learning and teaching (cf. Nikula 2010: 106), this bears the question of how teachers could judge as to whether the lesson planned matches student proficiency in the manner intended. Accordingly, teachers need knowledge and methodologies to measure language proficiency. The CEFR (see chapter 2.2) provides descriptions of proficiency on "different layers of specificity and abstractness" (Harsch 2017: 251), giving an overall perspective on proficiency, which is then broken down into increasingly specific descriptions of proficiency per skill

but also on vocabulary, grammar, and communicative strategies (ibid.). With this standardized scale at hand, providing teachers with information on every skill addressed in class, teachers could measure language proficiency and also evaluate their horizon of expectation towards students when planning lessons. This preparatory examination of language, however, requires teachers to thoroughly plan every single aspect of discourse, e.g., when planning on discussing a task in class, the teacher has to plan for possible student contributions. Moreover, teachers would have to determine a threshold at which student contributions are considered sufficient and thus, define the criteria for student success. The companion volumes to the CEFR are intended to provide teachers with the information needed to do just that (cf. Council of Europe 2020: 28). Each skill, subdivided into the six CEFR levels (A1 – C2), is described with the help of so-called *can-do-statements* giving teachers an idea of what can be expected of students at a certain level (ibid.). Accordingly, teachers can set their horizon of expectation in alignment with these can-do-statements and then have to take the proficiency and ability of their students into consideration to ensure the planned lesson is not overtaxing but "slightly beyond what the learner currently can handle" (Ortega 2009: 63).

Given the character of classroom discourse as spoken production, the language or level of language required for the productive aspect of tasks has to be considered individually by the teacher. Influential factors here are the horizon of expectations, the type of communication with its shares of interaction, dialogic or authoritative, the type of knowledge given, and with that the share of BICS and CALP (see Figure 4.1). Another approach to assessing the language requirements of spoken interaction in class is to place the featured discourse function onto the CLIL Matrix. In adapting Cummins' initial approach, Coyle et al. designed a two-dimensional grid featuring both cognitive and linguistic demand as axes. Those axes are subdivided into low and high demand, allowing all discourse functions on that grid to comprehend the requirements attached to each discourse function. The CLIL Matrix helps "ensuring that learners will be cognitively challenged yet linguistically supported" (Coyle, Hood & Marsh 2010: 43f.).

All those aspects point at extensive and detailed analysis (ibid.), which would or even should be part of lesson preparation. As this has to be considered as time-consuming and potentially overtaxing for teachers without explicit training in language teaching or even bilingual teaching, teachers consequently trust the ready-made material to be appropriate for the targeted grade as teaching material passes through approval processes. However, the question remains whether this material's face value meets the level of skill students master at the given time and, likewise, the level of skill targeted for the given grades.

4.4 Current status of research

In the light of the development of bilingual teaching, research on bilingual teaching and CLIL has become expansive (cf. Nikula 2010: 106). After the initial praise of bilingual programs (cf. Schlutow 2016: 6, Merino 2016: 17), even declaring it as the silver bullet of language teaching (cf. Wolff & Sudhoff 2015: 7), researchers started taking closer looks at different aspects connected to teaching and learning a content subject in a foreign language. Fields of interest focused on, i.a., the suitability of various subjects for the approach(es), possible influences on student motivation in connection to specific subjects, content learning and content-related competences, as well as specific skills and their development in select subjects taught in the foreign language.

Despite the general assessment of subjects of the natural sciences being less suitable to teaching in the foreign language (cf. Mentz 2010: 36), Mentz concluded that basically all subjects qualify for bilingual teaching. Furthermore, Viebrock also took this position, attesting suitability to Geography (cf. Viebrock 2007). Bonnet added to this focus of research by attesting suitability for Chemistry while looking at content learning in the bilingual setting (cf. Bonnet 2004). This aspect of research, the content learning in bilingual teaching, was a contentious issue especially among scholars in history sciences. While Hasberg, among others, voiced concerns that content learning would suffer due to teaching in a foreign language (i.a., cf. Hasberg 2004), Handro also detected difficulties in content learning. She stated that History classes are perceived as overtaxing in the traditional language of schooling already (cf. Handro 2013: 333). Students were struggling with putting their ideas and arguments into subject-appropriate language when discussing historical aspects (ibid.). Hasberg observed similar problems and pointed out that even strong students in History struggle with classroom discourse, especially with increasingly abstract topics. He even voiced concerns about whether discourse on such a level is possible in a foreign language at all (cf. Hasberg 2004: 230). On the contrary, however, studies by Müller-Schneck (2005) and Lamsfuß-Schenk (2008) were able to dispute those reservations and show at the same time that learning, especially with distinct aspects of History, such as *Fremdverstehen* (i.e., the understanding of the other), benefit from the higher degree of elaboration of the content presented in a foreign language. Golay came to a similar conclusion for Geography. His study showed that students in a bilingual track displayed comparable if not slightly advanced competences in the content subject although no additional lesson in the language of schooling was offered for bilingual students (cf. Golay 2005). For the natural science subjects, Bonnet investigated the chances for developing content subject competences through bilingual teaching in Chemistry, stating that although research is incomplete it appears that bilingual teaching yields comparable results in the science subject as teaching it in German (cf. Bonnet 2004). Likewise, Piesche concluded that bilingual teaching in science subjects, here Physics, does not increase the development of competences in the content subject, but discovered that bilingual teaching yielded increased motivation among students for the content subject (cf. Piesche 2016) and, with that, she addressed a prominent field of research within CLIL and bilingual teaching.

In bilingual strands, student motivation was perceived as being elevated, especially in the natural sciences. In her recent study, Rolletschek looked into bilingual teaching in Biology, primarily focusing on the motivation of girls for the science subject and their performance. She concluded that the motivation increased for the content subject but not for bilingual teaching as such (cf. Rolletschek 2021: 43). For Mathematics, Lipski-Buchholz could show that interest in language learning increased motivation for Mathematics taught in a foreign language. Additionally, she could not detect any problems with content learning, concluding that Mathematics is a suitable subject for bilingual teaching (cf. Lipski-Buchholz 2018). In his longitudinal, large-scale study, Rumlich also looked at motivation as a factor for student performance in bilingual tracks. He concluded that the better performance and higher motivation of students were not because of bilingual teaching but because those students passed through a selection process that targeted high-performing and motivated students upfront (cf. Rumlich 2016). Rumlich's longitudinal study looked at further aspects connected to bilingual teaching. With regard to student performance, including the development of language skills, he concluded that, again, the positive results other researchers arrived at were rather due to the selection process of students in bilingual tracks than the beneficial effect of teaching content subjects in a foreign language (ibid.: 434). Other studies, however, disregarding the aspect of the effect of selection, arrived at the conclusion that bilingual teaching is indeed beneficial for the development of skills in the foreign language. With a focus on a single subject and skill, Doll analyzed reading skills in bilingual History classes, in specific reading skills of subject-specific texts. The study was conducted at a Brazilian-German school. He realized that general reading skills cannot automatically be transferred to subject-specific reading, albeit bilingual reading skills were comparably advanced. He attested that bilingual schools have a high potential for content learning and for developing skills (cf. Doll 2002). The combined analysis of more than one skill was provided by Llinares and Whittaker in their study on oral and written production in Social Sciences (2007) focusing on Geography and History. They concluded that students, especially in the early stages of bilingual teaching, rather imitate language patterns whenever possible than producing language themselves. Students "hardly ever used" elaboration, a central part of developing content subject competence (Llinares & Whittaker 2007: 90). They further criticized that with content subject teachers, who are not explicitly trained for language teaching, language learning is expected to happen as a side product of teaching the content in the foreign language (cf. Llinares & Whittaker 2009). This presumption hints at problems, especially with oral production in those two subjects, since, according to Ortega, input alone does not facilitate language learning (cf. Ortega 2014: 268). Coetzee-Lachmann enhanced the scope of the researched skill by focusing on written discourse in bilingual Geography, detecting problems in the use of subject-specific register. Additionally, in the attempt to "contribute to the demystification of subject-specific discourse" (Coetzee-Lachmann 2007: 236), she broached the topic of expectations towards students' accomplishments in that skill and subject and stresses the importance of discourse for bilingual learning. Most prominently in the German context, the so-called *DESI* study surveyed the level of competences of students in grade 9 considering all four skills (reading,

listening, speaking, and writing), intercultural competence, and language awareness (cf. Klieme et al. 2006). According to the *DESI* group, students in bilingual tracks are one to two years ahead of their peers when it comes to language competences (ibid.: 60).

In summary, numerous fields in CLIL and bilingual education are researched. However, results, especially concerning student performance and motivation contradict one another. Likewise, disagreement on the suitability of selected subjects for bilingual or CLIL teaching seems to prevail. While studies, on the one hand, concluded that bilingual teaching is beneficial for both students' language proficiency and content competence, other studies discovered problematic aspects in the context of language use and vocabulary growth and, consequently, student performance. All in all, the picture generated by research on CLIL is not conclusive. A possible explanation for that may be, among other factors, that there is no standardized approach to teacher training, teaching, assessment, and curricula. Furthermore, there is no widely agreed-upon standard on student performance. In general, the attempt to describe a horizon of expectations regarding student proficiency in bilingual teaching is rarely exercised. No study analyzed more than one subject and more than one skill while focusing on prerequisites to successfully participate in classes taught in a foreign language (cf. Mentz 2015). Consequently, "although research […] has grown steadily during the recent years" (Nikula 2010: 106), the number of large-scale, systematic studies on language competences and proficiency of bilingual students is scarce (cf. Piske 2015: 110). A study by the *European Center of Modern Languages*, ECML on "language skills for successful language learning", however, resembles such a large(r)-scale study.

4.5   Study "Language skills for successful language learning"

The study by the European Center of Modern Languages (ECML) called *Language skills for successful subject learning – CEFR-linked descriptors for mathematics and history/civics* was carried out by a group of European linguists around Eli Moe from Bergen University. As "an Enlarged Partial Agreement" (European Center of Modern Languages 2022) of the Council of Europe, founded in 1994, the ECML focuses on language teaching and supporting member states to "implement language education policies" (ibid.). In total, "thirty-three countries currently subscribe" (Moe et al. 2015: inside cover) to the work and support of the ECML. The specific study's task was to identify minimal standards in the traditional language of schooling for migrant students "to do well" in Mathematics and History/Civics (ibid.: 19). The overall intention was to identify the required levels of language competence as outlined in the CEFR. Furthermore, the study should "raise awareness of the challenges that young language learners meet" (ibid.) when the language of instruction is a foreign language for which they hardly receive language support (ibid.). Further, the focus was on the quality of language, especially on the "academic use of language" in the content subject, which, according to Moe et al., differs greatly from the language used in language learning classes (ibid.). In the attempt to also link a selection of discourse functions to CEFR proficiency levels, supported by content subject experts from the subjects Mathematics and History (ibid.: 16), the study sought to answer which CEFR levels migrant

students in age groups 12/13 and 15/16 would need to master to *do well* (ibid.: 19). Successful participation or "doing well" was conditioned on the type of language students need to master, differentiated between academic and everyday situations (ibid.: 16).

Surprisingly, the result for 15/16-year-old students was in unison level B2 (cf. Moe et al. 2015: 39) for both subjects, while for the lower age group, the required level of language skills was determined to be at level B1. However, the authors conceded that in cases disagreement on a distinct CEFR level for a skill, the lower level was selected, although several cases accumulated a majority voting for a higher CEFR level (ibid.: 41). The study concludes that "language requirements are the same in History and Mathematics" (ibid.) for successful participation in the subject classes for migrant students. Likewise, all four skills require the same level of language proficiency.

To arrive at that rather unexpected result of the required levels for both subjects, the individual skills within the two subjects and a comparison of the results per subject to determine whether one subject may be more challenging than the other – language-wise (cf. Moe et al. 2015: 19), questionnaires on language descriptors covering all four skills with an equal number of items (ibid.: 29) were put together including the assessments of "31 teachers from 21 countries" (ibid.: 28). At the time, language descriptors were not yet standardized or published along with the CEFR in 2012/2013. Language and content subject experts, together with another group of CEFR-specialized researchers (ibid.: 26), worked on a revised, more specified version of those questionnaires (ibid.: 29). In a further step, language teachers of migrant students were involved. The expert group asked this group of experienced teachers to verify CEFR levels initially identified and indicate their idea of the required minimum level of language proficiency for their respective content subject (ibid.). In detail, the questionnaires were comprised of a selection of discourse functions, initially identified by the content subject experts, to which a CEFR level was allocated per skill. Lastly, teachers all over Europe were asked to both indicate whether the given discourse functions with their allocated CEFR levels were relevant for their individual content subject and to assign CEFR levels or much rather their descriptors to the individual skills addressed per discourse function (ibid.: 34).

When considering the obtained results, it has to be brought to attention that this study was initiated in 2012 when a volume of descriptors did not yet accompany the CEFR for the skills at the distinct levels (cf. Moe et al. 2015: 9). This approach to determine CEFR levels with the help of descriptors was most likely not as widely known as it may be today. However, the used can-do-statements were quite similar to the ones used today. When assessing the relevance of the results to the German educational landscape, it has to be pointed out that the study involved only one German CEFR expert and no German teacher (ibid.: 34, 38). Furthermore, the study seems to shift the benchmark. Initially, it talked about identifying requirements to "do well" (ibid.: 9) but then continued with "successful participation" (i.a., ibid.: 48). The study did not elaborate on either benchmark further than defining it as the language students need to master (ibid.: 16). The shifting and not clearly defined benchmark hampers the evaluation of the results and a possible transfer to the German context. Additionally, German studies

either hinted at or clearly stated that learning a content subject in a foreign language bears problems on both content learning and language usage (see chapter 4.4). As was discussed above (see chapter 4.4), History imposes challenges on students in the actual language of schooling (primarily German). Therefore, it appears somewhat unreproducible that Mathematics and History require the same level of language proficiency.

In addition to not involving German teachers, the study and the items worked with remained on a general level working with discourse functions and connected descriptors per skill. The specific teaching material was not incorporated, although both text and tasks play a significant role in teaching (see chapters 4.1 and 4.2). Moreover, the interdependences of language proficiency, tasks, and content performance (cf. Trenkic & Warmington 2019) cannot have been respected in this approach. Interpreting the results presented is difficult when neither "do well" nor "successful participation" are clearly defined, which may have been too complicated with all the different countries involved, all of which certainly have their own approach to educational success. Lastly, the group of teachers answering the questionnaire is unfortunately not described in more detail. It may be possible that those teachers were either language or content subject teachers, most likely not explicitly trained to teach a content subject in a foreign language. Although the afore-discussed ECML study has shortcomings, it points in the direction of research currently not covered in research on teaching in a foreign language, be it bilingual education or teaching migrant students.

In summary, the research so far has yielded debatable results in every aspect being researched. While Rumlich convincingly explained the reasons of the advanced performance of students in bilingual tracks, and thus put the initial praise of bilingual programs into perspective, results for content learning and competence, especially in History, are contradictory and inconclusive. Likewise, research results focusing on the different language skills are equally dispersed. Only research on the suitability of subjects for bilingual teaching seems to consistently arrive at the conclusion that basically all subjects are suitable. To date, no research has been conducted on either the language requirements each subject imposes on students or has attempted to identify a beneficial sequence of subjects to optimally foster language learning. This desideratum is addressed with thesis which will be further elaborated on in the following chapters.

5 Research questions and design

5.1 Research questions

The desideratum identified in the previous chapter, i.e., the absence of research with a comprehensive approach to determining language requirements in bilingual teaching and consequently, the deduction of a suitable sequence of subjects evokes four research questions:

I Is it possible to derive a sequence of subjects to be introduced in bilingual strands that reflects the level of proficiency students achieve to promote both subject and language learning as equal entities and thus be a pathway to CLIL?

To come up with a justified sequence of subjects which considers students' development of skills in the foreign language with the aim to avoid possible (negative) consequences for L2 acquisition and content learning, the identified desideratum has from different research angles. Language requirements consist of a combination of reading, listening, speaking, and writing and thus, each skill has to be looked at singularly. In identifying the required levels for each skill and imposed by the teaching material and comparing those findings to targeted levels, a sequence can be derived. The identified sequence then incorporates the analyzed subjects in an order that not only corresponds with the language competences targeted by the educational standards of language learning but also allows for true CLIL teaching, i.e., follow the above-described (see chapter 3.1) bidirectional approach of teaching language through content and vice versa. This sequence and outline could serve as base for future consideration when it comes to (re-) shaping existing bilingual strands or setting up a new strand, both aiming at CLIL, e.g., through highlighting and demonstrating the actual language learning potential in subjects which, to date, do not hold a prominent position in the canon of bilingual subjects, e.g., Mathematics.

Accordingly, the initial research question can only be sufficiently answered through further research questions focusing on finer grained aspects:

II What is the (average) level of language proficiency required by textbook material in bilingual teaching in Rhineland-Palatinate for the distinct grades?

With the centrality of texts as outlined in chapter 4.1, the level of complexity of texts is of interest with the aim of successful participation of students in mind, especially when the language used in those textbook deviates greatly from everyday spoken language (see chapter 4.2) and thus, increases the required level of language skill. This increase, caused not only by the difference between spoken and written language but also by the differences between the four chosen subjects, Mathematics, Biology, Geography, and History, decisively influences the required level of achievement in order to successfully participate in bilingual classes. Furthermore, various types of text within a single subject are of interest in this context, as, e.g., History, employs source text material which deviates considerably from the descriptive texts which can add another layer of complexity since those source texts can employ historical vocabulary. Research results concerning the textbook material in use for bilingual teaching

provide information on the complexity of such material but also allow for an assessment of the suitability of the material, especially with regard to the age group or grade the material is used with. In the context of the aspired sequence of subjects, findings on the complexity and connected language requirements give a first indication for an order among the subjects selected. The results obtained by the text analysis are one part of the overall evaluation of subjects and their language requirements. The other part is determined by answering the third research question.

III What is the (average) level of proficiency required by tasks and their discourse functions within the textbook material for the distinct grades?

Since textbooks not only come with information on the subject matter but also with tasks for students to carry out, those tasks need to be analyzed as well. This research question shifts the focus from the required receptive skill of reading to the required level of the productive skills of speaking and writing. The research connected to this question determines required levels of skill to successfully or much rather satisfyingly carry out the tasks from a language point of view. Acknowledging the interdependence of language proficiency, tasks, and content performance (see chapter 4), it is important that the required level of language skill is appropriate for the targeted age group. As tasks initiate discourse in the classroom, they are of equal importance as text input (see chapter 4.2). With language being used as tool while students are still developing their proficiency and the importance and dual-functionality of language in bilingual teaching (see chapter 4.2), it is important to respect the aspect that language is only "slightly beyond what the learner currently can handle" (Ortega 2009: 63). Again, the CEFR and its distinct levels of skill serve as a measure to identify the required level of receptive and productive language skills per task and, in aggregation, per subject. The findings on required levels for both reading comprehension as well as spoken and written production alone does not allow for an assessment of the suitability of the material for the targeted age groups. Likewise, a founded deduction of a sequence of subjects needs a comparison to the actual or expected, by means of curricula, level of language skills within the selected age groups. Therefore, another research question needs answering in the pursuit of identifying a sequence of subject to allow for a pathway to CLIL:

IV Does the required level of skill within bilingual teaching correspond with the levels targeted in the Rhineland-Palatinate curriculum for English? And if not, could bilingual students' proficiency still be sufficient at the given grades to successfully participate in bilingual classes of the selected subjects?

To answer this research question, the results obtained for RQ II and III have to be compared to the targeted level of skill per grade as set by the Standing Conference of Ministers of Education's educational standards of language learning (see chapter 2.2). The intention is to identify both single discourse functions as well as subjects which demand a level of language too high for the chosen grades. With the above discussed importance of speaking and writing (see chapter 4.2), this information is of

interest since it is debatable to what extent language acquisition can be realized if the required level of skill is too high. For the comparison of the identified CEFR levels required and those targeted, the findings of the DESI study that students in bilingual tracks display advanced language skills by one year compared to their peers (cf. Klieme et al. 2006: 60) has to be taken into consideration. Accordingly, the results of the quantitative text analysis and the qualitative analysis of tasks need to be combined to provide enhanced and amplified results and compared to the targeted levels of language as per the curriculum.

In summary, the results of the research conducted in connection to RQs II – IV combined provide an answer or much rather a proposal to answer RQ I.

5.2   Research design
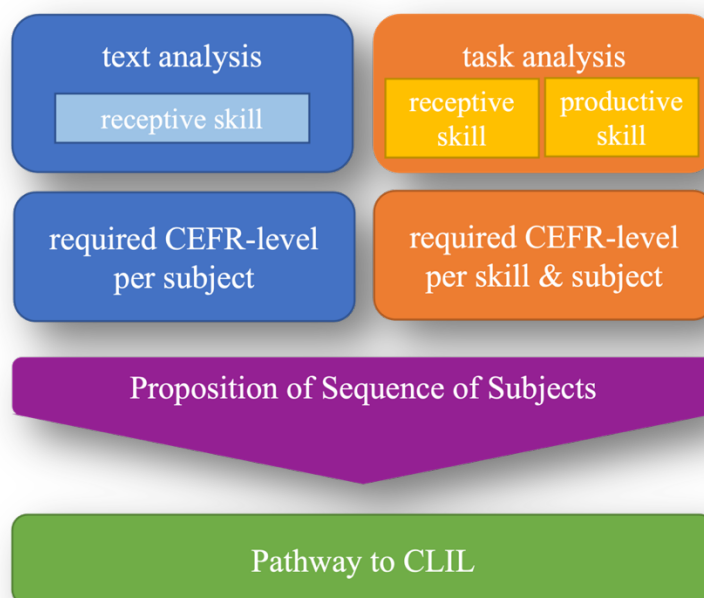
Figure 5.1 - Research Design Overview

The chosen research design as displayed above in figure 5.1 reflects the above-discussed research questions. The blue part addresses the information to be obtained to answer RQ II looking at the level of text complexity and thus, the required level of receptive skills in the four subjects and their various text types. The analysis of tasks and thus, the answer to RQ III is illustrated by the orange part. Both parts' findings and the comparison of the results to the set targets through the curricular standards (cf. KMK 2003) lead then to the proposition of a sequence of subjects, which in turn answers both RQ IV through the comparison and RQ I with the sequence, displayed here through the purple part. All answers and results combined contribute to the suggestions for a pathway to CLIL, represented here through the green part. Accordingly, the visual above displays not only the individual analyses necessary to answer the research questions but also the individual objects of investigation of each analysis to arrive at a pathway to CLIL. In summary, the research questions or, moreover, the information required to answer

those called for a tripartite research design. Additionally, each part consists again of several parts or steps of analysis.

As a first step, age groups or grades to be included into the analysis have to be chosen. Since parallels to the ECML study discussed in chapter 4.5 seem to be given, i.e., determining the required "language skills for successful subject learning" for migrant students aged 12-13 and 15-16 (Moe et al. 2015), this study adapts the age groups onto the context of bilingual education. Accordingly, the first age group in Germany oftentimes marks the end of what is called *Orientierungsstufe* (grades 5 and 6) (cf. Ministerium für Bildung, Wissenschaft, Weiterbildung und Kultur 2018: secs. 4, §18 SchulO RP). At the same time, bilingual preparatory courses end and bilingual education in selected content subjects starts in grade 7. The second age group aligns with grade 9, which for German schools marks the year before the *Sekundarstufe 1* ends, i.e., the equivalent of Level 2.44 to the *International Standard Classification of Education* (ISCED) defined by the UNESCO as "Lower secondary education. Sufficient for level completion, with direct access to upper secondary education" (UNESCO Institute for Statistics 2012: 37).

With the selection of subjects as described in chapter 3.1, adding Mathematics to the canon, this study enhances the range of subjects in comparison to the ECML study. This selection was based on a comprehensive list published by the federal state of Rhineland-Palatinate on schools offering a bilingual strand and research of the schools listed identifying subjects most frequently offered (see chapter 12.1). Accordingly, the four subjects, Mathematics, Biology, Geography, and History for the age groups of 12-13 and 15-16 and grades 7 and 9, respectively, form the focal point for the single parts of this study. In coherence with this selection, textbook material is analyzed (see chapter 6.2). The level of language complexity of textbook material for German grades 7 and 9 is of initial interest to answer the first research question. To obtain the required measure, the CEFR levels, a quantitative approach is chosen following the default a priori categorization (cf. Dörnyei 2007: 33). Since the textbook material has to be analyzed for various criteria (see chapters 4.1 and 6.3) and the results have to be assessed in accordance to the CEFR levels in comparison to the learning goals set by educational standards of language learning (see chapter 2.2), the help of automated tools for this analysis is indispensable. To determine the required CEFR level(s), a computer-based analytical tool which follows a quantitative approach, *Text Inspector* (Textinspector.com 2020), was selected. The desired information on text complexity as outlined in chapter 5.1, is obtained through the analysis of a selection of texts from the textbooks (see chapter 6.2.).

The third research question requires information obtained through the analysis of the tasks, precisely the receptive and productive skills (see Figure 5.1). Therefore, an analysis of valence and intensity (cf. Burwitz-Melzer & Steininger 2016), i.e., evaluating language or text along the lines of standardized scales as well as the analysis of (expected) learner language (cf. Marx & Mehlhorn 2016) is conducted. For these analyses, both receptive and, more intensely, expected productive skills are in focus to assess text along the lines of pre-set scales, in this case again the CEFR levels. For each task, a

tripartite analysis was conducted starting with reading followed by speaking and writing (see also chapter 8.2). To apply a systematic approach to this analytical step, a coding manual is created following a deductive approach (cf. Kuckartz 2012: 60), using the CEFR levels and their corresponding descriptors (cf. Council of Europe 2018) as categories and the so-called can-do-statements as distinct criteria (see chapter 12.7). Additionally, a horizon of expectations for the tasks and their individual discourse function is recorded through mapping all discourse functions per subject onto all feasible can-do-statements and their associated levels (see chapter 8.2.1 and chapter 12.6) allocating text to standardized categories in accordance to Mayring (cf. 2008: 93). Furthermore, this approach resembles a process of "understand[ing] the situations through the eyes of the participants" (Cohen, Manion & Morrison 2011: 32), i.e., what students need to be capable of when it comes to language production, which is a helpful step in preparation for the actual analysis of the tasks. The volume of tasks and their discourse functions requires an approach to assess the results not only on the individual level but also on an aggregated level per either discourse function and subject, or single skill in combination with a discourse function (see chapters 6.4.2 and 8.2.3). Averages were calculated for the individual skills to arrive at a CEFR level per grade. Likewise, maximum relative frequencies were calculated for discourse functions to be compared in an attempt to determine as to whether a distinct discourse function is equally demanding in the four selected subjects (see chapter 8.2.3).

While the analyses connected to RQ II and III determined the required language proficiency for both the textbook material and the tasks, the fourth research question aimed at a comparison of the yielded results and the levels of proficiency set as target (see chapter 2.2). Expected learner language was accordingly evaluated against the background of minimally required learner proficiency (cf. Marx & Mehlhorn 2016: 301). In order to identify possible discrepancies of the level of language required by the textbooks and their tasks in contrast to the targeted language levels as outlined in the curricula for the respective age groups, a gap analysis compares of two sets of data: the results from the analyses on the one hand and the targeted level which were to be compared to one another with the help of triangulation (cf. Schramm & Knorr 2016). The findings on the advancement of bilingual students (see chapter 4.4) have to be respected since students in bilingual tracks were found to be one year ahead of their peers attending non-bilingual tracks or classes (cf. DIPF & Leibniz-Institut für Bildungsforschung und Bildungsinformation 2020: 59). As a result of the comparison, the research questions can be answered for both the required proficiency in comparison to the targets set by the curricular standards and – as a consequence of the former – the likelihood of successful student participation in each of the selected subjects.

As a final step within this tripartite research design, i.e., the analysis of text, tasks, and the comparison to the targeted language levels, the first research question can be answered as a deduction from the results of RQ II, III and IV. Based on the analysis of both textbook material and tasks and the evaluation against targeted proficiency levels, a sequence of bilingual subjects can be identified which respects the development in language proficiency among students. At the same time, this sequence opens

a path to develop bilingual teaching onto another level: CLIL teaching. Respecting students' language proficiency and its development by this sequence, teaching the selected content subjects in a foreign language can also be used to teach the language with the help of the content subject. This way, teaching in a foreign language can progress from bilingual teaching to actual CLIL teaching applying the dual focus (see chapter 3.1).

5.3 Research validity – task analysis

The study on the linguistic requirements imposed by the tasks of the textbook material is neither strictly qualitative nor clearly cut quantitative. Much rather, this study follows a mixed-methods approach working with expectations to be evaluated. Still, those evaluations and assessments have to be reliable to the most possible extent. Consequently, this study employs validity measures that accommodate this mixed-methods approach. An initial question to the research's validity is the sample's adequacy (cf. Cohen, Manion & Morrison 2011: 198), which in this study's case should be answered satisfyingly based on the number of tasks analyzed. The selected textbook material for the first part of this study also provides the tasks for this part (see chapter 8.2). Furthermore, the described methodologies (see chapters 5.2 and 8.2) to yield the desired information, for both the preparatory steps and the evaluation, proved to be appropriate since the research questions could be answered based on the data created. Following VanderStoep and Johnston's idea of consensual coding in a Consensual Qualitative Research system (cf. 2009: 189), the study employs a team of researchers, experts for the selected subjects, to conduct the assessment along the lines of the coding manual based on standardized scales (see both chapter 12.7 and fold-out page at the end of this study). Cases of discrepancies between the initial assessment and the controlling iteration of the assessment are discussed individually, and adjustments are made if required.

6 Language proficiency requirements – textbook material

Textbooks are a central aspect in teaching in Germany (see chapter 4.1), which also holds true for Rhineland-Palatinate. Research on textbook complexity with a linguistic focus is a rare find, especially so in the bilingual context, for which textbooks are only available from a limited choice for the abovementioned prototypical subjects (see chapter 3.2). While today textbooks for English as a foreign language (EFL) are explicitly designed for the achievement of the CEFR level targeted in the curricular standards, textbooks for bilingual or CLIL teaching do not seem to be consciously
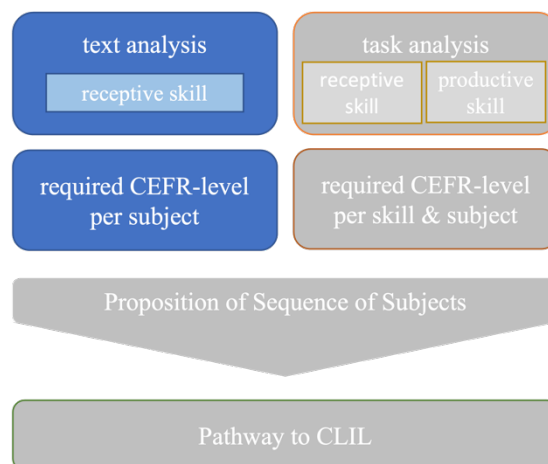
Figure 6.1 - Design Text Analysis

aligned with this benchmark. As "there is growing evidence that vocabulary size measures correlate well with overall measures of language ability" (Milton & Alexiou 2009: 196) and German curricula also give reference on the amount of vocabulary targeted in each school year (cf. Ministerium für Bildung, Wissenschaft und Weiterbildung 2000), the bilingual textbooks should follow suit and align their level of complexity.

6.1 Design

The selection of text material for this study was based on those textbooks most commonly used in Rhineland-Palatinate (see chapter 12.2). The linguistic requirements and their corresponding CEFR level had to be analyzed as described in chapter 5. As this analysis focused on the actual text, only the receptive skill of reading was addressed. The texts were analyzed applying various methods and approaches as described below. With teachers not specifically trained in bilingual or even CLIL teaching (cf. Fein 2021), textbooks serve as a (didactic) guideline, source of material, and supporting aspect in bilingual teaching (cf. Sudhoff 2015: 270). It is then all the more important, that textbooks' linguistic demand is appropriate for students' capabilities, and that, should this not be the case, teachers are aware of the gap between their students' language skills and the textbook requirements to ensure successful participation for students. Consequently, material was selected for each subject. Per subject, two chapters were picked per age group, i.e., two chapters from the textbook aimed at grade 7 and two chapters from a textbook aimed at grade 9. As outlined in chapter 5.2, grade 7 also marks the onset of bilingual teaching in Germany. Additionally, to be able to trace the progress students make throughout the schoolyear, the two chapters were selected in a way to correspond to the supposed beginning and end of a schoolyear. This selection reflects the progress in case it is incorporated in the level of language complexity of the textbooks. The selected chapters of each textbook are subdivided into subchapters. The analysis followed this preset structure and subdivision, which not only made information and data handling less complex but also helped with the application of the automated analysis since Text

Inspector limits texts to 10,000 words. Accordingly, the presentation and discussion of the selected key figures for the criteria of analyses (see chapter 7) are given on a subchapter level. Additionally, in order to arrive at a comprehensive assessment of the texts analysis, Text Inspector's overall CEFR-value across all criteria is added to the results. Text Inspector was employed to analyze the above-described corpus of material assessing the text based on a selection of criteria (see chapter 6.3). To ensure the validity and reliability of the results yielded from using an automated application, each text sample was processed three times. Deviations in the results were investigated and clarified, e.g., differences in sentence length numbers occurred which were based on punctuation not being interpreted correctly by the application. In such cases, the affected text was manually corrected instead of copy-and-paste addition. Following the text analysis as such, numerical values were assigned to the identified CEFR levels to calculate average values within the categories across the subchapters, which in turn were used to detect an overall CEFR level for the chapter as such. For example, the findings for lexical sophistication for each subchapter were averaged to come to an overall assessment of this key figure for the entire chapter. Together with the overall assessments by Text Inspector, these averages allowed for an assessment of the complete textbook material per subject and grade.

6.2 Material selection for each subject

For each subject to be analyzed, textbooks and respective chapters were selected. To identify appropriate textbooks, a preliminary survey was conducted looking at school's websites and their book lists to identify those textbooks used most frequently (see chapter 3.2). This data was available for a total of 28 schools offering bilingual teaching in Rhineland-Palatinate (see chapter 12.2). The identified textbooks served as a source to compile the corpus for this study (see chapters 6.2.1- 6.2.5). The selected texts conform to the above-described characterization of texts per subject (see chapter 4), e.g., the difference in History between descriptive texts and historical source texts; or the difference between their intention, describing either procedures, i.e., giving instructions on what to do and how to do it, or reports to structure or organize content knowledge (cf. Llinares, Morton & Whittaker 2012: 112).

6.2.1 History

For bilingual History teaching, three comprehensive textbooks and various module booklets are available. The textbooks, covering the full range of curricular topics in History, are *Invitation to History*, Volume 1 & 2 by Cornelsen publishing (cf. Weeke 2011, 2011a), *Exploring History for Bilingual Classes 1 & 2* by Westermann Publishing (cf. Kröger & Lohmann 2007; Kröger 2009), and *Connect* by Schöningh publishing. *Connect*, however is rarely used in Rhineland-Palatinate (see chapter 12.2) and hence, the selection process focuses on the remaining two. *Exploring History for Bilingual Classes* by Westermann publishing is now available as a second edition published in 2018 but, to date, rarely used. Cornelsen publishing offers *Invitation to History* first released in 2016. Both textbooks follow a chronological approach to History teaching, offering both subject-related skill practice as well as

scaffolding. The selection of topics covered revealed that *Invitation to History* presents a composition of rather atypical topics when it comes to History teaching in Germany. In addition to the French Revolution, classically featured in History textbooks, *Invitation to History* covers also the American Revolution (cf. Weeke 2011: 6-23). Likewise, this textbook contains a chapter on Great Britain in the context of industrialization (cf. ibid.: 24-31). Usually, textbooks in Germany discuss mainly the German history and developments and only briefly broaden the topic to Great Britain. According to the Rhineland-Palatinate curriculum for History, teachers can choose topics freely within the focal points of the given epoch (cf. Ministerium für Bildung, Wissenschaft, Weiterbildung und Kultur 2021: 88ff), so discussing the American Revolution instead of the French Revolution is basically possible. The choice of topics, however, puts *Invitation to History* at a disadvantage because the developments in Europe are not dealt with. Those developments, however, are an important aspect to fulfill the mandate of History teaching (see chapter 2.1), i.a., the education of "responsible and active citizens" to preserve "historical and cultural heritage in Europe" (Beacco 2007: 3), and, thus, should not be left out. Not only the better fit to the curriculum but also the fact that *Exploring History* has been on the market for a longer time may be the reason why this book is more often used in schools than *Invitation to History*. As schools work mostly with the first edition of *Exploring History* it was also selected to provide the texts to be analyzed.

The selection of topics to be analyzed was based on the current curriculum for History published in 2021. The topics to be analyzed are chosen in slight deviation from the age groups preset by the ECML study. While the ECML study looked at the age groups of students between 12 and 13 years as well as 15 and 16 years of age, especially the first group had to be adjusted. Since bilingual teaching of History only starts in year 8, students are usually between 13 and 14 years old. Hence, the topic "Absolutism and Enlightenment" (cf. Kröger & Lohmann 2007: 6f.) was chosen as it is taught in the beginning of the year (Ministerium für Bildung, Wissenschaft, Weiterbildung und Kultur 2021: 82). According to the targeted CEFR level as described above (see chapter 2.2), students should be at level A2, which is targeted at the end of year 7. As described in chapter 5.2, the selection of chapters respected the assumed development of language skills throughout a school term which is why the second topic to be analyzed was "Industrialization" (cf. Kröger & Lohmann 2007: 54f.), taught towards the end of the schoolyear. Accordingly, the following topics for the second age group are "The Cold War and German Division" (cf. Kröger 2009: 72f.) and "Superpower Rivalry" (ibid.: 128f.). Both topics are covered towards the end of grade 9 or early grade 10 at which time students usually are between 15 and 16 years old.

6.2.2 Geography

For Geography, two comprehensive textbooks along with a large selection of thematic modules are on offer. One is *Around the World* available in two volumes by Cornelsen publishing (cf. ibid.: 2014, 2014a). Just like their History textbook, *Around the World* offers scaffolding for material, methodological practice for study skills as well as language support in form of chunks, wording suggestions and examples. The book covers topics outlined in the curriculum for grades 6 and 7 (cf. Ministerium für Bildung, Wissenschaft, Weiterbildung und Kultur 2021: 22) while volume 2 covers topics for grades 8 and 9. It is, however, a little difficult to clearly map the topics in *Around the World* to the current curriculum since this not only addresses content knowledge but also competence-oriented work such as spatial orientation, methodological competences, and, i.a., communicative competence. This structure is not reflected in *Around the World*. Content and competences are reflected in the other comprehensive textbook, *Diercke Geography*, published in 2013 and 2014 along with or after the revision of the Rhineland-Palatinate curriculum. This may be the reason why 16 schools offering Geography in their bilingual track prefer this textbook.

For this study, chapters on "Living on a dynamic earth" (Löbmann 2014: 8f.) and "Living in Different Climate Zones" (ibid.: 31f.) were selected, again respecting the assumed language development in students throughout the course of the school year. Those topics have to be covered in *Lernfeld II* according to the curriculum (cf. Ministerium für Bildung, Wissenschaft, Weiterbildung und Kultur 2021: 57f.), which is subject of teaching in grades 7 and 8. To represent topics to be covered in grade 9, chapters on "Global Disparities" (Ahrend 2013: 24f.) and "Europe Changes" (ibid.: 104f.) located with *Lernfeld III*, taken from Volume 2, were added to the corpus.

6.2.3 Biology

Bilingual teaching in Biology is not as established as History or Geography. Out of the 28 schools in Rhineland-Palatinate, only five offer Biology as a subject throughout the full school year (see chapter 12.1). This is reflected in the textbook material available. Currently, a single comprehensive textbook for Biology is in use in Rhineland-Palatinate while again a larger number of topic module brochures is offered by publishers. Hence, this study focused on the comprehensive textbook *Discover Biology Volume 1* (cf. Mathews & Olmesdahl 2010). This textbook covers all topics of Biology teaching throughout grades 7 to 10 (cf. Ministerium für Bildung, Wissenschaft, Weiterbildung und Kultur 2014a). Contrary to History and Geography, the Biology curriculum does not allocate certain topics to a grade in which those topics should be covered. It is hence the schools' decision when to introduce which topic. This flexibility may be beneficial for schools but posed as a problem for this study since the effort to select topics with regard to the language development is difficult to realize for Biology. To make a suitable selection, a closer look at the textbook's structure is necessary. *Discover Biology* is subdivided into three main units. The first unit covers "Human Biology", the second focuses on "Flowering Plants" while the third unit is on "Cells, Tissue and Organs" (cf. Mathews & Olmesdahl 2010). With the idea

of revealing a development in language skills in mind, topics are chosen from unit 1 on Human Biology since remaining in this context may show that later topics refer to and build upon earlier topics. Hence, the chapters on "Respiration" (cf. Mathews & Olmesdahl 2010: 30f.) and "Blood and Blood Circulation" (ibid.: 44f.) are added to the corpus representing topics that could be addressed in grade 7. Likewise, "Food and Nutrition" (ibid.: 58f.) as well as "Digestion and Absorption" (ibid.: 75f.) were selected for grade 9. Topics for both grades were selected after reconciling with local Biology teachers. All selected chapters feature subchapters on subject-related skills, e.g., "Working with models" (ibid.: 36), activity suggestions, e.g., "Measuring changes in your pulse rate" (ibid.: 54), and revision parts.

6.2.4   Mathematics

Mathematics as a CLIL-subject is not yet supported by German publishers and hence, a different approach to material selection had to be taken. To single out topics for which English material had to be collected, the curriculum for Mathematics was consulted (cf. Ministerium für Bildung, Wissenschaft, Weiterbildung und Kultur 2007). As was the case with History, topics from the beginning and end of grades 7 and 9 were selected: Algebra 1 covering basic competences on data, graphs, tables, and mathematical terms (ibid.: 55–62) and Geometry covering angles, bisectors, vertices, and medians (ibid.: 49–54) for grade 7. For grade 9, the continuance of the aforementioned topics was selected: Algebra 2 covering simple or linear equations and inequations, linear optimization, and the procedures of substitution and accumulation (ibid.: 77–78) while Geometry in grade 9 deals with square shapes, the Pythagoras theorem, basic algebraic calculations, etc. (ibid.: 74–76). To support and confirm the selection, the German textbook *Mathematik – Neue Wege* by Schroedel Publishing (cf. Körner 2016, 2016a) served as a sample to reconcile. This particular textbook is used by most of those 28 schools offering bilingual strands (see chapter 12.2). Having selected the topics on Algebra 1 and 2 as well as Geometry, those topics were matched with native-English material based on the technical terms used. Initially, curricula in the US (cf. Dossey, Halvorsen & McCrone 2012), the UK (cf. Department for Education 2013a), and Australia (cf. NSW Standards Authority 2019) were consulted to ensure that topics selected are taught in those countries as well. For accessibility reasons, British mathematics textbooks addressing the corresponding topics were selected. In total, samples from seven different text- and exercise books were compiled to add to the corpus featuring three volumes of *Elevate Mathematics 11-14* (cf. Baker 2008; Baker & Hogan 2008; Baker & Hogan 2008a), *Framework Maths 7c* (cf. Capewell 2002), *CGP Mathematics for Key Stage 3 Book Three* (cf. CGP 2014), *GSCE Mathematics Workbook Higher Level* (cf. Parsons 2000), and *Edexcel - GCSE Mathematics - Higher Course* (cf. Casson 2001). All textbooks selected support students to achieve the General Certificate of Secondary Education, the UK graduation equivalent to the German "Realschule" or Rhineland-Palatinate's "Realschule Plus" after 10 years of education.

The three volumes of *Elements* are designed to cater to differentiated levels of difficulty for the same mathematical topics of key stage three. In the UK curriculum, key stages describe the approach to

organize "blocks of years" (Department for Education 2013b) with key stage three comprising grades 7 to 9. Each volume of *Elevate Mathematics* is aimed at one grade within that key stage. *Frameworks Maths 7c* also targets key stage three, here mainly year 7, and is designed to be a course for basic knowledge. Book three of *CGP Mathematics for Key Stage Three* features material representative for the level of difficulty applied in year 9, i.e., the end of key stage three. The *GCSE Mathematics Workbook Higher Level* compliments aforementioned textbook with further practice exercises. Lastly, *Edexcel* is a comprehensive textbook to prepare for the GCSE examination at the end of year 10 or key stage four, respectively. The topics selected for grade 7, Algebra 1 and Basic Geometry, were covered by exercises and tasks taken from *Frameworks Maths 7c*, *Elevate Mathematics 11-14*, precisely from the volumes for levels 3 to 4 and level 5 to 6 as well as from *CGP Mathematics for Key Stage 3, Book Three*, and *GSCE Mathematics Workbook Higher Level*. For grade 9, the topics within Algebra 2 and advanced Geometry, were also covered in *Elevate Mathematics* for level 4 to 5, *CGP Mathematics for Key Stage 3, Book Three*, and the *Edexcel* preparatory course book. It has to be pointed out that the quality of texts analyzed for Mathematics differs remarkably from the texts in the other subjects. The History, Biology, and Geography books feature longer, coherent informative texts on the topic at hand. In Mathematics, however, no such informative texts are given. Moreover, the text analyzed was the sum of exercises and tasks and their texts. An overview of the topic mapping is attached to this thesis in the chapter 12.4.

6.2.5   English

Since selected key figures of the analysis done by *Text Inspector* refer to scales created for the native-speaking context, such as the Flesch-Reading-Ease index (see 6.3.3), it was considered as both appropriate and essential to also analyze texts from English classroom material to adjust the scale to German learners. Therefore, *English G21* by Cornelsen publishing (cf. ibid.: 2010, 2010a, 2011, 2011a, 2011b) was analyzed using the *Text Inspector*. This textbook is widely used in Rhineland-Palatinate. *English G21* is a comprehensive set of textbooks for English teaching. It is used throughout grades 5 to 10, covering thematic aspects connected to Great Britain, the US, and Australia and in parallel, these books introduce grammatical phenomena and vocabulary along the lines of the thematic contexts. This way, the six volumes aim at developing students' language proficiency towards the targeted CEFR levels B1/B2 (cf. Ministerium für Bildung, Wissenschaft, Weiterbildung und Kultur 2014b: sec. 5). Each book is subdivided into a number of units, which convey selected topics and grammatical aspects. Each unit features an introductory part, vocabulary and grammatical work, both exercises and tasks, as well as a main text. For grades 6 to 10, the main texts of two units, one from the beginning and from the end of the book, were analyzed to determine a reference for the scales mentioned above. The textbook for grade 5 was excluded since preparatory courses for bilingual strands in Rhineland-Palatinate only start in grade 6, hence only material from grades 6 and higher grades was required for a comparative analysis.

6.3   Methodology

"Quantifying vocabulary richness gives the sense of a quest for the Holy Grail […]"

(Malvern et al. 2004: 3)

The initial discussion of key figures to determine text complexity, such as readability scores, lexical sophistication and diversity, or mere quantitative measures (see chapter 4.1), showed that analyses of texts need the support of automated tools. Although the sources of the information on the above-described key figures are available online, such as the *English Vocabulary Profile* and *the English Grammar Profile* by Cambridge University (cf. Cambridge UP: 2015; 2021), or the corpora for frequencies for both British and American English, manually analyzing text and calculating key figures should not be done. Today, the options of automated analysis of text are given and considered far more reliable and precise (cf. Green, Khalifa & Weir 2013: 25). One such automated tool already mentioned above, working on the word and sentence level, is the so-called *Text Inspector*, which not only calculates key figures but also assigns a CEFR level to each key figure (cf. Textinspector.com 2020). With the support of *Text Inspector*, the analysis of the selected material was conducted. This online analysis tool was developed by Prof. Stephen Bax (1960 – 2017)[5], who focused on "the use of technology for language learning" (Textinspector.com 2020). He initially developed this tool to provide measurements on readability, complexity, lexical diversity, and CEFR level (cf. Textinspector.com 2020). *Text Inspector* analyses texts uploaded with regard to various aspects, such as Type-Token-Ratio, diversity values, corpus-based evaluation for BNC[6], COCA[7], EVP[8] and AWL[9]. To come up with a reliable evaluation of the text uploaded, *Text Inspector* also makes use of syllable, word, and sentence counts and calculates various readability scores such as Flesch-Reading-Ease, Flesch-Kincaid Grade, and the Gunning-Fog Index, which all align the calculated score with corresponding experience levels of native readers. Not only does *Text Inspector* examine texts with regards to the skill, e.g., a written text, it also aligns the results of key figures with the CEFR levels. For example, the tool compares the text to the renowned English Vocabulary Profile provided by Cambridge University Press. It identifies "the level of each word, phrase, idiom and collocation found in the text according to the CEFR on a scale from A1 – C2 […]" (Textinspector.com 2020). The fact that *Text Inspector* is transparent with the applied algorithms and the alignment of the computed results with the CEFR added to the desired level of validity of the analyses. Furthermore, the fact that the British Council uses *Text Inspector* as an instrument of assessment (cf. Owen, Shrestha & Bax 2021: 2) adds to the status of this automated tool as being both reliable and used in standardized testing.

For the purpose of this study, a selection of key figures contributing to the determination of the level of linguistic complexity was made. The figures which contributed to an assessment of the textbook

---

[5] https://www.timeshighereducation.com/people/obituary-stephen-bax-1960-2017
[6] British National Corpus
[7] Corpus of American English
[8] English Vocabulary Profile, developed in cooperation with Cambridge University Press and Cambridge English Language Assessment
[9] Academic Word List, based on Averil Coxhead's MA thesis at the School of Linguistics and Applied Language Studies, Wellington, New Zealand

material are lexical sophistication, lexical diversity, lexical or propositional density, the Flesch-Reading-Ease index, and the CEFR levels assigned to distinct grammatical features. That way, assessments on the word level, the sentence level, on quantitative aspects, as well as the text as a whole was ensured.

6.3.1 Lexical sophistication

Lexical Sophistication is the "appropriate choice of low frequency vocabulary items that include "the use of uncommon words and […] uncommon words that allow writers to express their meanings in a precise and sophisticated manner" (Malvern et al. 2004). Lexical sophistication gives information on the volume of a learner's active vocabulary and the ability to use it (cf. Malvern & Richards 2002: 85). To analyze the lexical sophistication of a text, several sources to compare the choice of words of a text can provide information for comparison. Accordingly, *Text Inspector* uses, i.a., the English Vocabulary Profile to compare the words used in, e.g., a text, with the collection provided by the English Vocabulary Profile (EVP). The collection is based on "several hundred thousand examination scripts written by learners from all over the world that is added to every year" (Cambridge UP 2021). For the assessment part, *Text Inspector* uses the lowest given rating per word provided by EVP (Textinspector.com 2020), which as such respects that words and their various meanings may have different levels of complexity within the CEFR. EVP complements its data set with information from the Cambridge English Corpus to determine "what students can or cannot do at each [CEFR] level" (ibid.). "Unlisted" words are those words which are not listed and rated within the EVP and most often refer to individual or place names, non-English words, or numericals.
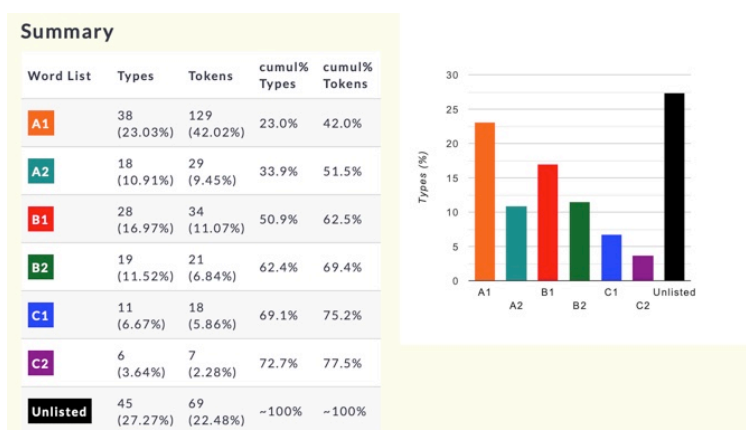
Figure 6.2 - Screenshot from Text Inspector on a sample text assessment for lexical sophistication based on EVP

According to *Text Inspector* experts, this group, however, does not affect the overall CEFR-value[10]. When it comes to deriving an overall CEFR-rating from the comparison to EVP, the A-band has a different effect than the B- and C-Band: while higher percentage values in the A-band imply a lower text complexity, they imply a higher text complexity in the B- and C-band. Calculating the CEFR level for the text at hand, *Text Inspector* takes figures from all three into account. Precisely, for the EVP assessment, the text is analyzed for both types and tokens and their percentage share with regards to the CEFR level (see Figure 6.2 above). From this allocation, *Text Inspector* calculates an average value to assign a CEFR level for EVP-based lexical sophistication. Lastly, *Text Inspector*

---

[10] The author of this study inquired about various questions concerning the algorithms of *Text Inspector* to both better understand the processes involved and to clarify such issues. Unlisted words, e.g., were not incorporated into the calculation of CEFR-levels since they are not assessable through the references such as EVP (Michael Bax, personal communication, January 20, 2022).

aggregates the single results of each analysis to calculate an overall assessment for the analyzed texts (cf. Textinspector.com 2020).

In addition to the comparison of text to the EVP, an analysis with regards to frequency of the words used was done with the help of *Text Inspector*. The frequencies of words, or the rarity of words, provided further information on the sophistication of vocabulary. Therefore, large corpora, such as the British National Corpus (BNC) and the Corpus of Contemporary American English (COCA) can be taken into consideration to calculate CEFR levels for each word based on its frequency. As schools in Rhineland-Palatinate mostly work with textbooks and material based on British English, excluding the English textbook in year 8 which is about America, this analysis here focused on the frequencies and their assigned levels derived from the British National Corpus. For the BNC, *Text Inspector* assesses the texts on a word basis looking at "how commonly used they are" (ibid.). For each word used in the texts to be analyzed, *Text Inspector* determines the lexical frequency (Lexical Frequency Counts (LFC)) and retrieves the rank of each word in accordance (ibid.). The span of ranks ascertained serves as the basis to determine a median position, e.g., if the first and lowest rank of a word within the text was 400 while the highest rank was 800, the median position would be 600. It has to be mentioned that the lowest rank indicates a high frequency relative to the highest rank, which indicates a lower frequency (cf. BNC Consortium 2007). Around that median position, a word which ranks closest to that position in terms of frequency is then identified as the median word. This median word is then analyzed for its CEFR level with the help of the information given in the *English Vocabulary Profile* (EVP). Picking up the example from above, the word closest to the median position of 600 is the adverb "expressly" (Davies 2004) at position 598. Within the EVP, this word is rated at CEFR level B2. Consequently, a high position, implying a less frequent use, within the corpus results in a higher CEFR level for lexical sophistication whereas a lower rank or position within the BNC, implying a more frequent use, results in a lower CEFR level. To further analyze the text for its linguistic demand or required CEFR level, *Text Inspector* helped to calculate the percentage of words above the determined median word and its assigned CEFR level. With relatively high percentages above the assigned level, it is rather likely that the lexical diversity applied in this specific text requires higher levels of language competence, e.g., if the assigned level is A2+ with a total of 55% below or equal to that level, the remaining percentage of types indicate that there is a considerable share of more demanding types students have to work with (a smaller percentage may be of unlisted or unknown words). It has to be noted, though, that not every word allocated to a frequency rank in BNC is rated in EVP. Words like "raspberry", "gut", "resulting" or "others", e.g., are allocated with a rank within the BNC but do not appear in EVP. If *Text Inspector* comes across any of these words, it marks it as "unlisted" referring to the non-existent entry in EVP.

Another key figure to be calculated through *Text Inspector* bases on the Academic Word Lists compiled by Averil Coxhead in 2000. Again, types and tokens are reconciled with those lists and the frequency of occurrences is counted (see Figure 6.3) (cf. Textinspector.com 2020). Finally, the sum of all words detected both in the text and on one of the lists is calculated to arrive at a CEFR level. The higher the total amount or the percentage of the text respectively, the higher the CEFR level, i.e., for the sample text used as basis for Figure 6.3, the percentage of all types used in the text to also be found on the AWL or phrases lists sums up to 11,71% (see first line, "AWL total"). This value translates to CEFR level C2+ in accordance to *Text Inspector*. It is important to understand that vocabulary size alone does not manifest a CEFR level an EFL learner achieves. Scores related to a learner's lexis moreover indicate

Word List	Types	Tokens	cumul% Types	cumul% Tokens
AWL Total	39 (11.71%)	83 (10.06%)		
AWL 1	12 (3.60%)	22 (2.67%)	3.6%	2.7%
AWL 2	10 (3.00%)	29 (3.52%)	6.6%	6.2%
AWL 3	3 (0.90%)	12 (1.45%)	7.5%	7.6%
AWL 4	1 (0.30%)	2 (0.24%)	7.8%	7.9%
AWL 5	4 (1.20%)	8 (0.97%)	9.0%	8.8%
AWL 6	1 (0.30%)	1 (0.12%)	9.3%	9.0%
AWL 7	4 (1.20%)	5 (0.61%)	10.5%	9.6%
AWL 8	4 (1.20%)	4 (0.48%)	11.7%	10.1%

Figure 6.3 - Screenshot from Text Inspector on a sample text assessment for lexical sophistication based on AWL

"[…] knowledge and attainment rather than [being an] absolute determiner" (Milton & Alexiou 2009: 210).

6.3.2 Lexical diversity measures

Lexical diversity indicates the grade of variation of vocabulary used in a text (cf. Johansson 2008: 62). Classically, this is expressed by a measure called Type-Token-Ratio (TTR), which is also calculated by *Text Inspector*. However, TTR was not used in this study since the numbers this figure provides may be misleading as "text samples containing large numbers of tokens give lower values for TTR and vice versa" (ibid.). Moreover, text or rather their TTR values are only really comparable when the samples analyzed are of the same length (ibid.). It is also rather problematic especially with assessing textbook material and their linguistic demand that "although the number of word tokens increases substantially with the speaker's […] age increases, the TTR drops" (ibid.: 63). Accordingly, lexical diversity measurement should not be regarded as a plain counting and ratio-calculating method but needs also to take composition, the deployment of words to a text, into consideration (cf. Treffers-Daller, Parslow & Williams 2018) as "it is not just about the range of words a reader possesses, but also about the ways in which words are deployed in a text" (ibid.: 4). Therefore, it is mandatory to analyze the full text-sample and thus, avoid truncation to apply the same sampling method to all text excerpts to be analyzed and to apply a method that yields comparable results independent of the sample size and to ensure methodological coherence (cf. Duran 2004: 222). Jarvis identified six distinct aspects that characterize or determine the perceived complexity of a text: "variability, volume, evenness, rarity, dispersion, and disparity" (2013: 18). Although the analysis of lexical diversity covers only one of Jarvis' aspects, namely diversity, it still yields enough information on a text to determine the variability (cf. Treffers-Daller, Parslow & Williams 2018: 5), which in turn provides information on requirements of a text

towards a reader's receptive skills. Treffers-Daller et al. (2018) align lexical diversity measures with CEFR scores providing the required information for this study.

A diversity measure calculated by *Text Inspector* is the MTLD value. Although the key figure provided by the "measure of textual lexical diversity" (MTLD) still has its shortcomings with regard to text length and the character of the value being a probability (cf. McCarthy & Jarvis 2010: 381) it is regarded as being robust (Malvern et al. 2004). The MTLD value is calculated sequentially, measuring the TTR for each word in a text and working to identify a "mean length of sequential word strings that maintain a given TTR value (0.720)" (ibid.: 384) until this
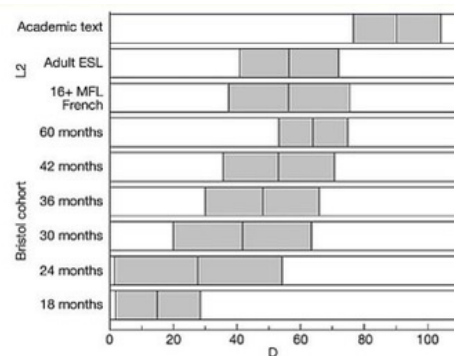
Figure 6.4 - Means and sub-ranges (10th–90th percentile) of D for various cohorts 8Malvern et al. 2004: 238)

TTR value drops to the given value 0.720, making the string at this value the first factor (cf. Treffers-Daller, Parslow & Williams 2018). MTLD measurements do not cut off words after the "factor count increased the value of 1" (ibid.), meaning that all words in a text are taken into consideration and further factors are calculated until the complete text is factorized. The MTLD finally is calculated by dividing the number of words by the total of factors computed (cf. Treffers-Daller, Parslow & Williams 2018). This approach delivers a complete assessment of a text and a comprehensive impression of the complexity of the text at hand. Malvern et al. created a scale to allocate the level of skill when it comes to language usage for both native speakers as well as language learners. This scale indicates the typical diversity measure of selected age groups or groups of learners, based on the findings in their study "Lexical Development and Language Development" of 2004. However, it has to be noted that MTLD does not respect the content complexity or cognitive challenge stemming from the words but measures the diversity of vocabulary used in a text (cf. Malvern et al. 2004: 384f.). The calculated value D for the analyzed textbook material was compared to that scale to deduct the text's diversity based on the information on age and the typical diversity measure provided by Malvern et al.

6.3.3   Lexical or propositional density

While lexical diversity is concerned with the amount of different expressions in a text, *lexical density* "provides a measure of the proportion of lexical items" (Johansson 2008: 61). The term *propositional density* describes "the amount of information conveyed in an object or environment per unit element" (Lidwell 2009) and goes back to Noam Chomsky's *Syntactic Structures* (ibid.). In other words, lexical density indicates the way information is packaged (cf. Johansson 2008: 65) within a text, meaning that a text with a high number of content words transports significantly more information than a text with a high number of function words (ibid.). Additionally, it has to be noted that lexical sophistication alone cannot be seen as a reliable indicator for lexical density or vice versa. Lexical density, as indicated above, may be branched out to variants paying attention to the density of nouns, verbs, adjectives, etc.

(ibid.). *Text Inspector* differentiates between nominal and verbal elements per sentence to determine lexical density (cf. Textinspector.com 2020). Nominal elements are nouns and adjectives grouped together while verbal elements look at all elements or items derived from a verb (cf. Kapust 1998). Both are counted per sentence, indicating the density of information per sentence in a text. It has to be taken into consideration that if a decisive amount "of information is conveyed in a small amount of text, the reader can become confused" (Ziafar & Namaziandost 2019: 819). This is especially a potential scenario if readers are learners of the text's language. Consequently, *Text Inspector* rates complexity being high the higher the density measures (cf. Textinspector.com 2020). This way, *Text Inspector* finds another approach to determine this key figure to assess readability or text difficulty. *Text Inspector* does not follow the idea of finding proportions of a whole text but rather on a sentence level "measuring the number of unique semantic units" (Ziafar & Namaziandost 2019: 816). It appears all the more suitable for textbook material to work on a sentence level since during teaching, sometimes single sentences are focused on.

6.3.4   Flesch-Reading-Ease index

Another indicator that helps to determine the complexity and demand of a text is the Flesch-Reading-Ease Index. It accommodates for different levels of complexity in reading. The index indicates the level of complexity of a text on a semantic level and expresses this as readability. Factors like choice of words, i.e. vocabulary, length of sentences, and propositional density, i.e. the ratio of content vs. function words are taken into consideration (cf. Flesch 1948). Initially created for the assessment of marketing and advertisement, the index was soon introduced to other contexts, such as education. Although today criticized for its "inability to take into consideration deeper levels of text processing" (Ziafar & Namaziandost 2019: 816), it is still considered a helpful tool within the educational community if analyses are enhanced by aspects that cover the "blind spots" of in-depth text processing (ibid.). Furthermore, the Flesch-Reading-Ease index is still deemed reliable when it comes to the prediction of difficulty of a text (cf. DuBay 2004: 54–56).

The index is based on a 0-100 scale, with high scores indicating a lower level of complexity while a lower score indicates high levels of complexity. To arrive at those scores, the following calculation has to be conducted:

$$206.835 - 1.015 \times (words/sentences) - 84.6 \times (syllables/words)$$

(Flesch 1981: 147).

The results of this algorithm arrive at the order of scores, which also ascribes a school grade to each level (see Table 6.1 below). It has to be noted, however, that this grade refers to students whose language of schooling is English per se.

Score	School level	Notes
100.00-90.00	5th grade	Very easy to read. Easily understood by an average 11-year-old student.
90.0–80.0	6th grade	Easy to read. Conversational English for consumers.
80.0–70.0	7th grade	Fairly easy to read.
70.0–60.0	8th & 9th grade	Plain English. Easily understood by 13- to 15-year-old students.
60.0–50.0	10th to 12th grade	Fairly difficult to read.
50.0–30.0	College	Difficult to read.
30.0–10.0	College graduate	Very difficult to read. Best understood by university graduates.
10.0–0.0	Professional	Extremely difficult to read. Best understood by university graduates.

Table 6.1 - Flesch-Reading-Ease Score (Flesch 1949: 149)

The readability formula applied here has proven its validity over the decades (cf. DuBay 2004: 19). However, weaknesses should not be ignored as it does not respect aspects such as prior knowledge or actual reading competences into consideration (ibid.). It has to be noted that researchers came to the conclusion that although students may be in grade 9, their reading skills are much rather at grade 7 level (ibid.: 16).

The values from the analyses on Flesch-Reading-Ease address reading competences obtained among native speakers of English, i.e., learners' skills in their L1. As this study, however, focused on learners' skills in a foreign language, the scale used for native-speaking students had to be adjusted. To arrive at an appropriate scale for students in Germany, the main reading texts from German English textbooks, in specific the broadly-used *English G21* by Cornelsen publishing, were analyzed using *Text Inspector*. The values yielded in this analysis were matched with the scales provided by Flesch to adjust the numbers to the German context. To align the results of the analysis, the CEFR levels were assigned to the various grades on the adjusted scale (see chapter 6.4.1 below).

6.3.5 Grammatical aspects

Grammar plays a central role in human communication (cf. Newby 2013) and hence, the analysis of grammar within the selected textbook material was required. In English teaching, grammatical phenomena such as the various tenses, aspect, modality, conditionals, and the passive are introduced progressively following the English curriculum[11]. The passive voice, e.g., is introduced in grades 7 or 8 (cf. Ministerium für Bildung, Wissenschaft und Weiterbildung 2000: 45), hence the textbook by Cornelsen follows suit and introduces the passive in unit 3 of its grade 7 textbook (cf. English

Figure 6.5 - English Grammar Profile, Excerpt for "Passive"

G21, A3 2010: 156f.) and develops further and more sophisticated usages of grammar throughout years 8 and 9.

For this analysis, all textbook material was examined thoroughly for all grammatical phenomena used in the texts. First, the phenomenon was checked against the curriculum and after that, the English Grammar Profile[12], the equivalent to the English Vocabulary Profile, was consulted to obtain a CEFR level for the grammatical form at hand. It was important here to respect the various possible uses of the grammatical form, i.e., "Can use the passive with 'by' in a relative clause […]" (Cambridge UP 2015: sec. passive) to determine the correct CEFR level. Especially with grammatical phenomena used in the selected textbook material at a point at which said phenomenon as per the curriculum would have to be regarded as unknown to students, the analysis discussed the assigned CEFR level in detail (see chapters 7.1 to 7.4 for further information).

6.4 Data treatment

Usually, textbooks come with scaffolding, mostly helpful chunks or language activation cues. All but the Mathematics textbooks offered vocabulary help, either monolingually or as translations. Furthermore, teachers presumably provide students with additional vocabulary help, especially for technical terms of the subject. To assess the text not only in their non-scaffolded form but also assess the scaffolds, the texts were analyzed in two iterations. The first iteration examined the texts in their original version. A second iteration worked with altered texts, altered as such that those terms provided by the textbooks – already translated or explained through scaffolding – were replaced by either the

---

[11] During the composition of this study, a new and revised version of the English curriculum was published. This version does not longer imply specific grammatical phenomena and their point of teaching. The information given here is hence based on the old version of the English curriculum in Rhineland-Palatinate, first published in 2000.

[12] This thesis has made use of the English Grammar Profile. This resource is based on extensive research using the Cambridge Learner Corpus and is part of the English Profile program, which aims to provide evidence about language use that helps to produce better language teaching materials. This information has to be given in any paper using the EGP. See http://www.englishprofile.org for more information.

translation or the explanation. Consequently, the readability of the texts suffered to a great extent but for analyzing purposes, i.e., to reflect upon how complex a text could still be when worked with in class, this iteration was inevitable.

For History, the scaffolded terms were replaced by the explanation, e.g. the term *heraldry* is explained as "the study of the coats of arms of old families" (Kröger & Lohmann 2007: 9), accordingly *heraldry* was replaced by this explanation within the text. In total, 44 terms were replaced for Exploring History 1, Chapter 1 whereas Chapter 4 contained 75 replacements. In Exploring History 2, 15 terms were replaced in Chapter 3 and 25 in Chapter 4 respectively. Especially in History, it has to be noted that the terms explained are not necessarily technical terms in the literal sense. Oftentimes, words of rather everyday language are explained such as "commodities" being "goods" or "to falter" meaning "to weaken" (Kröger 2009: 85).

For Geography, the quality of vocabulary help is different from Exploring History. *Diercke Geography for bilingual classes* offers scaffolding only by listing key terms important for the text at hand. Those key terms are then collected in the textbook's chapter and simply translated into a single German equivalent. Following the same approach, those key terms were replaced within the text with their German equivalent, e.g. *trench* is replaced with "Graben" within the text on tsunamis (cf. Löbmann 2014: 18). Chapter 1 of Volume One features 69 terms in total, Chapter 2 provides 81 key terms. The chapters selected from Volume 2 for grade 9, *Global Disparities* and *Europe Changes* feature 44 and 32 key terms, respectively.

In Biology, the quality of technical terms changes once more. Here, technical terms are actual technical terms which also in their German version are not intelligible per se. Oftentimes, those terms are derived from Greek or Latin, e.g. *enzyme* or *bronchus* (i.a., cf. Mathews & Olmesdahl 2010: 30), a term that can also occur in German textbooks. *Discover Biology* also lists key terms essential for the text or topic at hand. Other than *Diercke Geography*, *Discover Biology* directly translates the key terms, i.e., there is no collection of vocabulary in the end of the book but translations in immediate proximity to the text that uses the term. These translations replaced the technical terms in the second iteration on the Biology texts run by *Text Inspector* for examination.

Mathematics material, however, needed additional data treatment. As outlined in 6.2.4 there is no German comprehensive textbook material for bilingual Mathematics classes and hence, various textbooks used in Great Britain provided the corpus to be analyzed. Mathematics teaching includes not only the use of technical terms but also the explanation of these terms. Hence, the material does not outline key terms much less translations. To mirror the approach for the three other subjects and their material, the technical terms were replaced by the definitions provided by Selkirk's *Mathematics Handbook* or *Oxford's Advanced Learners Dictionary* (OALD). Selkirk's handbook is explicitly directed at " 'A'-level and undergraduate students" and provides students with "clear and concise definitions" (Selkirk 1991: back cover). The online version of the OALD offers specialized group-entries for Mathematics and its sub-topics. While some technical terms of the teaching material are also

used in the German context, e.g., *horizontal* or *vertical*, other terms, however, do not occur in German textbooks, e.g., *quadrilateral*. The explanations or definitions gathered from OALD and Selkirk were listed in a document (see chapter 12.5) and replaced the original terms in the text for the second iteration.

During the analysis, especially types reported as unlisted within the English Vocabulary Profile (see chapter 6.3.1) underwent additional treatment. Every unlisted type was double-checked again with the website for EVP as such. In case the type could be retrieved from the website, the allocated CEFR level was entered into *Text Inspector* and the analysis' results. Names of locations, numbers extended with letters (e.g., 1990s), but also German terms were deducted from the count of unlisted types. Afterwards, based on the new total of types unlisted, a new percentage of unlisted words was calculated.

In the light of the above-described approach on data treatment, it seems as if the process deviates with each subject. The results as they were obtained may at a first glance not be comparable since in History and Mathematics, English explanations replaced the original terms in the textbook material while in Geography and Biology, translations were used to treat the data. This may hold true, especially since German terms were not rated or assessed by *Text Inspector*. Yet still, this modus operandi was chosen to reflect upon the actual teaching in class using the selected material. Teachers would use what the textbooks provide to work the vocabulary. Especially in Mathematics, teachers might work with definitions in English to teach the matter at hand. The intention to adhere as authentically as possible to the actual teaching and working with the texts analyzed legitimates the varying replacement procedures chosen for the four selected topics. The results as described in chapter 7 reflect on the texts in their altered form, i.e., after the data treatment.

6.4.1 Adjusted scales for Flesch-Reading-Ease index

Before presenting the results for each subject, it is essential to first introduce the adjusted scale for the Flesch-Reading-Ease Index. As this adjustment was based on the analyses of two main texts featured in English G21 for each grade, the analyses' results of those are listed first. The analyses for *English G21* yielded CEFR levels between A1+ to B1 as the overall level of complexity for the texts. Below, detailed results are listed per grade, textbook (Ax, with x indicating the number of the book and year of English instruction) and text (Tx, with x indicating either the first or second selected main text), e.g., for grade 6, the textbook A is A2 with the first text T taken from the early unit in this book, resulting in A2, T1(cf. English G21 2011: 22f.; 105ff):

Grade	Volume & text number (Ax, Tx)	Lexical Sophistication (CEFR) – average values			MTLD-measure	Flesch-Reading-Ease	Overall CEFR
		EVP	BNC	AWL			
6	A2, T1	A2	A1+	A2	64.79	96.60	A1+
6	A2, T2	B1	A2	A2	57.54	95.71	A2
7	A3, T1	B1+	A2	A2	70.09	91.82	A1+
7	A3, T2	B1	A2	A2	68.57	93.29	A1+
8	A4, T1	A2+	A2	A2+	69.92	94.21	A2
8	A4, T2	B1	A2	A2+	65.36	80.72	A2+
9	A5, T1	B1+	B2	A2	80.99	95.94	A2
9	A5, T2	B1	A2+	A2	88.70	91.60	A2+
10	A6, T1	B1+	B1	B1	99.80	78.99	B1
10	A6, T2	B1	B1+	A2+	98.79	77.01	B1

Table 6.2 - Overview of Text Inspector results for English G21 A2 - A6

The Flesch-Reading-Ease index also associates grades to the respective levels (see 6.3.3). Accordingly, above-listed results would also allow for a grade-related evaluation. Since the original index refers to native-speaking students and not to learners of the target language, this original allocation of grades to a scale should not be transferred onto the context of foreign language learners. The results for the analysis of *English G21* imply that students are supposed to achieve a comparable level with a year delay, i.e., Flesch-Reading-Ease index values of 100 to 90 are appropriate for students in grade 6 (while in the original version, grade 5 was allocated to this range).

Allocating CEFR levels to the Flesch-Reading-Ease scale, the adjusted version to be used is as follows:

Flesch-Reading-Ease Score	grade in the original scale	adjusted grade for learners of English	CEFR level (per *Text Inspector*)
100 - 90	5th	6th	A1+
89 - 80	6th	7th	A2
79 - 69	7th	8th & 9th	B1
68 - 57	8th & 9th	10th - 13th	B2
57 – 54	10th to 12th	10th - 13th bilingual	C1
< 54	college, college graduate, and professionals	university level	C2

Table 6.3 - Adjusted scale of grades in comparison to original scale of grades for Flesch-Reading-Ease

6.4.2 Calculating CEFR levels

The analysis was conducted to determine a CEFR level for each selected key figure as described in chapter 6.3. To be able to make robust statements and draw conclusions on the effect and implication of those CEFR levels, it was necessary to calculate average values per key figure. Hence, point values were assigned to each CEFR level. The assigned point values and their incremental increase respect that the ascribed competences of the distinct CEFR level do not follow a linear progression, i.e., the description of the competences per level of skill in the *Companion Volume With New Descriptors* to the CEFR reveal that the progress from A1 to A2, e.g., involves comparably less growth in competences than from level A2 to B1, and especially B2. The spoken skill, e.g., increase from A2 with "[…] a simple description […]" to the competence of "[…] fluently sustain a straightforward description" with level B1 (Council of Europe 2020: 69). Moreover, the skills in reading or much rather their differentiation within the CEFR descriptors show the progress more clearly: while at level B1, learners "can recognise the significant points in straightforward newspaper articles", the considerable increase of skills at level B2 is evident with learners being able to "[…] understand articles and reports" (Council of Europe 2020: 63) and even differentiate between factual information and biased positions expressed in text form (ibid.). Consulting the more differentiated descriptors for *Reading for Information and Argument* as an example (ibid.: 61f.), the development and the width becomes evident. For level A1, reading comprehension is limited to "understand[ing] short texts on topics of personal interest […] written in simple words and supported by illustrations" (ibid.: 63) with the aim of coming to a basic, general

understanding of the text. Level A2 enhances this to, e.g., "understand[ing] a short factual description […] within [own] field, provided that it is written in simple language [without] unpredictable detail." (ibid.). Further, the aim at this level is to be able to "identify specific information in simple written material" (ibid.). Consequently, a broader lexis is required for topics beyond the readers field of personal interest along with reading comprehension capabilities that do not rely on visual support.

Spoken production skills improve from the ability to "make him/herself understood in very short utterances, even though pauses, false starts and reformulation are very evident" at level A2 to the ability to "keep going comprehensibly, even though pausing for grammatical and lexical planning and repair is very evident, especially in longer stretches of free production." at level B1 to even the ability to "produce stretches of language with a fairly even tempo; although he/she can be hesitant as he or she searches for patterns and expressions, there are few noticeably long pauses." (Council of Europe 2018: 29) at level B2 targeted towards the end of secondary education. This is consistent with the described augmentation for the productive skill of speaking as outlined in *Overall Spoken Production* (ibid. : 69f.). While at level A1, speakers are expected to be able to "produce short phrases about themselves […]" (ibid.), or for *Sustained Monologue: Giving Information* (ibid.: 71) "giving a simple description […] using basic words, phrases and formulaic expressions" (ibid.), at level A2 the range of vocabulary and skill is enhanced to "[…] a short series of simple phrases […]" (ibid.: 69) and "[…] sequential connectors […]" (ibid.: 71). Although the vocabulary used for speech acts in both cases is based on a corpus of personal interest and environment, the increase is given. Hence, assigning the point value of 1 to level A1 and point value 3 to level A2 was justified. As level A1+ marks the transition phase from A1 to A2, the point value 2 reflected that development as well.

The spread of the assigned numerical values reflects upon the development. This span is further supported by the curricular standards set by the curriculum for English and the standards for the German Abitur as outlined in "Verwaltungsvorschrift des Ministeriums für Bildung, Wissenschaft, Weiterbildung und Kultur vom 14. Januar 2014 (9413 C / 944 A – 51 410/34/35)" (Ministerium für Bildung, Wissenschaft, Weiterbildung und Kultur 2014b). Both set the targeted level B1 in grades 7 to 10 (see also Table 7.15 - Targeted CEFR levels per grade), proving that the development of competences to fulfill the requirements by that CEFR level is broad and thus, takes time. Likewise, the spread within the C-band is chosen to be smaller by single point increments since spoken language, e.g., only changes with regards to the pauses being no longer noticeable (ibid.). Only for D1, an additional level added by *Text Inspector* for academic language (cf. Textinspector.com 2020), a difference of three points is made to reflect the increased complexity that comes with academic language.

This spread was not only used to calculate averages for text complexity but also for the assessment of the language proficiency required by the tasks which are subject of RQIII. The points were assigned as follows:

CEFR level	Assigned Point Value
A1	1
A1+	2
A2	3
A2+	4
B1	6
B1+	7
B2	10
B2+	11
C1	12
C1+	13
C2	15
C2+	16
D1	18

Table 6.4 - CEFR levels and assigned point values for calculations

7   Textbook material – study results

To implement chapters 6.2 and 6.3, this chapter describes the analysis of the selected texts using the key figures chosen from the results *Text Inspector* yields. First, the results from the text analysis of the original texts are described per subject and selected chapter, followed by outlining differences caused by data treatment as discussed in chapter 6.4.

7.1   History

Exploring History 1, Chapters 1 and 4

With its specific role and characteristics of language in History (see chapter 4), the level of difficulty or the linguistic demand of the texts used in class is crucial for learning outcomes. With History solely based on text and the main way to convey knowledge and understanding of historical aspects is through and with text, the material should reflect upon the proficiency learners have while working with it. This essential role language in the form of texts plays in History called for the analysis conducted to differentiate between text that is written exclusively for the textbook and text that comes in the form of historical sources. While the former can be adjusted to the proficiency of the target audience, i.e., students, the latter cannot and may not be altered. Main texts and sources were analyzed separately to examine the requirements each genre imposes on students, especially since source texts, up front, were assumed to be more challenging (see chapter 4.1).

For chapter 1 on *Absolutism and Enlightenment* and its four subchapters and source texts (cf. Kröger & Lohmann 2007: 6-20), the overall assessment by *Text Inspector* for CEFR levels ranged between B2 and C1 per subchapter and their sources. Of the total 15 metrics used by *Text Inspector*, lexical sophistication for EVP, e.g., ranged between B1+ to D1, with D1 being an additional CEFR level assigned especially to texts of academic quality. For comparability and information value, these values were translated into CEFR levels, given as an overview (see Table 7.1). The percentage rate of words at level A1 among the distinct text segments ranged between 27.36%, and 42.34%. The lower percentage rate of 27.36% indicates that the majority of words used were more demanding. This translated into to CEFR level D1. The higher percentage rate of 42.34% resulted in an overall CEFR level of B2. The next analytical step looked for the percentage rates of words within level B2. For words at the B2 level, the analysis found a spread throughout the subchapters between 7.14% (B1+) and 14.05% (C2+). Finally, when analyzing the texts for words or types of the C1 level, 1.83% (B2+) was the lowest percentage rate found in the texts, while 3.78% (C2+) was the highest rate. It has to be noted, though, that the conclusion based on these values in the respective CEFR-band is diametrical: while higher percentage values in the A-band imply a lower text complexity, high percentage rates of words within the B- and C-band imply a higher text complexity (see also chapter 6.3.1). In other words, texts with a majority of words being ranged in the A-band are considerably less complex than text whose majority of words is allocated in the B- or C-Band. This implication explained the assessment of D1 with comparably low percentage rates in the A-band in those specific texts. Looking at frequencies of the used words referring to the

BNC and the corresponding position of the word at the median position, the positions per chapter of this median word ranged between 280 and 650. Position 280 indicated a relatively high rank, while position 650 was considerably lower in rank. These ranks resulted in CEFR levels ranging between A2+ and C2. Accordingly, the percentage of words above those levels varied from 28.28% for A2+ to 0% above level C2. Between 13.06% and 24.45% of words were not listed in EVP but. The number of types within the texts found on the Academic Word Lists (AWL) was between 8.56% and 4.5%, or between C2 and B1+. In summary, the range required for this chapter for lexical sophistication was between A2+ and C2 (D1).

Further, the propositional density was calculated to determine the amount of information conveyed per subchapter. Values between 3.08 and 5.44 were identified for nominal elements per sentence, corresponding to CEFR levels between A2 and C1+. At the same time, the verbal elements per sentence varied from 1.85 to 2.75, with corresponding CEFR levels between A2+ and C1.

The analysis for the key figure of lexical diversity, MTLD, generated values that ranged between 66 and 104. According to the scale Malvern et al. developed, those values imply required lexical diversity to be similar to those observed in native children of 60 months and a diversity applied in academic texts. Finally, the Flesch-Reading-Ease indices spanned from 62, equivalent to B2, to 47, implying C2 level of readability, or reading skills to be achieved in grades 10 to 13 and to university level. With the CEFR levels identified and assigned, texts within the first chapter of this textbook, designed for students in 8th grade, require reading skills between A2+ and C2, which translates to students having to read texts which are short and simple, requiring students to only "understand specific information in simple language" (cf. Council of Europe 2018: 61). On the other end of the spectrum, students have to be able to work with "very long and complex texts", even "specialized texts" with a "varying type of discourse" on an "unfamiliar topic" (ibid.: 62-64).

A table listing the calculated values per key figure and subchapter can be found in the appendix (see chapter 12.3). MTLD is not expressed in terms of a CEFR level since its assessment follows the scale Malvern et al. created.

Key Figure	Subchapter							
	1.1	1.1	1.2	1.2	1.3	1.3	1.4	1.4
	Text	Source	Text	Source	Text	Source	Text	Source
EVP A1 (% of types)	D1	C2	C2	C2	C2	B1	C2	C2
EVP B2 (% of types)	C1	C2	C1	C1+	C2	C2	C2+	C1+
EVP C1 (% of types)	C2	C2	B1+	C2	C2	C2	C1	C1+
BNC 50th	C2	A2+	A2+	C1+	B1+	A2+	B1	A2+
AWL (% of types)	C2	C1	B1+	C2	B2	B1+	B1+	C1
Nominal Elements	A2	A2+	A2+	B1	B1+	A2	B1+	B1+
Verbal Elements	A2+	A2+	A2+	A2	A2+	A2	B1	A2+
Flesch-Reading-Ease	C2	C2	B2	C1	C2	B2	C2	B2+
Overall CEFR	C1	B2	B2	C1	B2+	C1	C1	B2

Table 7.1 - Exploring History 1, Chapter 1 - CEFR levels

With the MTLD not being expressed in terms of a CEFR level, an average value was still calculated to be 86.51, indicating a diversity comparable to that of academic text. The last figure to be averaged was the Flesch-Reading-Ease arriving at C1 or 58.29, which in accordance to Table 6.3 translated to a reading level to be achieved during grades 10 to 13 at a German school. Based on the CEFR levels of each key figure in combination with the overall assessment per subchapter, and as an effort to assign a CEFR level for the whole chapter, the individual results aggregated to a tendency towards CEFR level C1.

The grammar used within this subchapter covered most phenomena in the curriculum for grades 7 and 8. Some forms, however, were used although potentially overtaxing for students and would need scaffolding or even grammar-oriented language work. One phenomenon in the first chapter was the use of a non-finite adverbial subordinate clause with '-ing': "[…] after a long time of being persecuted […]" (Kröger & Lohmann 2007: 8), which qualifies as a participle clause with a conjunction. This phenomenon is only introduced in grade 9 and was assessed as CEFR level B2 since it requires students

to be able to "use a non-finite subordinate clause with 'after' + 'having/being' + '-ed' form, before a main clause, to refer to past time." (Cambridge UP 2015: sec. CLAUSES, subordinate). Furthermore, the text features relative clauses, which in their basic form are introduced in grade 7 but the ones in *Exploring History* use both the pronoun as subject and "whose" as a relative pronoun. Those advanced forms are introduced in year 9 according to the curriculum (cf. Ministerium für Bildung, Wissenschaft und Weiterbildung 2000: 92). The assigned CEFR level was A2 for the relative clause using the pronoun as subject and B2 for relative clauses using "whose" as relative pronoun (cf. Cambridge UP 2015: sec. CLAUSES, relative).

Text Inspector's analysis of chapter 4 discussing *Industrialization* (cf. Kröger & Lohmann 2007: 54-76) resulted in CEFR levels of B1 to C2+, an overview of the individual results is given in Table 7.2. A detailed look at the selected key figures revealed both low and high CEFR levels. For lexical sophistication, the subchapters and textual sources showed CEFR levels between A2+ and C2. For the comparison against A1 level words in EVP, the analyzed texts or moreover, the percentage of words used in those texts per subchapter ranged from 54.11% and 30.67%, and A2+ and C2, respectively. For B2-level words, *Text Inspector* identified percentage rates between 6.49%, which resulted being rated as level B1+, and 13.5%, which was rated as a share to be equivalent with C2. Finally, for C1-level words, the results ranged between 0.87% of all words, resulting in a B1+ rating, and 7.36% as the highest value causing the CEFR rating to be at D1, implying academic level. The analysis for BNC measurements yielded median results between positions 310 (A2) and 910 (C2+). The percentage of types above those CEFR levels was between 20.35% for A2 and 0% for C2+. The percentage of unlisted types ranged from 11.69% to 18.97%, meaning that these words may well occupy a rank in the BNC frequency lists but are not assessed within the EVP. The analysis of the subchapters for AWL revealed that the subchapter with the lowest count of types on either of the AWLs was at 2.55%, translating to a CEFR level at B1, while the subchapter with the highest count had a percentage of 8.74% types to be found on the AWLs, thus rated at C2.

For all seven subchapters analyzed, the propositional density for nominal elements transporting noun-related information was between 2.40 and 7.89 elements per sentence. This ratio translated to CEFR levels between A2+ and D1. For verbal elements, the count range started at 1.25 verbal elements per sentence and ended at 3.11 elements per sentence within the subchapters. Here, CEFR levels between A2+ and B2 were assigned. Concerning lexical diversity, MTLD values ranged between 58.91 and 120. Those values place the subchapters between lexical diversity or richness of native children displayed at 30 months and required in academic texts. Lastly, the Flesch-Reading-Ease index computed for the seven subchapters ranged from 73, equivalent to B1 or 8th-grade reading, and 38.02 with level D1 for the university level. Other than in the first chapter, the fourth did not use any grammatical forms students would not have learned by that time. A comprehensive table containing all numerical values for an overview of the above-described findings can be found in chapter 12.3.

The overall CEFR level for the subchapters differentiated for texts and sources resulting in a span from B1 to C2+. This span indicated, however, that the required level of language is inconsistent. While some subchapters may be just appropriate at level B1 or B1+ towards the end of the schoolyear, others are too demanding. Further implications and interpretation of such wide spans are given in chapter 7.5. Outside of the CEFR level assessment, the average value for MTLD arrived at 97.51, which on the scale by Malvern indicated that the texts were of academic quality. Lastly, the Flesch-Reading-Ease index averaged around C1 at a value of 55.89, translating into reading competences expected throughout school grades 10-13 of a bilingual track. With the eight out of 14 subchapters rated at level B2, and three within C-band, the overall assessment was that this chapter required level B2. That places both chapters, chapter 1 with a tendency towards C1 and chapter 4 towards B2, above the targeted language level at A2 at the end of grade 8 (cf. KMK 2003: 11f.).

The key figure values described above translate to the following CEFR levels not including MTLD-values since *Text Inspector* does not assign CEFR levels for lexical diversity measure MTLD:

Key Figure	Subchapter													
	4.1 Text	4.1 Source	4.2 Text	4.2 Source	4.3 Text	4.3 Source	4.4 Text	4.4 Source	4.5 Text	4.5 Source	4.6 Text	4.6 Source	4.7 Text	4.7 Source
EVP A1 (% of types)	B1+	C2	C2	C2	C2	B1	C2	C2	B1+	A2+	C2	C2	C2	C2+
EVP B2 (% of types)	C1	C2	C1	C1+	C2	C2	C2+	C1+	C2	B1+	C2+	C1+	C2	C2+
EVP C1 (% of types)	C2	C2	B1+	C2	C2	C2	C1	C1+	C2	B1+	B1+	B2	C1+	D1
BNC 50th	C1	C2	B2+	B1+	C2	C1	B1	B1	B1	A2	A2+	B1+	B1	C2+
AWL (% of types)	C2	C1	B1+	C2	B2	B1+	B1+	C1	C1	B1	B1+	B2+	C1	C2
Nominal Elements	A2	A2+	A2+	B1	B1+	A2	B1+	B1+	A2	A2+	B1	B2	A2+	D1
Verbal Elements	A2+	A2+	A2+	A2	A2+	A2	B1	A2+	A2	B1	A2+	A2	A2	B2
Flesch-Reading-Ease	C2	C2	B2	C1	C2	B2	C2	B2+	B2+	A2+	C2	C2	B2+	D1
Overall CEFR	B2	C1	B1+	B2+	C1	B1+	B2	B2	B2	B1	B2	B2	B2	C2+

Table 7.2 - Exploring History 1, Chapter 4 - CEFR levels

Exploring History 2 – Chapters 3 and 5

The second volume of *Exploring History* is aimed at grades 9 and 10 in German secondary education. Chapters selected for this analysis were chapter 3, dealing with *The Cold War and German Division* (cf. Kröger 2009: 72f.) and chapter 5 on *Superpower Rivalry* (ibid.: 128f.). Chapter 3 is subdivided into seven subchapters which again consist of main text and source texts to be analyzed separately. The overall CEFR levels *Text Inspector* assigned to the subchapters and their sources ranged from B2 to C2, for which an overview is given in Table 7.3. As this overall rating includes key figures that were not selected for this analysis; however, the selected individual figures will be broken down.

Starting with EVP A1 types and their percentages for chapter 3, *Text Inspector* determined rates between 46.32% and 23.03%, along with corresponding CEFR levels between B1+ and D1. For EVP B2 types, the analysis calculated percentage rates starting at 7.37% up to 16.39%, defining a span of CEFR levels from B2 to D1. The last figure computed for EVP-based ratios was the EVP C1 type percentage. The total amount of types counted to be at C1 was between 3.16% and 4.51%, translating to CEFR levels between C2 and C2+. Continuing with the comparison to the BNC, positions of the median word ranged from 310 to 1050. Therefore, CEFR levels between A2 and D1 were assigned. The share of types counted to be above the assigned levels and thus both supporting the assigned level and highlighting the share of types which were assumed to be more demanding ranged from 33.33% for the position at 310 and level A2 and 0% at position 1050. The percentage of types not listed in the EVP and thus not being assigned a level of difficulty ranged from 16.95% to 24.14%, making almost a quarter in that specific case. Unlisted types here were, e.g., "dungarees" or "stacked". Having covered the BNC for the types' frequencies, the comparison to the AWL revealed that the chapter also features words to be found on those lists. The percentage of words featured in both sources, AWL and text analyzed, ranged from 6.25% to 13.04% or in terms of CEFR levels, the difficulty level was graded to be between C1 and C2+. Moving on to the key figure not translated onto the CEFR but expressing lexical diversity on its own scale, MTLD was calculated. For MTLD, values were computed ranging from 70.09 to 125.04. These values for diversity measure D implied again required skills to be comparable to the lexical richness displayed in English-speaking children of 60 months of age and that of academic text according to the scale by Malvern et al.

On a sentence level, the analysis yielded results for nominal elements per sentence between 2.91, marking CEFR level A2 to 6.41 in a source text marking CEFR level C2+. Verbal elements per sentence ranged from 1.62 and CEFR level A2+ to 3.23 at CEFR level C2. Lastly, the Flesch-Reading-Ease index for the individual sections of chapter 3 ranged from 25.52, indicating reading capabilities required at university level or D1, to 63.07, which requires students' reading capabilities to be achieved in grades 9 and 10 and translates to CEFR level B2. The table below (see Table 7.3) provides a comprehensive overview of the individual CEFR levels bar the values and their implications for MTLD. The results of the analysis as such, i.e., the numerical values, are given in chapter 12.1.

Key Figure	Subchapter											
	3.1 Text	3.1 Source	3.2 Text	3.2 Source	3.3 Text	3.3 Source	3.4 Text	3.4 Source	3.5 Text	3.5 Source	3.6 Text	3.6 Source
EVP A1 (% of types)	D1	C2	C2+	B1+	D1	C2+	D1	C2+	C2+	C2+	D1	C2+
EVP B2 (% of types)	C2+	C2	C2	B2	C2+	C2+	C2+	D1	C2+	C2+	C2+	C2+
EVP C1 (% of types)	C2+	C2+	C2+	C2	C2+	C2	C2	C2+	C2+	C2	C2+	C2+
BNC 50th	C2+	A2+	D1	D1	A2+	B2+	C2	B2	C2	C2	C2	A2
AWL (% of types)	C2	C2	C1+	C1	C2+	C2	C2+	C2+	C2	C2	C2+	C2+
Nominal Elements	C2+	A2	B1+	B2	C2	A2+	B2	B1	B1+	B2	B2	B1
Verbal Elements	C1+	B1	B1+	B1	C2	A2+	C1+	A2+	B1	B1	B1	A2+
Flesch-Reading-Ease	D1	C2	C2+	B2	D1	C2+	D1	C2	C2	C2	D1	C2
Overall CEFR	C2+	B2	C2	C1	C1+	B2+	C2	C1	C1+	C1	C1+	B2

Table 7.3 - Exploring History 2, Chapter 3 - CEFR level

The overall levels for both texts and sources ranged from B2 to C2+ with nine out of 12 subchapters within the C-band. Consequently, an average overall CEFR level for this chapter would approximate level C1 given the frequency of this level being assigned to half of the subchapters. However, this level is targeted at the end of grade 13 for either students of bilingual tracks or students of an advanced course (cf. Ministerium für Bildung, Wissenschaft, Weiterbildung und Kultur 2014b: sec. 5). The average value for lexical diversity MTLD was 91.42, clearly indicating the lexical diversity of academic text required in accordance with the scale by Malvern et al. The average CEFR level for the Flesch-Reading-Ease index was calculated based on the findings described above to be at C2+ and a score of 45.20, corresponding with reading skills expected at university level.

Grammatically, this chapter did not employ phenomena outside the curriculum's ascribed range. Sentences mainly used the simple past with rare occasions of past progressive in a passive construction as in "was being fought" (Kröger 2009: 80) or using past tense forms together with the infinitive "was to take" (ibid.). Relative clauses used combinations of a preposition and a relative pronoun, as in "for whom" (ibid.). Since the texts narrate the developments leading up to the division of Germany after World War 2, the past perfect was used frequently.

Chapter 5 on *Superpower Rivalry* (cf. Kröger 2009: 128-146) was analyzed likewise. Here, the results yielded from the five subchapters and their texts and sources were again comparable with the results for chapter 3. The analysis started with the percentage values of words or types at the EVP A1 level. The highest calculated percentage was 41.54%, or level C1. The lowest computed value in the A1-band was 26.22%, meaning a little over one-quarter of words were at level A1, which indicated a CEFR level of D1, with almost 75% of all words classified above A1 per the EVP. For the EVP B2 band, the lowest percentage value indicating the lowest CEFR level was 8.09% or level B2+, while the highest share of the B2-band was at 13.72%, with a corresponding level of C2+. Compared to the B2-band, the EVP C1-band percentage values between 1.69% and 5.43% seemed considerably lower. However, their impact on the CEFR level was noticeable, requiring levels between B2 and D1. Continuing with the figures calculated for BNC, *Text Inspector* placed the median position for each subchapter, or the texts and sources separately, between positions 415 and 930. These two positions marked the beginning and end of the range for the required CEFR levels to be at A2+ and C2+. For those subchapters, the percentage of types above the assigned CEFR levels spanned from 27.42% for A2+ to 0% for C2, which is evident as there is no further differentiation above the C2-band in EVP. The share of words unlisted in EVP was between 16.05% and 24.92%, with terms like "armistice" or "counteroffensive" but also French loan words like "d'étente" being included in those percentage shares. The ratio of words found both in the texts analyzed and the AWLs was between 5.98% with an assigned CEFR level of B2 and 11.31% and a level of C2. The values calculated for MTLD were comparable to those of Malvern et al. for the group of 60-month-old children of their Bristol cohort (cf. Malvern et al. 2004) at a value of 66.64 and 103.46 compared to academic text.

On the sentence level, the calculated results for nominal elements per sentence started at 2.48 with an assigned CEFR level at A2 and moved up to 8.72 at level D1. Verbal elements per sentence were calculated from 1.89 at CEFR level A2+ to 3.94 at C2+. Reading the texts and sources of chapter 5 required reading skills to be between B2+ with a Flesch-Reading-Ease index being at 57.89, and D1 with the index at 33.34. Those requirements can be achieved from the end of year 10 onward and onto university level. The table below shows all CEFR levels put together; the corresponding numerical values are to be found in chapter 12.3. The overall assessment in terms of CEFR levels per subchapter ranged from B2 to C2.

Key Figure	Subchapter								
	5.1	5.1	5.2	5.2	5.3	5.3	5.4	5.4	5.5
	Text	Source	Text	Source	Text	Source	Text	Source	Text
EVP A1 (% of types)	C2+	C2	C2	C1	C2+	D1	C2+	C2	D1
EVP B2 (% of types)	C2+	B2+	C2	C2	C2+	C2+	C2	C2+	C2+
EVP C1 (% of types)	C2+	B2	C1+	D1	C2	D1	C2	D1	C2+
BNC 50th	C2	A2+	C1+	C2	C1	C2+	C2	B2+	C2
AWL (% of types)	C2	B2	B2+	B2+	B2+	C2	C1	C2	C1
Nominal Elements	B1	B1	A2	B1	B1	C2	A2+	D1	C1
Verbal Elements	A2+	B2+	B1	C2	B1	B1	B1	C2+	B1
Flesch-Reading-Ease	C2	C1	B2+	B2+	C2	C2+	C2	D1	C2+
Overall CEFR	C1	B2	B2+	B2+	C1	C2+	C1+	C2	C1+

Table 7.4 - Exploring History 2, Chapter 5 - CEFR levels

The calculated average CEFR level based on the numerical values (see chapter 12.3) for the Flesch-Reading-Ease index arrived at C2 and university level with a score of 48.58, tying in with the average value calculated for MTLD being 81.09 and hence, requiring academic level skills.

The grammar applied in this chapter was again suitable for the targeted group of students frequently using the simple past and past perfect. Source texts used conditional 1 and 2 ("If we attack Cuban missiles […]" (Kröger 2009: 135)) while eyewitness reports mainly employed the simple present (ibid.: 139). In the attempt to place the chapters as a whole at a CEFR level, chapter three allowed for an assessment of approximating level C1 or even C1+ with eight out of 12 values between C1 and C2. For chapter 5, six out of nine results were within the C-band which again placed this chapter as whole rather in the more demanding CEFR-band at C1+. The implications of those findings concerning the second and fourth research questions will be discussed in chapter 7.5.2.

7.2   Geography

The degree to which Geography relies on text is not as extensive as is the case with History (see chapter 4.1). While most phenomena, events, or statements made in the (non-digitized) past cannot be revisited or reproduced via video recording for History teaching, geographical aspects can be observed, sometimes even live, or traces of events are still visible. Geography teaching relies on text and textbooks; hence, the textbook material analysis for bilingual teaching. The analysis did not differentiate between primary texts and source texts as the latter were not given.

Diercke Geography Volume 1– Chapters 1 and 2

The first chapter, focusing on *Living on a Dynamic Earth* (cf. Löbmann 2007: 8-28), consists of five subchapters. All were analyzed for the same key figures described in 6.3 using *Text Inspector*. For the EVP levels indicating lexical sophistication, A1 types ranged between 33.78% and 39.83%, indicating CEFR levels between C2+ and C2. The lower amount again resulted in a higher text complexity as opposed to lower values in the EVP B2-band, which mean lower text complexity. For chapter 1 and its subchapters, the EVP B2-band percentages started at 7.49% and went up to 11.89%. The assigned CEFR levels spanned from B2 to C2. Furthermore, the values computed for the EVP C1-band varied between 0.68% and 1.81%, or the CEFR level between B1 and B2.

Looking further at the comparisons to the BNC, *Text Inspector* calculated the positions at or around the median type to range between 290 and 885. This range translated into CEFR levels or the requirement to be competent in those levels between A2 and C2+, creating a wide span throughout this chapter. The percentage of types exceeding the CEFR level assigned ranged from 25.86% (A2) to 0% (C2+) with a share of unlisted types between 16.16% and 12.85% featuring terms like "asthenosphere", a village's name in Germany "Windischenbach" but also "aftershocks", or "cherished". Especially the latter two are ranked in the BNC but EVP does not provide a CEFR level for these. Continuing with the percentages of words found in both the texts and the Academic Word Lists, *Text Inspector* produced results between 4.79% and 9.81% of words found on one of the lists. The assigned CEFR levels here ranged from B2 to C2. The measurement concerned with lexical diversity, MTLD, indicated the texts were placed with lexical diversity measures displayed in the age groups of 24 to 42-month-old children at 41.01 for MTLD. The upper values were calculated to be at 85.32 for MTLD, covering the range of lexical diversity displayed in academic texts.

The first chapter displayed a lexical density on the sentence level with nominal elements ranging from 1.65 to 3.30 or CEFR levels from A2 to B1. Verbal elements per sentence counts were between 0.77 and 1.45 and A1+ and A2. Lastly, the Flesch-Reading-Ease indices spanned from 71.70 at the lowest to 57.82 at the highest value, translating into reading skills requirements between A2+ and B2+.

All determined CEFR levels are listed in the table below, which serves the purpose of an overview. The corresponding numerical values are to be found in chapter 12.3.

Key Figure	Subchapter				
	1.1	1.2	1.3	1.4	1.5
EVP A1 (% of types)	C2+	C2	C2	C2	C2
EVP B2 (% of types)	B2	C1+	C2	C2	C2
EVP C1 (% of types)	B1	B1+	B1+	B2	B2
BNC 50th	A2	C2	A2+	C2+	C2
AWL (% of types)	B2	C1+	C1	B2+	C2
Nominal Elements	A2	A2+	B1	A2+	A2+
Verbal Elements	A1+	A2	A2	A2	A2
Flesch-Reading-Ease	A2+	B2+	B2+	B2	B2+
Overall CEFR	B1	B2+	B2	B2+	C1

Table 7.5 - Diercke Geography Volume 1, Chapter 1 - CEFR levels

Similar to History, averages were also calculated for the Flesch-Reading-Ease index which resulted in a value of 63.5, translating into CEFR level B1 targeted at the end of 8th grade according to the adjusted scale (see Table 6.3). The average value for MTLD, 72.9, suggested the requirement of lexical diversity for all subchapters comparable to that of academic texts. With a three out of five subchapters rated to be either B2 or B2+, this CEFR level was also adopted as an approximant overall level for the chapter as a whole.

The second chapter chosen for the analysis follows the above-described chapter immediately and deals with *Living in Different Climate Zones* (cf. Löbmann 2007: 33-68). It consists of nine subchapters with varying lengths in text. All subchapters, however, feature graphs, maps, and pictures. Employing *Text Inspector* and analyzing the selected key figures, calculations arrived at percentages for EVP A1-rated types from 44.32 to 31.64%, indicating a CEFR level span from B1+ to C2+. For the EVP B2-band, percentages were between 3.66 and 11.94% of all words featured in the respective text. Those results translated into CEFR levels between B1 and C2. As for the EVP C1-band, *Text Inspector* calculated percentage rates for the subchapter to go from 0.89% at the lowest to 2.81% at the highest, resulting in CEFR levels between B1+ and C2. Applying the comparison of each type featured in the

texts to the BNC database, *Text Inspector* placed the median positions per subchapter between 360, suggesting a rather frequent use, to 1080, which resulted in assigned CEFR levels between A2+ and D1. This marks again a wide span that must be respected when discussing the meaning of those findings in 7.5.3. While with the upper levels C2, C2+, and D1, no other types were rated to be above those, 20.13% of types were rated to be of higher linguistic demand than A2+. The share of unlisted and thus unranked types went from 13.08% to 23.72% caused by terms such as "aridity" and "elevation" but also subject-related terms like "permafrost", "Steppe", or "subpolar". The percentages of types matching the entries in the AWLs ranged from 3.64% marking CEFR level B1+ to 9.66%, with a corresponding level C2. The values calculated for MTLD showed a lexical diversity comparable to that of 42-month-old native children or adult ESL at 55.13 as the lowest and reaching the lexical diversity comparable to that of academic texts with a calculated value of 89.26 as the highest determined value. On a sentence level, the propositional density with nominal elements was between 2.26 and 3.62, or CEFR levels between A2 and B1+. The count for verbal elements per sentence ranged from 0.61 to 1.17, marking CEFR levels between A1+ and A2. As the final key figure provided by *Text Inspector*, the Flesch-Reading-Ease index for this chapter and its subchapters went from 68 with level B2 up to 42.64 and level C2+.

Again, a comprehensive overview of CEFR levels is considered to be helpful at this point (for numerical values, refer to chapter 12.3):

Key Figure	Subchapter								
	2.1	2.2	2.3	2.4	2.5	2.6	2.7	2.8	2.9
EVP A1 (% of types)	B2	C2	C2	B2	C2	C2+	C2	B1+	C2+
EVP B2 (% of types)	B1	C2	C2	B1+	C1+	C2	C1+	B1+	C1+
EVP C1 (% of types)	C1+	B2+	B1+	B1+	C1	C2	B1+	B1+	B1+
BNC 50th	A2+	C2	C2	C2	D1	C2+	B1	B1	C2+
AWL (% of types)	B1+	C2	C2	B2	C2	C2	C2	B2	C1+
Nominal Elements	A2+	A2	A2	A2	A2+	A2	B1	A2	A2
Verbal Elements	A1+	A2	A2	A2	A2	A1+	A2	A2	A1+
Flesch-Reading-Ease	B2	C1	B2+	B2	C2+	B2+	B2+	B2	C2
Overall CEFR	B1+	C1	B2+	B2	C1+	C1	B1+	B1+	C2

Table 7.6 - Diercke Geography Volume 1, Chapter 2 CEFR levels

The nine subchapters' overall CEFR level ranged from B1+ to C2. For MTLD, the average value arrived at 68.46, a diversity equivalent to native-speaking children at 60 months or the end of the spectrum of adult EFL with 6 hours of contact time per week. The average Flesch-Reading-Ease index value was computed and resulted in 58.85, a value located at reading skills to be achieved at the end of 10th grade and translating into B2 on the CEFR scale. Looking at the overall CEFR levels per subchapter, five out of nine chapters were rated to be in the B-band, while four were in the C-band. If, however, in an attempt to arrive at an assessment of the entire chapter, B1+ and C2/C1 were to level one another out, the second chapter approximated level B2 for the overall CEFR level.

Neither chapter displayed an extensive use of potentially unknown grammatical features. The texts mainly use the simple present to describe geographic facts and circumstances. Eyewitness reports on the 2004 tsunami (cf. Löbmann 2014: 17), e.g., made use of simple past or past perfect. Relative clauses were only rarely used. The only grammatical feature more advanced than the competences to be

expected from students in years 7 and 8 were adjectives in their comparative form (ibid.: 13). According to the curriculum, this would only be introduced in years 9 and 10 (cf. Ministerium für Bildung 2021: 67). Mostly, comparatives were used "before the noun" or combined as "the same comparative adjective […] to indicate change over time […]" (Cambridge UP 2015: sec. ADJECTIVE) and was assessed as CEFR level B1. In summary, both chapters tended towards level B2/B2+ in the overall assessment taking the overall CEFR level per subchapter into consideration.

<u>Diercke Geography Volume 2, Chapters 2 and 7</u>

Chapter 2 on *Global Disparities* (cf. Ahrend 2013: 24-38) features eight subchapters on the topic, all having approximately the same length. Characteristically of this textbook, all subchapters come with substantial graphical material such as maps, graphs, and pictures. However, when looking at the single key figures representing lexical sophistication, the percentage of used vocabulary allocated with the EVP A1-band was between 45.12% and 24.22% and translated into levels between B1+ and D1. Again, these results supported that the lower the percentage of A1-band was, the more complex the text and its CEFR level rating were perceived. Further on, the percentage of types within the EVP B2-band ranged from 11.11% to 20.18%, with corresponding CEFR levels from C2 to D1. Lastly, the CEFR level span assigned to the percentages of types within the C1-band was between C1 and C2+ based on values ranging from 2.17% to 4.38%.

With regard to the median position of types in the texts in comparison to the frequencies listed within the BNC, the position of types was between 435, marking level A2+ and 975 at level C2+. Types' percentages being assessed to be above the lowest CEFR level assigned was 29.29%. For the highest levels, C2+ and D1, no types were listed to be above that level. Of the given types, those unlisted with EVP and their respective percentage share of all types used in the subchapters ranged from 11.01% to 17.84%, including "globaliz[s]ation", "outsourcing", but also - and unexpectedly - "resulting" ("result" was graded at level B1). The comparison of the texts to the AWLs resulted in percentages of 8.25% to 17.78% of all words that were to be found on those lists. The corresponding CEFR levels were between C2 and D1. The calculations of the value for MTLD indicating lexical diversity resulted in values between 67.44, equivalent to the lexical diversity of 42- to 60-month-old native-speaking children on Malvern's scale and 106.25 on an academic level. After the word-level analysis, the analysis moved on to lexical or propositional density. The number of verbal elements per sentence ranged from 1.12 with an assigned CEFR level of A2 to 2.16 at level B1. Likewise, the number of nominal elements per sentence was between 3.39 and 5.83 or between CEFR levels A2+ and C2. Looking at Flesch-Reading-Ease figures as a last step, those ranged from 47.17 at level C2 to 31.83, marking academic level D1.

The respective CEFR level are summarized in the table below; information on the numerical values is to be found in chapter 12.3.

Key Figure	Subchapter							
	2.1	2.2	2.3	2.4	2.5	2.6	2.7	2.8
EVP A1 (% of types)	C2+	C2+	D1	C2+	C2+	C2+	B1+	C2
EVP B2 (% of types)	C2+	C2+	D1	D1	D1	D1	C2	D1
EVP C1 (% of types)	C1	C2+	C2	C1	C2	C1+	C2+	B2+
BNC 50th	B1+	C2+	C2+	D1	C2+	C2	A2+	C2+
AWL (% of types)	C2	C2+	D1	C2	C2+	C2+	C2	D1
Nominal Elements	B1	C1	C1+	C2	A2+	C1+	B2	B1+
Verbal Elements	A2	A2	A2+	A2+	A2+	A2+	B1	B1
Flesch-Reading-Ease	C2	C2+	D1	D1	C2	D1	C2+	C2+
Overall CEFR	B2+	C1+	C2	C2	C1+	C1+	B2+	C1+

Table 7.7 - Diercke Geography, Volume 2, Chapter 2 - CEFR levels

Lexical diversity averaged for MTLD at 81.72; a value found on the scale to be equivalent to academic texts. Lastly, the average Flesch-Reading-Ease index was computed to be 40.58, which can be found in the adjusted scale (see chapter 6.4.1) within university-level values and hence, CEFR-value C2+. Lastly, the overall levels per subchapter were between B2+ and C2 with four out of eight subchapters rated at C1+. This allows for the overall assessment of this chapter as a whole at level C1+ as well.

Chapter 7 on *Europe Changes* (cf. Ahrend 2013: 104-120) consists of eight subchapters, just like the previously analyzed chapter. All subchapters were again of comparable length except for the first, which had fewer pages and, thus, less text. Just like chapter 2, this subchapter also features many graphs and maps but also comes with tables and a considerably lower number of pictures. For chapter 7, the values for the EVP A1-band were between 36.46% and 21.81% resulting in CEFR levels between C2 and D1. Percentages for the EVP B2 level were between 11.12% and 18.62%, translating into CEFR levels between C2 and D1. Types allocated at the EVP C1 level spanned from 3.15% to 5.03%, assigning the same CEFR levels as the aforementioned EVP B2-band. On a word level, comparing all types used with the frequency tables of the BNC resulted in ranks for the types closest to the median positions between 495 and 990 and equivalent CEFR levels between B2 and C2+. The percentage of types above the assigned CEFR level ranged from 25.10% to 0%. Those subchapters were rated at level C2 or C2+.

Additionally, the percentage of unlisted types varied from 6.7% to 11.96%. Aside from names of cities, countries, or committees, terms like "roaming", "convergence", or "disparities" were not listed, along with "farmland" and "others". The percentages of types from the texts found within the various Academic Word Lists ranged between 6.15% at the lowest, rated C1, to 20.21%, rated D1. The measurements for lexical diversity arrived at values from 53.30 to 115.21, which describes a lexical diversity that can be found in native-speaking children between 24 and 42 months of age as well as with adult ESL and – at the upper end of the determined range – with academic texts. The propositional density of this chapter and its subchapters ranged from 1.57 verbal elements per sentence to 2.54, from level A2+ to B1+. Nominal elements, however, were counted to amount between 3.57 and 6 per sentence resulting in CEFR levels between B1 and C2. For all subchapters, the Flesch-Reading-Ease index was calculated to be between 55.26 as the highest value and lowest complexity, with CEFR level C2, to 31.81 as the lowest value and highest level at D1. The overall CEFR levels assigned by *Text Inspector* went from C1 to C2+. Put together, the analysis results expressed in CEFR levels were as follows:

Key Figure	Subchapter							
	7.1	7.2	7.3	7.4	7.5	7.6	7.7	7.8
EVP A1 (% of types)	C2	D1	C2+	D1	D1	C2+	C2+	C2+
EVP B2 (% of types)	C2+	D1	D1	D1	D1	C2	D1	D1
EVP C1 (% of types)	C2	C2	D1	D1	D1	C2	C2+	C2
BNC 50th	C2	B1	B2	C2	C2+	B2	B2	C1+
AWL (% of types)	C1	D1	C2+	D1	D1	D1	C2+	C2+
Nominal Elements	B2	B2	B1	B1+	B2+	C2	C2	B2
Verbal Elements	A2+	A2+	A2+	A2+	B1+	A2+	B1+	B1
Flesch-Reading-Ease	C2	C2+	D1	D1	C2+	C2+	D1	C2+
Overall CEFR	C1	C1	C1	C1+	C2	C1+	C1+	C1+

Table 7.8 - Diercke Geography, Volume 2, Chapter 7 - CEFR levels

The scores calculated for MTLD arrived at an average value of 81.3, which compared to the diversity found in academic text. Lastly, the Flesch-Reading-Ease index was at 40.88 on average, corresponding

again to university level and a level of complexity to be placed at C2+. Based on the overall CEFR levels of the individual subchapters, the overall CEFR level for chapter 7 as a whole could be classified as CEFR level C1+, just like chapter 2 above.

Both chapters employed mostly the simple present and the present perfect to describe the conditions and facts for the EU and in the context of globalization. Occasionally, a gerund form was used as well as the present progressive. The use of the progressive form as participle was one aspect that did not cohere to the curriculum since its use as a "subordinate clause to refer to time" (Cambridge UP 2015: sec. CLAUSES) is only introduced in year 9 or 10. The English Grammar profile ranked this grammatical aspect as CEFR level B1. Furthermore, the chapters both featured adjectives in their comparative and superlative form, including "most" as an indicator (cf. Ahrend 2013: 24). Although this was assessed as level A2 only, it is introduced in years 9 or 10 according to the curriculum (cf. Ministerium für Bildung 2021: 67).

7.3 Biology

In Biology, texts mostly describe bodily functions, natural phenomena, animal behavior, and botanical topics. Compared to History and Geography, Biology textbooks comprise a relatively larger number of technical terms, often originating in Latin or Greek expressions. The textbook analyzed, *Discover Biology* Volume 1, applies a two-step translation for those. English technical terms are translated into the German technical term, which often is a Latin-based equivalent, and then provides a translation or a transfer into German. Furthermore, Biology teaching often employs the practical application of knowledge utilizing experiments and observations. Hence, it is necessary to introduce related terminology and linguistic tools to conduct such experiments. As mentioned in chapter 6.2.3, *Discover Biology* Volume 1 covers all topics addressed in the curriculum for grades 7 to 9. Hence, this analysis focuses on chapters 2 and 3 for the first age group and 4 and 5 for the second.

Discover Biology Volume 1 – Chapters 2 and 3

The selected chapters introduce the topics of respiration and blood circulation. Chapter 2 features six content subchapters as well as a skill-related activity subchapter that serves as an instruction for topic-related experiments and associated activities to document experiment outcomes (cf. Mathews & Olmesdahl 2010: 30-43). The chapter concludes with a revision. The analysis examined eight subchapters for the selected key figures outlined in the chapter on methodology (see chapter 6.3).

Starting with the English Vocabulary Profile (EVP) and its three category levels, A1, B2, C1, *Text Inspector* calculated the percentage of types within the subchapters allocated within the first band, A1, between 49.65% and 49.74%, which in turn meant CEFR levels between B1+ and C2. The percentages of types allocated in the EVP B2-band went from 11.81% as the highest value down to 5.59% as the lowest, reflected in levels C2 and B1+ accordingly. Lastly, CEFR levels determined for types within the C1-band were between C2+ and B1, based on percentages between 4.33% and 0.70%.

When computing the selected texts to determine the frequency of words used within the BNC and find the median position, the analysis yielded results per subchapter between positions 190 and 800. These translated into assigned CEFR levels from A1+ to C2, indicating considerable differences between the subchapters. Likewise, the percentages of those types rated at a CEFR level above the one assigned due to the median's position was between 37.89% (for level A1+) and 0% (for level C2). Moreover, the percentages for unlisted words ranged from 5,88% to 12,83%, mostly being Greek- or Latin-based words like "bronchiole", which even do not occur on BNCs frequency ranking. The comparison of the texts with the AWLs found between 1.29% and 4.78% of all types on one of those lists. Most of the types were found in AWL 1, which is concerned with general academic language (cf. Coxhead 1998). The resulting percentages translated into corresponding CEFR levels ranging from A2+ to B2+. When moving on from lexical sophistication to lexical diversity, *Text Inspector* calculated MTLD diversity values between 51.59 and 102.75. These values indicated lexical diversity requirements comparable to those observed in the competences of native-speaking children aged 30 months (quite advanced) to 42 months (average diversity). Alternately, this lower level of MTLD can be observed on average in adult ESL with six contact hours per week. The upper value calculated indicated that lexical diversity was comparable to academic texts.

On a sentence level, the number of nominal elements per sentence varied from 1.66 to 4.6, with assigned CEFR levels between A2 and B2. Verbal elements per sentence were between 1.15 to 2.8 and A2 to C1. Regarding readability, Flesch-Reading-Ease indices for the subchapters ranged from 78.49 at level A2+ to 58.07 at level B2+.

The following table lists all CEFR levels (numerical values are given in chapter 12.3):

Key Figure	Subchapter							
	2.1	2.2	2.3	2.4	2.5	2.6	2.7	2.8
EVP A1 (% of types)	C2	B1	B1+	C2	B1+	C1+	C1	C2
EVP B2 (% of types)	B2+	B2	B1+	C2	B2+	C2	C1	B1+
EVP C1 (% of types)	B1+	B1	B1+	B2+	C1	B1+	C2+	B1+
BNC 50th	B1	A2	A1+	A2	B1	B1+	C2	B1+
AWL (% of types)	B1+	A2+	B1	B1+	B2+	B2	A2+	B1+
Nominal Elements	A2+	A2	A2+	A2+	A2	B2	B1	A2
Verbal Elements	B1	B1	B1+	A2+	A2	C1	B1	A2+
Flesch-Reading-Ease	A2+	A2+	B2	B2+	A2+	B2	B2+	A2+
Overall CEFR	B1+	B1	B1	B2	B1	C1	C1	B1+

Table 7.9 - Discover Biology Volume 1, Chapter 2 - CEFR levels

The mean value for measurements on lexical diversity D was rounded to 68, equating with a diversity found in advanced stages for 60-month-old native children and advanced adult ESL. Lastly, the average calculated for the Flesch-Reading-Ease index was 68.5, which translated to B2. The overall CEFR levels required by the individual subchapters ranged from B1 to C1. With five out of eight subchapters being rated in the B1-band, the overall CEFR level of the complete subchapter approximated also level B1 as three out of five subchapters within the B1-band were at level B1.

Chapter 3 features eight subchapters on the thematic issue of blood and its circulation along with a subchapter on subject-related skills and activities, i.e., observation of one's pulse and a revision of all contents (cf. Mathews & Olmesdahl 2010: 44-57). Both chapters 2 and 3 also feature schematic drawings, foremost physiognomic cross-sections of body parts discussed in the chapter, e.g., "Outer and inner structure of the heart" (Mathews & Olmesdahl 2010: 57). Although relatively short with, on average, two pages dedicated to each subchapter, the key figures were calculated as follows.

The analysis of types to allocate them with either of the three EVP-related categories revealed that between 33.79% and 46.72% of all types were evaluated to be within the EVP A1-band, the corresponding CEFR levels ranged from C2+ to B1+. Types rated for the EVP B2-band made up between 8.08% and 11.80% of all types resulting in levels between B2+ and C2. For the EVP C1-band, the percentages calculated ranged from 0.75% to 4.17% or in terms of CEFR levels from B1 to C2+. The analysis of the BNC to determine the median positions and their closest words for each subchapter yielded positions ranging from 180, allocated to level A1+, to 625, allocated to level C1+. The percentage of types above the assigned CEFR levels ranged from 37.61% (at level A1+) to 1.09% (at level C1+). Unlisted types went from 5.18% to 12.96%, including "phagocyte" and "thrombocyte". The comparison with the AWL revealed that between 2.76% to 9.59% of all types were on either of those lists. The percentage values resulted in CEFR levels to be assigned ranging from B1 to C2. Measurements for lexical diversity (MTLD) revealed that the texts featured in the subchapters ranged from 30.13 to 76.91, which for the lowest value was equivalent to diversity measure ranges observed in native-speaking children aged 24 to 60 months up to diversity measures observed at the beginning of the range for academic texts.

On the sentence level, the calculations for nominal elements per sentence yielded results between 2.42 and 3.87 with assigned CEFR levels between A2 and B1. Verbal elements were less frequent, resulting in a range from 1.45 to 2.27 per sentence and corresponding CEFR levels between A2 and B1. Readability scores varied from 70.52 and level B1+ to 52.85 and level C2 on the Flesch-Reading-Ease index.

Overall, the results looked as displayed in the following table (see chapter 12.3 for the corresponding numerical values).

Key Figure	Subchapter									
	3.1	3.2	3.3	3.4	3.5	3.6	3.7	3.8	3.9	3.10
EVP A1 (% of types)	C2	C2	B1+	C1	B1+	C2	B2+	B1+	C2+	C2
EVP B2 (% of types)	C1	C2	B2+	C2	B1+	C1	C2	C1	C1	C1+
EVP C1 (% of types)	B2	C2+	B1+	C2	C2+	B1	C2	B1+	B1+	B1+
BNC 50th	B1	A2	A2	B2	A1+	B1	C1+	A1+	A1+	A2+
AWL (% of types)	B2	C2	B2	B2	C1	B1+	C2	C1	B1	B2+
Nominal Elements	B1	B1+	B1	A2+	A2+	B1	B1+	A2	B1	B1+
Verbal Elements	B1	B1	A2+	A2	A2+	B1	A2+	A2	A2+	A2+
Flesch-Reading-Ease	B2+	B2+	B2	B1	C2	B2	B2	B2	B2	B2+
Overall CEFR	B2	B2+	B1+	B2	B1+	B1+	B2+	B1+	B1+	B2

Table 7.10 - Discover Biology Volume 1, Chapter 3 - CEFR level

Outside of the CEFR classification, the average MTLD value was calculated at 49.02, which describes an average lexical diversity for this chapter comparable to the span observed among advanced 24-months-old and average 42-months-old native-speaking children as well as average adult ESL learners. The analysis calculated at an average Flesch-Reading-Ease index to be 63.99, which, according to the adjusted scale (see chapter 6.4.1), placed the complete chapter at grades 9 & 10 or an average CEFR level of B2. Following suit and attempting to place the overall CEFR level of all subchapters combined at a single CEFR level, the majority of subchapters was rated at B1+. Since the remaining subchapters ranged from B2 to B2+, the overall CEFR level approximated B1+ on the verge to B2.

Both chapters of *Discover Biology* used grammatical features on an appropriate level. The phenomena of respiration and blood circulation were primarily described using the simple present. Occasionally, the passive was used here, along with present perfect. Selected tasks made use of either conditional 1 or 2. Likewise, the modal "could" was used to express ability. Relative clauses with the pronoun "which" as an object and selected demonstratives were used. All those are covered in years 7

and 8 according to the curriculum (cf. Ministerium für Bildung 2021: 45). Only the use of adjectives in their comparative and superlative forms exceeds what would be taught in the targeted grade for these two chapters. The usage of those adjective forms was graded as CEFR level B1.

Discover Biology Volume 1 – Chapters 4 and 5

While chapters 2 and 3 delve into topics of fundamental bodily functions, chapter 4 focuses on *Food and Nutrition* (cf. Mathews & Olmesdahl 2010: 58-73), and chapter 5 covers the subject of *Digestion and Absorption* (ibid.: 75-89). Chapter 4 includes six thematic subchapters along with a section on relevant topic-related skills and activities. Chapter 5 is more extensive, comprising nine subchapters and a section dedicated to skills and activities.

The analysis of chapter 4 for the key figures, based on the EVP assessments, yielded percentage values within the A1-band between 49.21% and 42.73%. A higher percentage in this band indicates a lower complexity of vocabulary. Hence, the CEFR values ranged from B1 to B2+. For the consecutive bands B2 and C1, higher values translate to higher complexity and required CEFR levels. For band B2, the percentages went from 4.35% to 8.96%, or in terms of levels, from B1+ to C1. Lastly, the computed values for C1-level words in the EVP were between 0.91% to 4.71% translating into B1+ and C2+. The comparison of the words used in the texts to the BNC database revealed that the determined position of the median for each subchapter ranged from position 240 to 660. Higher positions for the median indicate more frequent use of the words closest around that position and thus can be seen as more commonly known and less demanding. Based on that, the CEFR level here was between A2 and C2. The percentage of types rated above the particular CEFR level ranged from 26.70% to 0.00%. These figures indicate the share of types within a distinct subchapter above the identified CEFR level based on the BNC frequency. The higher the share, the more complex the text can be regarded. This aspect had to be considered when discussing the results and their implications in chapter 7.5. The percentages of types not listed and rated within the EVP ranged from 5.00% to 12.28%, including types like "poultry", "raspberries" ("raspberry" is also not listed), "binge", or "clerk". Following the analysis for BNC, the comparison of the texts to the Academic Word Lists revealed that the level for lexical sophistication was between A2 and C1 since between 1.09% and 7.25% of the types in the subchapters were also found on the AWLs. Concerning lexical diversity, the calculations based on the subchapters' texts resulted in values for MTLD ranging from 72.94 to 91.73, which indicated lexical diversity levels comparable to those of quite advanced 60-moth-old native speaking children or adult learners of English as a foreign language at the lower value to that of an academic text for the higher value.

Looking further at the ratio of nominal elements per sentence, *Text Inspector* calculated values from 2.67 to 3.65 elements per sentence resulting in the assignment of CEFR levels between A2 and B1. As for verbal elements, the results were between 1.42 and 3.47 elements per sentence, with a range in levels between A2 and B2. Grading the readability of the texts revealed Flesch-Reading-Ease index scores going from 74.54 to 58.63, expressed as CEFR levels between B1 and B2+. According to the

adjusted scale in chapter 6.4.1, those values corresponded to reading skills targeted in grades 8 and 9 and 10 to 13, respectively.

The quantity of figures discussed is better displayed in a table to give an overview of the analyses' results (corresponding numerical values are to be found in chapter 12.3):

Key Figure	Subchapter						
	4.1	4.2	4.3	4.4	4.5	4.6	4.7
EVP A1 (% of types)	B1	B1+	B1+	B2+	B1+	B2	B1+
EVP B2 (% of types)	C1	C1	B2	B1+	B1+	B1+	B1+
EVP C1 (% of types)	C2+	B2	B2+	B1+	B2	B1+	B2
BNC 50th	A2	A2+	B2	A2+	A2+	C2	B1+
AWL (% of types)	C1	C1	C1	B1	B2+	A2	B2+
Nominal Elements	B1	A2+	A2+	A2	A2	A2	A2+
Verbal Elements	B2	A2+	A2+	A2+	B1	A2	B1
Flesch-Reading-Ease	B2+	B2+	B2	B1+	B2+	B1	B1+
Overall CEFR	B2	B2	B2+	B1	B2	B2	B1+

Table 7.11 - Discover Biology Volume 1, Chapter 4 - CEFR level

The measurements of lexical diversity (MTLD) resulted in an average diversity of 82.20 across all subchapters. According to Malvern's scale, this value is comparable to the diversity displayed in an academic text. The average for the Flesch-Reading-Ease index was 65.43. This value corresponded to reading skills targeted in grades 10 to 13 or level B2. The overall CEFR level across all subchapters approximated level B2 as the majority of subchapters, five out of seven, were rated to be at this level.

The following chapter 5 features a total of 10 subchapters to be analyzed, nine of which are topic-related, while one covers skills and competences for Biology. Among these ten subchapters, the percentages for types within the EVP A1-band were between 46.32% at level B1+ and 31.11% at level C2+. For the EVP B2-band, the values ranged from 6.62% to 16.34%, which corresponded to a span within the CEFR levels from B1+ to D1. The span for EVP C1-band types was almost as wide, with levels between B1+ and C2+ and percentages between 1.20% and 4.02%. The calculations for the positions of the median words within the frequency tables of the BNC yielded results between 210 and 560, corresponding with CEFR levels between A1+ and B1+. Above these assigned CEFR levels were between 10.18% (at B1) and 39.55% (at A1+) of all types per subchapter. The share of types unlisted

with EVP ranged from 5.26% to 12.13%, i.a., "activate(d)", "starch", "windpipe", or "regeneration". The percentage rates of words featured in both the texts and the AWLs varied from 3.03% to 9.74%, equivalents in CEFR levels between B1 and C2. Computing the texts for lexical diversity, *Text Inspector* yielded MTLD values starting at 42.19, which marks a diversity comparable to that of advanced 24-month-old native speakers to the highest value calculated for MTLD at 114.23, which was equivalent to the lexical diversity in advanced academic texts.

Propositional density was calculated for nominal elements per sentence to start at 1.94 elements and increase to 5 elements per sentence. These values corresponded to CEFR levels between A1+ and B2+. Verbal element counts yielded the number of elements per sentence to vary between 1.06 and 2.22 and levels between A2 and B1+. Reading requirements were on CEFR levels from A2+ to C2 based on the calculated Flesch-Reading-Ease index scores between 79.04 and 53.09. Bringing all key figures together, the results for chapter 5 were (with the numerical values for this overview given in chapter 12.3):

Key Figure	Subchapter									
	5.1	5.2	5.3	5.4	5.5	5.6	5.7	5.8	5.9	5.10
EVP A1 (% of types)	C1	B1+	B1+	C2	C2+	B2+	C2	C2+	C2+	C2
EVP B2 (% of types)	C2	C2	B1+	C1+	C1	B1+	C2+	C2	D1	C1
EVP C1 (% of types)	B1+	B1+	C1	B2+	B1+	C2+	C1+	B1+	C2+	C2
BNC 50th	A2+	A2+	A2	B1+	B1	A2	B1	A1+	A2+	A2
AWL (% of types)	C1+	B1+	B1+	C1	B1	B2	C1+	C1	C2	C1
Nominal Elements	A2+	A2+	A2+	B1	B1+	B1	B2+	B1	B1+	A1+
Verbal Elements	B1+	A2+	B1	B1	B1	B1	A2+	A2+	B1	A2
Flesch-Reading-Ease	B2+	B1	A2+	B2+	B2	B2	C2	C1	C2	B1
Overall CEFR	B2+	B1+	B1	B2+	B2	B1+	B2+	B1+	B2+	B1+

Table 7.12 - Discover Biology Volume 1, Chapter 5 - CEFR level

Although single key figures were assigned considerably high CEFR levels, the overall assessment by *Text Inspector* for subchapter within chapter 5 ranged from levels B1 to B2+. Comparing the average measure of lexical diversity, which was 68.07, to the scale, this value indicated that the texts required, on average, a lexical diversity comparable to that displayed by very advanced 60-month-old native

speakers or equally advanced adult learners of English as a foreign language. The average score for the Flesch-Reading-Ease index at 63.85, and the average CEFR level was B2. The overall CEFR levels per subchapter were evenly distributed between the B1- and B2- band making a clear tendency for the complete chapter expressed as a single CEFR level impossible.

Grammatically, all used forms were in line with what the curriculum sets as the curricular standard. Most descriptive texts used the simple present, the present progressive, and occasionally the passive, conditional 1, and the gerund.

Concluding, chapters 2 and 3 arrived at a lower CEFR level than chapters 4 and 5. The lower CEFR levels assigned for chapters 4 and 5, however, were on par with chapters 2 and 3. Chapter 7.5.4 discusses the indication of each value, the averages and key figures for Biology material.

7.4   Mathematics

Other than the three subjects above, the analyzed material for Mathematics is a combined corpus taken from eight different textbooks published in the UK. To cover the topics outlined in 6.2.4, Algebra 1 and Basic Geometry for grade 7 and Algebra 2 and advanced Geometry for grade 9, tasks from those various textbooks were compiled to create a text that was comparable by quantity to those texts analyzed for the other subjects (see chapter 12.4 for the selection of sources). Accordingly, the description of the results follows rather the line of the topics covered and less the line of subchapters.

Grade 7 – Algebra 1 and Basic Geometry

Algebra 1 features the basic aspect of graphs, tables, and terms. It introduces technical terminology of that specific field, relationships of data, and proportional as well as inversely proportional functions and operations to solve these. Angles and intersecting lines are the general topics in basic Geometry. This part includes various types of angles, the topic of parallelism, calculating the angular sum of not only but also polygons, and introduces perpendicular bisectors. In total, the analyzed text comprises 140 single exercises, i.e., the instructive text comprised a total of 637 sentences. As mentioned above, definitions replaced the technical terms; see chapter 6.4.

With the substitutes, the percentage of types allocated to the EVP A1-band for Algebra 1 was 49.21% corresponding to level B1, while Basic Geometry arrived at level A2 with 55.91% of all types located in the A1-band. EVP B2-band types comprised 7.29% of Algebra and 6.07% of Geometry, assigning CEFR levels B2 and B1+ accordingly. Lastly, Algebra arrived at B1+ for 1.41% of all types assessed to be within the EVP C1-band, while for Geometry, 2.08% were attributed to that band translating into level C1 here. When comparing every word from the texts to the BNC, the median position for Algebra arrived at 363 with a CEFR level of A2+. For Geometry, the median position was at 302, corresponding to A2. For Algebra, 24.52% of all types were above the assigned level, referring to the median position, while for Geometry, 21.93% were above that level. The number of unlisted types arrived at 7.20% for Algebra and 6.56% for Geometry. Examples of unlisted types were, among others,

"arrowhead", "sized" (since "(to) size" is not listed either), "digit", "edged" (type also not listed for the verb use), and "coordinate". The categorization "unlisted" refers to words not being assigned a CEFR level within EVP although they are very likely ranked within the BNC frequency table. Comparing the texts to the Academic Word Lists revealed that 6.78% of the types were featured for Algebra. This percentage translates into CEFR level B2. For Geometry, however, the CEFR level was A2+, with 2.58% of all types included in the AWLs. Calculations of the MTLD value help to obtain further information on lexical diversity. For Algebra, the lexical diversity measure arrived at 33.02, a comparable value to the lexical diversity observed in averagely advanced 24-month-old native speakers. For Geometry, the MTLD was slightly higher at 36.02, which places the diversity at the same level as Algebra compared to Malvern's scale.

The analysis for nominal elements per sentence revealed a low count of these being at 1.44 for Algebra and 0.80 for Geometry, placing Algebra in the A2-band while Geometry was in the A1+-band. Verbal elements counted 0.75 per sentence in Algebra and 0.70 in Geometry, allocating both parts in the A1+-band. With the above-described text character, a collection of exercises and tasks, the Flesch-Reading-Ease index was calculated to be 74.05 for Algebra and 70.57 for Geometry. Allocating those values on the adjusted scale (see Table 6.3), Algebra and Geometry arrived at level B1 targeted in grades 8 and 9. Although for Mathematics, the number of calculated values and assigned CEFR level is lower than for the other subjects, a comprehensive overview in the form of a table is nevertheless helpful to get an impression of the linguistic demand here.

The numerical values on which the CEFR level are based are given in chapter 12.3.

Key Figure	Subchapters	
	Algebra	Geometry
EVP A1 (% of types)	B1	A2
EVP B2 (% of types)	B2	B1+
EVP C1 (% of types)	B1+	C1
BNC 50th	A2+	A2
AWL (% of types)	B2	A2+
Nominal Elements	A2	A1+
Verbal Elements	A1+	A1+
Flesch-Reading-Ease	B1	B1
Overall CEFR	B1	B1

Table 7.13 - Mathematics Corpus Grade 7 - CEFR level

The overall linguistic requirements for Algebra and Geometry arrived at CEFR level B1. The measure for lexical diversity was at 34.52 on average and indicated a lexical diversity comparable to that of averagely advanced 24-month-old native speakers. The calculation of the value for the Flesch-Reading-Ease index resulted in a score of 70.47, indicating reading complexity comparable to the targets set for grades 8 to 9 (see Table 6.3) and corresponding to CEFR level B1.

The corpus on both topics mainly employed the simple present and conditional type 1. Furthermore, passives occurred quite often as well as the modal "would" (cf. Parsons 2000: 142). All used grammatical forms were in line with the standards set by the curriculum (cf. Ministerium für Bildung, Wissenschaft und Weiterbildung 2000: 45). A single passive form, however, was not covered by the curriculum for grades 7 and 8. Elevate 1, level 5-6, employed the present participle "given" preceded by the auxiliary verb being in the progressive. This construction indicates an ongoing activity to express a temporal iteration: "Without being given some more information […]" (Baker 2008: 168). This form was assessed as CEFR level C1 and would be part of grammar teaching in grades 9 or 10.

<u>Grade 9 – Algebra 2 and Advanced Geometry</u>

In grade 9 at German secondary schools, Algebra 2 features systems of linear equations with increasing complexity and difficulty. It also discusses linear relationships along with the methodologies of substitution and accumulation to work with such systems of equations and reveal relationships of numbers (cf. Körner 2016a). Furthermore, linear optimization and linear inequations are a topic (cf. ibid., see chapter 12.4). Advanced Geometry introduces square shapes and calculations on those to introduce the Pythagoras Theorem later on (ibid.). Furthermore, advanced geometry includes elements of a right-angled triangle to finally lead to the introduction, understanding and application of the so-called pq-formula (ibid.). Accordingly, tasks were selected from the material described in chapter 6.2.4. The sum of exercises and tasks analyzed for those two parts was 246 consisting of 562 sentences.

Within this collection, the percentage of types from the EVP A1-band for Algebra was 49.66%, while 43.46% of Geometry were calculated for that band. These percentages were transferred onto CEFR level B1 for Algebra and B2 for Geometry. Looking at the EVP B2-band, *Text Inspector* identified 6.06% of all types to belong to that band in Geometry, and 7.95% were found in the Algebra part resulting in CEFR levels B1+ and B2+ respectively. In the third EVP-band for level C1, a share of 1.94% of types of the Algebra text was contributed to this category, while for Geometry, 2.18% were. With regards to the CEFR level in the context of EVP, the analysis assessed Algebra to be at B2+ and Geometry to be at C1. When comparing each word of the text to the tables in the BNC, the median of all types used for Geometry arrived at frequency position 307. Calculations for Algebra revealed that the position of the mean word was 266. Both positions indicated that the text is comprised of rather frequently used words, and thus the positions were rated to be at CEFR level A2. Of the featured types, 26.89% were above the assigned CEFR level for Algebra, while for Geometry, 23.72% were above. The percentage for unlisted types arrived at 7.38% for Algebra. For Geometry, the percentage was slightly higher at 11.63%. As unlisted types, "enlarge", "multiply", "orchard", and "grid" were listed. Comparing the collection for Algebra 2 to the AWLs resulted in the identification of 4.94% of the types used were also listed on the AWLs. This value transferred to level B1+. In the case of Geometry, 7.84% of words were also found on either of the Academic Word Lists, meaning that the CEFR level here was also B1+. The analyses concerned with lexical diversity resulted in values for MTLD for Algebra at 33.05, comparable to the lexical diversity Malvern et al. found with slightly advanced 24-month-old native speakers. For Geometry, the diversity measure was 41.88. This value corresponds to averagely advanced 30-month-old native speakers on Malvern's scale. This may seem surprising since the vocabulary used in Mathematics would most likely not be used by children of that age. But it has to be repeated that the MTLD measurement does not reflect upon the content and its complexity but on the number of different words used within a text (see chapter 6.3.2).

Counting the nominal elements per sentence for both parts resulted in the same CEFR level at A2, with Geometry arriving at 1.48 elements per sentence and Algebra slightly below that figure at 1.45. The results are likewise with verbal elements: for Algebra, the elements per sentence counted 0.85 at level A1+, and the same CEFR level applied to the Geometry part with 0.63 elements per sentence. In the Algebra part, the Flesch-Reading-Ease index score arrived at 66.87, which translated to a level of difficulty expected in grades 10-13 since it transferred onto CEFR level B2. For Geometry, the score was calculated to be 71.97 and thus resulting in CEFR level B1.

Overall, all determined and calculated key figures for both Algebra 2 and advanced Geometry for grade 9, the required CEFR level arrived at B1+ and B1 respectively for both topics. The average lexical diversity figure compared to the diversity of averagely advanced 30-month-old native speakers with a value of 37.47. The average score computed for the Flesch-Reading-Ease index over both parts was 69.42 marking the complexity students towards the end of grade 9 without bilingual teaching aim at a CEFR level of B1.

The individual results are more convenient to grasp in the form of a table (see chapter 12.3 for the numerical values).

Key Figure	Subchapters	
	Algebra	Geometry
EVP A1 (% of types)	B1	B2
EVP B2 (% of types)	B2+	B1+
EVP C1 (% of types)	B2+	C1
BNC 50th	A2	A2
AWL (% of types)	B1+	B1+
Nominal Elements	A2	A2
Verbal Elements	A1+	A1+
Flesch-Reading-Ease	B2	B1
Overall CEFR	B1+	B1

Table 7.14 - Mathematics Corpus Grade 9 - CEFR level

As was the case with the majority of the material for grade 7, the material for grade 9 did not contain any grammatical forms beyond the scope of the nominal standards set by the curriculum. Tasks and descriptions of mathematical rules, e.g., the Pythagoras Theorem (cf. Casson 2001: 186, 11), used mainly the simple present. Other tenses used were the simple past and the will-future. Many tasks used the conditional type 1 along with relative clauses, the gerund or passive constructions in the present tense. The only aspect of the abovementioned that was unexpected was the somewhat colloquial use of the combination of "get" plus the "-ing"-form as an informal participle clause as in "get learning" or "get working on" (cf. Parsons 2000: 40).

For both grades, the overall linguistic demand was at CEFR level B1, with only Algebra in grade 9 being slightly more complex at level B1+. This slight increase marked a development in linguistic demand appropriate for the grade and the assumed development of language competence among students.

7.5   Discussion of results

Before the results described above are discussed in detail, it is essential to recapitulate the implications the various key figures and their values have when it comes to assessing the demand a text imposes on students.

7.5.1   Implications of key figure results and curricular standards

The analysis features a selection of key figures with various focal points such as lexical sophistication analyzed with the help of the English Vocabulary Profile (EVP). Accordingly, a higher percentage of words categorized as level A1 means that a considerable part of the analyzed text comprised words with lower difficulty; hence, the text was rated as less demanding. With the other two levels, as mentioned in 6.3.1, higher percentage rates indicated a higher difficulty level; thus, the text was rated more demanding. The frequency measures of the BNC served as another aspect to arrive at an understanding of the demands and complexity of the text. Lower rank numbers were associated with a more frequent use of the word leading to a lower CEFR level. At the same time, higher rank numbers meant a higher CEFR level because these words are used less frequently and may be interpreted as more specialized, expert vocabulary. Furthermore, the analysis also looked at the number of words ranking above the word in the median position. The ratio of words above the detected CEFR level hinted at possibly increased complexity. In cases which showed a considerable percentage rate outside or above the initial ranking and assessment, the conclusion that this text was possibly comprised of an increased number of complex words was likely. With the comparative analysis of the text and the Academic Word Lists (AWL), it was possible to determine the share of the words used in the text also featured on one of those lists. The lexical diversity measures were aligned with the MTLD scale by Malvern et al. to compare the texts' complexity with the required skill level to the values detected in distinct age groups and groups of learners, respectively. Additionally, comparing the results of the analysis of the English textbooks (see chapter 6.4.1) ensured that the computed MTLD values and their corresponding position on the Malvern scale convey the appropriate assessment. It has to be kept in mind, though, that this measurement only informs about the quantity of different words used in a text, not their quality. Another indicator from the analysis was the density of both nominal and verbal elements per sentence. The higher the amount of both elements per sentence, the higher the density of information within a sentence, and the higher the complexity of a single sentence and the text as a whole. The Flesch-Reading- Ease Index reflected upon the readability of texts determining how easy or challenging an individual text is to read and understand. The Flesch-Reading-Ease Index combines the key figures described above into one measure, which in turn was aligned with a slightly adjusted scale for the German context (see Table 6.3 - Adjusted scale of grades in comparison to original scale of grades for Flesch-Reading-Ease). Lastly, exemplary grammatical aspects employed in the texts were analyzed with the help of the English Grammar Profile to allocate a precise CEFR level for the individual phenomena. Based on the English

curriculum by Rhineland-Palatinate (version of 2000), those findings were reconciled with these guidelines to assess the demand those grammatical aspects impose on students.

All results and their corresponding CEFR levels are to be aligned with the standards of language competence set by the curriculum for English and the standards for the German Abitur as outlined in "Verwaltungsvorschrift des Ministeriums für Bildung, Wissenschaft, Weiterbildung und Kultur vom 14. Januar 2014 (9413 C / 944 A – 51 410/34/35)" (Ministerium für Bildung, Wissenschaft, Weiterbildung und Kultur 2014b). This administrative directive states that by the end of year 10, students' language competences should fulfil the requirements stated with CEFR levels B1 with shares of B2 for bilingual tracks. Deducting a scale of CEFR levels per grade from this, students' competences obtained within those various grades should arrive at the following iteration:

Grade	Level
6	A2
7	A2/B1
8	B1
9	B1
10	B1/B2

Table 7.15 - Targeted CEFR levels per grade deducted from the curriculum and the curricular standards KMK

The fact that four grades are at level B1 reflects upon the extensive language development in those grades. Furthermore, it justifies the point values or rather their span assigned to the levels in the B-band (see chapter 6.4.2). The analysis of grammatical phenomena taught in the respective grades according to the curriculum (cf. Ministerium für Bildung, Wissenschaft und Weiterbildung 2000) further supports this finding with the help of the English Grammar Profile (EGP online). The easiest way to confirm this scale is to look at grammatical phenomena and their time of introduction in English classes. The past perfect e.g., is introduced in grades 7 or 8 (cf. Ministerium für Bildung, Wissenschaft und Weiterbildung 2000: 91) and is assessed in its basic use to be level B1 according to EGP online (cf. Cambridge UP 2015: sec. PASSIVE). The simple past is usually introduced in grade 6 and is assessed as level A2 (ibid.: sec. PAST). Hence, the scale deducted is appropriate, judging from grammatical aspects.

To gather the full meaning of the computed key figures per subject and their allocated CEFR levels, these were compared to the standards set by the curriculum and the administrative directive. Additionally, illustrations support comparing the analyses' results to the curricular standards for each chapter per subject. Bar charts help to understand the measured and assessed values and CEFR levels per key figure category. Added to that is the zone of expected skill placed as set by the curricular standards. Color-coded bars further help to understand the extent of the required skills and the possible mismatch with the actual expected level.

Before coming to a conclusion based on the comparisons, the idea behind Vygotsky's *Zone of Proximal Development* also has to be respected when assessing key figures, subchapters and the assigned CEFR levels. The Zone of Proximal Development has to be considered as the possible level of achievement with sufficient and skilled help (cf. Gauvain 2020). That is to say, if the targeted standard set is at CEFR level B1 while the analyzed material features a key figure at B2, it could still be within reach if students receive support and scaffolding from skilled and specially trained teachers. Solely the fact that a key figure or whole subchapter was a level above the targeted level of the learning group did not automatically imply that this key figure or subchapter is as overwhelming. The discussion of the implication of the individual result as per subject considered all of the above.

7.5.2   History

The results obtained from the analysis of *Exploring History 1*, chapter 1 indicated that the chapter and its texts were difficult to extremely difficult within the key figure categories. The key figures showed that most assigned CEFR levels were between B2 and D1 for the three EVP bands. These findings were based on the low percentage rates of A1-rated types and the comparingly high percentage rates of B2- and C1-rated types within those texts. Accordingly, concerning the lexical sophistication based on the English Vocabulary Profile, this chapter is most likely overtaxing for students in the lower age group between 12 and 13 years or in grade 8. Further, the computed positions of the median words within the BNC and the corresponding CEFR levels showed that the vocabulary used is mostly of higher frequency, rated between A2+ or B1, B1+ for most parts of chapter 1. Those ratings would imply an overall lower complexity for vocabulary. Nevertheless, at a second glance, these lower levels of complexity cannot be regarded as the final assessment of difficulty for the frequency measurement, as the majority of all words within the subchapters are below the computed rank of the median word, indicating that the said majority employed is used less frequently and thus, to be assessed as more demanding. The required levels of language competence determined by comparing the texts to the Academic Word Lists implied that the chapter at hand is quite challenging in this context.

With the majority of CEFR levels being either B2 or higher, this exceeds the expected receptive competences in the targeted age group of this textbook. The information density of the chapter's texts was mostly target-group-appropriate, with both nominal and verbal elements' CEFR level at B1. Lastly, the results for the Flesch-Reading-Ease rated as level C1 implied that the texts employed exceed students' competences and, thus, are most likely overtaxing if left as is without scaffolding and further assistance. Likewise, the computed average MTLD-value of 86.51 showed that this chapter's lexical diversity exceeds that of students in grade 7 or 8 since the texts were allocated on the scale to be of academic quality. Furthermore, looking at the MTLD-values computed for the English textbooks of the corresponding grades, the value for this History text was well above the values for the English texts.

Those ranged from 65 to 70; thus, the computed value for History material is too demanding. In an attempt to illustrate the skill mismatch, Figure 7.1 below depicts the computed values per text with maximum value for *English G21, A3/A4* (green) in contrast to the value determined for *Exploring History 1*, chapter 1 (pink):

Textbook					
English G21, A3/A4 (range)					
Exploring History1, Chapter 1					
MTLD-value	50 – 60	60 – 70	70 – 80	80 – 90	90 – 100

Figure 7.1 - MTLD-values compared for English G21, A3/A4 and Exploring History 1, chapter 1

Illustrating the findings and curricular standards together in one graphical overview helps to understand the assessment of this chapter as being too demanding and potentially overtaxing for students. This overtaxing bears the danger of either decreasing motivation for students when permanently confronted with too complex texts or missing the content-related learning goals when said content is too complex language-wise to be understood and processed.

To further the understanding of the impact of the average CEFR levels calculated per key figure based on the point values outlined in chapter 6.4.2, these are visualized in an comprehensive overview (see Figure 7.2). It displays the allocated and averaged CEFR levels per category and contrasts those with the curricular standards and the targeted CEFR level for the distinct grades as a bar chart. The bars use color-coding: orange for CEFR level C2, yellow indicates C1, while blue bars represent B2 and B1. Additionally, green bars stand for levels A1 and A2. The lighter coloring for BNC 50 was selected to imply that the majority of words used in this chapter ranked as being less frequent and, consequently, more complex words. The curricular standards or, moreover, the targeted CEFR level for the grade this textbook is designed for is added as a red zone to this overview to illustrate the extent to which the analyzed material in the distinct categories is too complex for students. It respects the aspect that bilingually taught students display advanced language skills (cf. Klieme et al. 2006: 60).

	A1	A2	B1	B2	C1	C2
	+	+	+	+	+	+
EVP A1						
EVP B2						
EVP C1						
BNC 50						
AWL						
Nominal Elements						
Verbal Elements						
Flesch-Reading-Ease						
	A1	A2	B1	B2	C1	C2
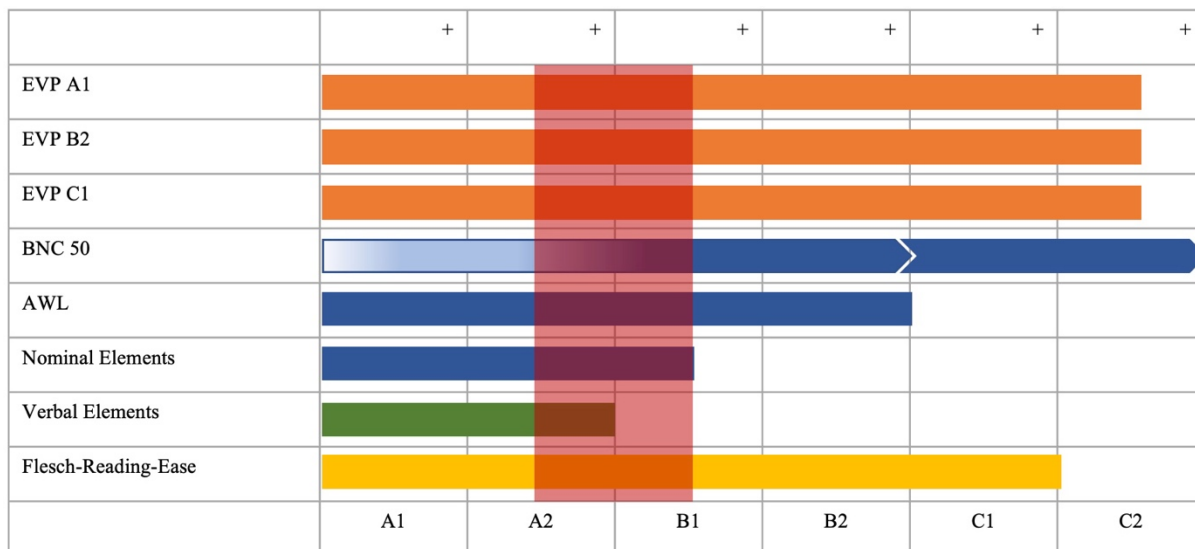
Figure 7.2 - Ratings per key factor category for Exploring History, chapter 1, compared to curricular standards

The overall assumption of chapter 1 being too demanding is further documented by the results per subchapter, either texts or sources. Looking at the overall CEFR levels as per subchapter (see Table 7.1) across all key figures, the ratings implied that these were above the targeted CEFR level for grades 7 and/or 8, too, with levels starting at B1+. It has to be pointed out, however, that the assumption that sources are per se more complex than the texts as such did not hold true for chapter 1. CEFR levels varied here independently of the quality or kind of text. This finding allowed for the assumption that the complexity here is mainly rooted in the cognitive aspect of language, i.e., the decoding of meaning transported with language. Contrasting the overall CEFR levels assigned per subchapter text and sources to the level targeted by the curriculum results in the following bar chart:

Subchapter	A1	A2	B1	B2	C1	C2
	+	+	+	+	+	+
1.1 Text						
1.1 Source						
1.2 Text						
1.2 Source						
1.3 Text						
1.3 Source						
1.4 Text						
1.4 Source						
	A1	A2	B1	B2	C1	C2
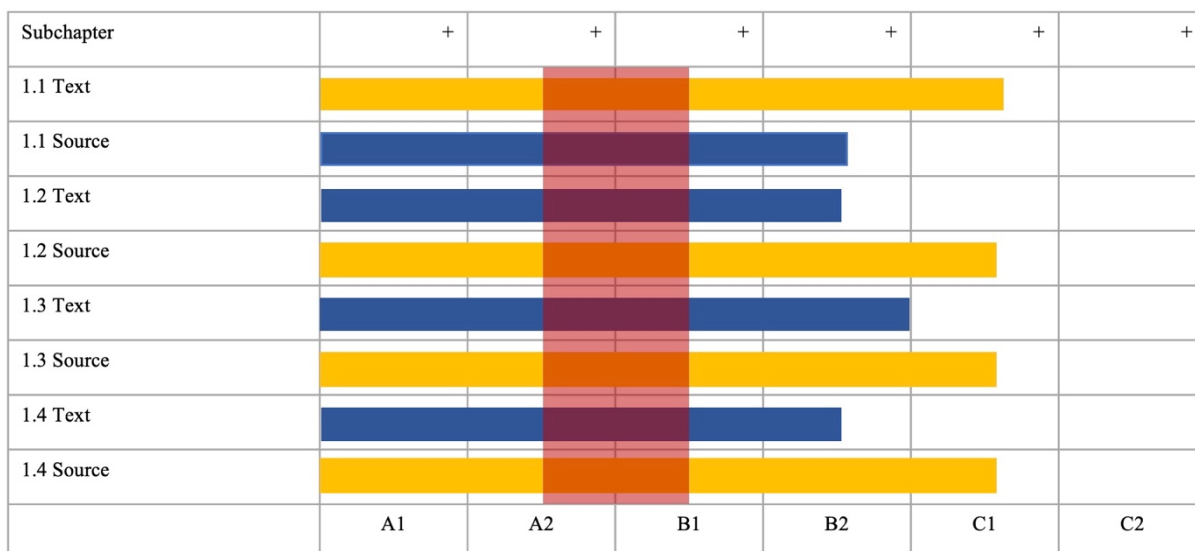
Figure 7.3 - Overall CEFR ratings per subchapter Exploring History 1, chapter 1 compared to curricular standards

With the red-colored zone marking the targeted CEFR level for grades 7 and 8 at the threshold of level B1, the overall CEFR level as given by *Text Inspector* per subchapter are clearly above that level. Additionally, the average computed from this overall assessment of the texts selected (see bottom line

of Table 7.1) is at level B2+. Although the extreme demand of single subchapters is levelled out, the required level of skills is still clearly above the targeted level. This discrepancy further attests to the interpretation of the key figure results of chapter 1 being too demanding.

The analysis of *Exploring History 1*, chapter 4 showed that the overall demand decreased from chapter 1 to chapter 4, although an increase could have been possible since this would have taken students' progress into account. However, looking at the results computed for the various categories, chapter 4 and its subchapters divided into texts and sources displayed CEFR levels that still impose a too high demand on students. This bears the danger of asking too much of students in the targeted age group. For the categories concerned with the English Vocabulary Profile (EVP), ratings across all three categories (level A1, B2, and C1) spanned from A2+, which was only assigned once, to the majority of ratings at C2 and even C2+. The assessments of the median positions for the distinct portions of the material (text or source) for the frequency in the BNC implied that half of the required levels meet students' possible competences while the other half is above the level students should achieve or potentially have achieved. While this did not seem too demanding, again, the percentage number of words or, much rather, their frequency counts that were below this median position and thus, used far less frequently and being more complex was at 90%. This indicates that most words were more complex than the level assigned to the average median being B2. This level was already above the expected competence level of B1 in grades 7 or 8. Continuing with the results for the Academic Word Lists, only 6% of the words used in the text were found on those lists. Presumably, the majority of words was not featured here. However, this should not lead to the conclusion that the texts were less challenging. The identified 6% already caused a rating at level B2, above of what can be expected. The key figures for propositional density were at a surprisingly low CEFR level. Both shares for verbal and nominal elements were at level A2+, which is entirely appropriate for the target group. Lower ratings of propositional density point at the conveyed information is embedded in a comparably large amount of text. This, in turn, decreases complexity. A look at the subchapters of chapter 4 with low CEFR levels for propositional density reveals that nouns and verbs are used repetitively within the context of Industrialization. This repetition leads to the conclusion that although the subchapter deals with various aspects of Industrialization, the choice of unique semantic units is limited but still entirely sufficient to transport the information. The single subchapter rated at level D1 features an excerpt of a translation of the papal encyclic *Rerum Novarum* (cf. Kröger & Lohmann 2007: 62) as a source dense with nominal elements. In contrast to the assessment of propositional density, the readability was at level C1 with an average Flesch-Reading-Ease index value of 55.89. With this value being allocated on the adjusted scale to be appropriate for grades 10-13, this key figure again shows that the chapter's texts and sources are too demanding to be worked with in grades 7 or 8.

	A1	A2	B1	B2	C1	C2
	+	+	+	+	+	+
EVP A1						
EVP B2						
EVP C1						
BNC 50						
AWL						
Nominal Elements						
Verbal Elements						
Flesch-Reading-Ease						
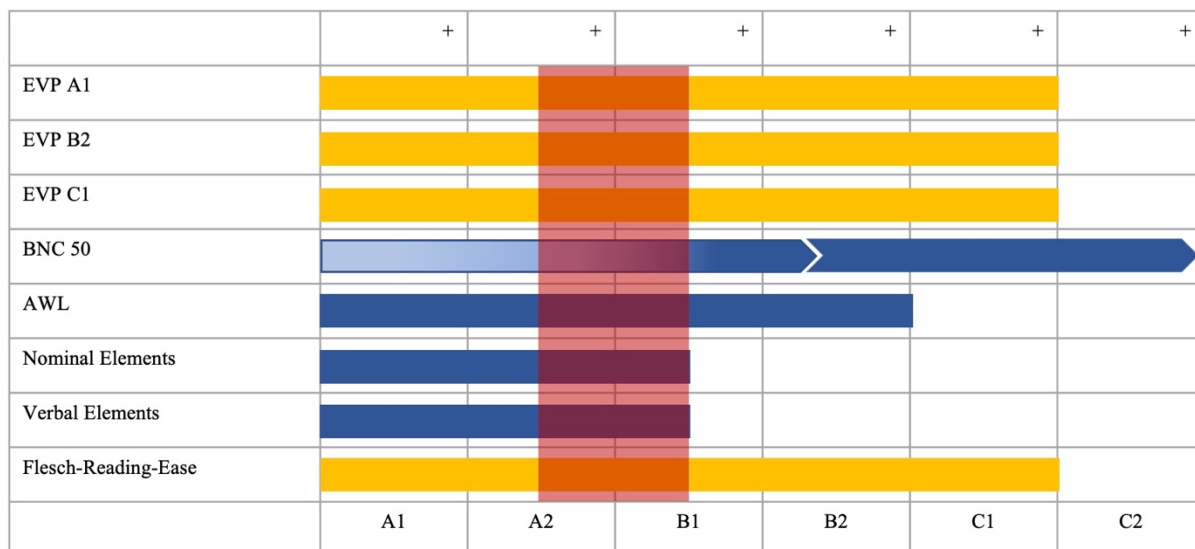
Figure 7.4 - Ratings per key factor category Exploring History 1, chapter 4 compared to curricular standards

All key figures and their averages but those for propositional density were demanding too much of a potential target group of students in grade 7 and 8. Even propositional density figures demanded a CEFR level targeted to be achieved earliest towards the end of year 8 in a bilingual track (see Table 7.15).

Additionally, the average MTLD value of 97.51 further supports this conclusion of exceeding demand. The average MTLD clearly places the chapter at an academic level and once again proves that the texts potentially ask too much of students aged 12 to 13, especially when the reference value based on the English textbooks ranges between 65 and 70.

Textbook					
English G21, A3/A4 (range)					
Exploring History 1, Chapter 4					
MTLD-value	50-60	60-70	70-80	80-90	90-100
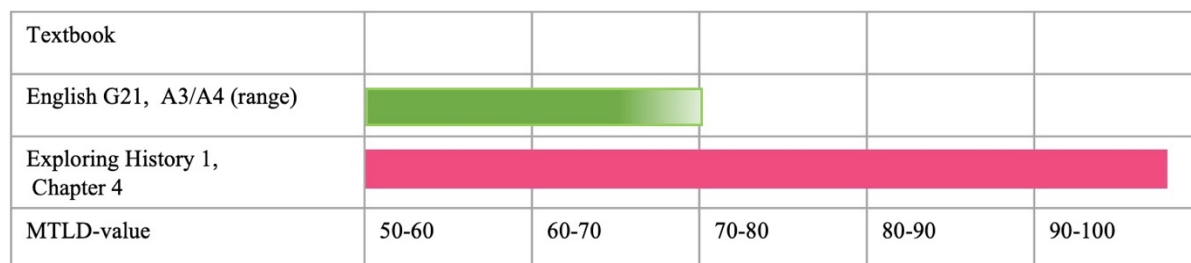
Figure 7.5 - MTLD-values compared for English G21, A3/A4 and Exploring History 1, chapter 4

The overview of the overall ratings per subchapter below illustrates that chapter 4 is slightly less demanding than chapter 2. However, with the potential level of competences in the targeted grades represented by the red-colored zone, it is obvious that despite the decrease in demand, the requirements still exceed students' assumed competences. Within chapter 4, only three subchapters' overall CEFR level was appropriate for the assumed slightly progressed level targeted within the curriculum. The other subchapters, both texts and sources, were again exceeding level B1 with the highest level assigned being C2+. The red-colored zone illustrates this skill mismatch, which could cause distinct problems in both language and content learning. On average, this chapter's overall CEFR level was B2, which is again a full level beyond the maximum targeted level of language competence. Referring back to Vygotski's *Zone of Proximal Development* mentioned in 7.5.1, this assessment of the subchapters, the single key

figures, the overall CEFR level, would not fall under what is intended with the ZPD. While single, well-considered challenges foster learning, continuous and extensive overtaxing most likely causes the opposite effect: discouragement to take up the challenge and thus, hindering the learning process.

Subchapter	+	+	+	+	+	+
4.1 Text						
4.1 Source						
4.2 Text						
4.2 Source						
4.3 Text						
4.3 Source						
4.4 Text						
4.4 Source						
4.5 Text						
4.5 Source						
4.6 Text						
4.6 Source						
4.7 Text						
4.7 Source						
	A1	A2	B1	B2	C1	C2
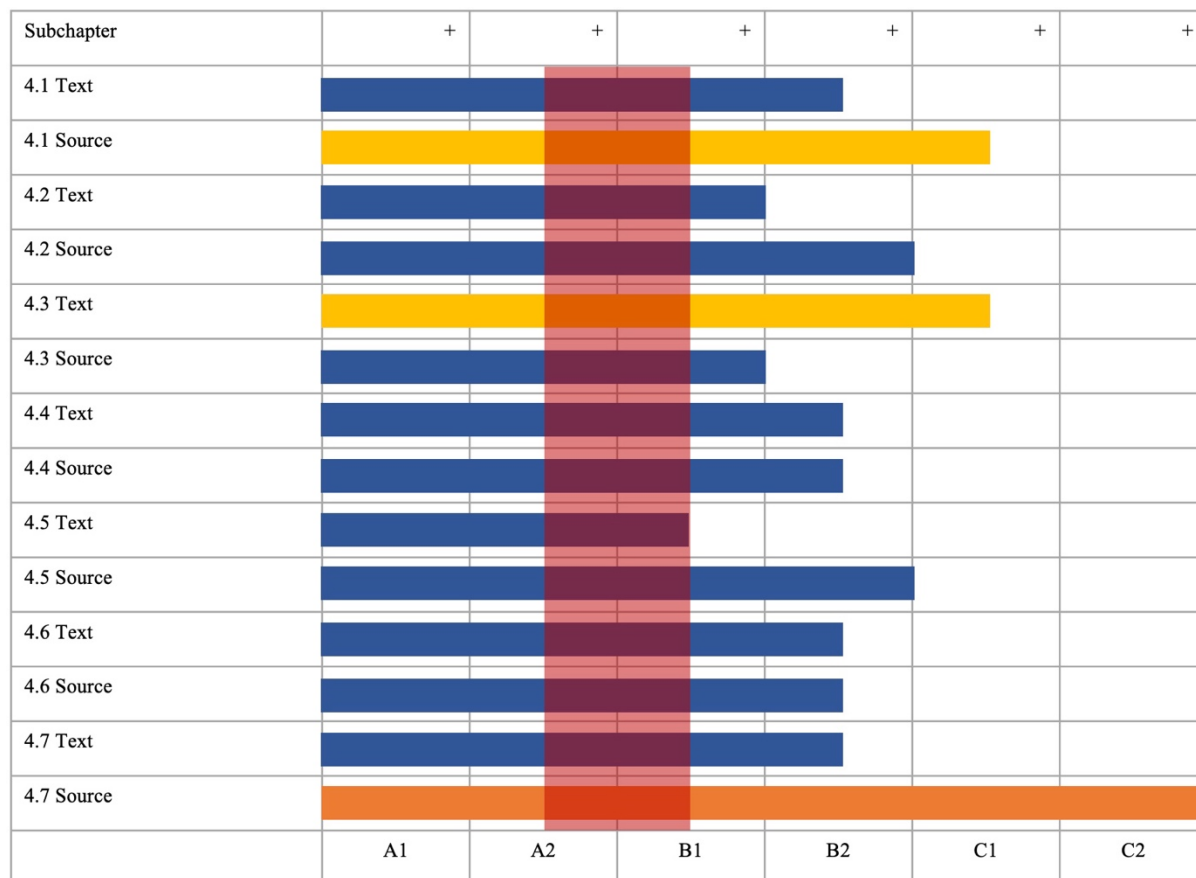
Figure 7.6 - Overall CEFR ratings per subchapter Exploring History 1, chapter 4 compared to curricular standards

For *Exploring History 2*, chapters 3 and 5, increasing demand was expected before analyzing the material since this textbook is issued for teaching in grades 9 and 10 with students aged 15 to 16. At that time, students will have had at least five years of instruction in English and thus, should be able to work with more demanding texts. The curricular standards still set the CEFR level to be expected for grade 9 at B1, while in grade 10 the transition onto level B2 is projected. It can be assumed that students in bilingual tracks have solid competences within the B1-band and may also have proceeded further to display competences placed in the B2-band. Accordingly, the shift of the red-colored zone in the overview in Figure 7.7, Figure 7.9, Figure 7.10, and Figure 7.12 reflects this progress.

The analysis of chapter 3 revealed that although students are supposedly advanced, the texts still exceeded the potential competences (see Figure 7.7). Correspondingly, the English Vocabulary Profile assessments placed all three categories within the C-band. The comparably low number of A1-rated words, making a slim third of the texts, led to a rating of C1+, which is clearly above students' capabilities. Likewise, the other two EVP bands exceed the level that can be expected of students at the beginning of year 9, with EVP B2 assessed to be at C2+ and the relatively low percentage of C2-

categorized words being rated C2+ as well. So, from the vocabulary perspective, this chapter is too demanding or requires overly extensive scaffolding and vocabulary work.

Compared to the chapters analyzed and discussed beforehand, the median position for *Exploring History 2*, chapter 3, was lower on the frequency scale, meaning increased complexity. This increased complexity ties in with the findings yielded from the EVP analysis, and the CEFR level assigned being C1 supports this finding. The fact that approximately 90% of all words employed in this chapter were more complex, i.e., below the median position, further attests to the assumption that this chapter is too demanding, requiring too high levels of vocabulary knowledge and lexical sophistication for the targeted learner group. Correspondingly, the results for the comparison of the texts and the Academic Word Lists revealed that with the 10% share to be found on those lists, the required CEFR level would be C2, which again is well above the level students can be expected to be at. The propositional density measures were again comparably lower, rated in the B-band but slightly surpassed the assumable curricular standards in grades 9 and 10. Were it not for the vocabulary, the texts – assessed for this key figure alone – would be appropriate with teacher support. Although the packaging of information appears fitting, the Flesch-Reading-Ease index again implies that the texts are too demanding. The computed values arrived at a CEFR level of C2+ with a score corresponding to readings that can be expected at university level but certainly not at the level the textbook was designed for. The grammatical aspects, however, were all within the range to be expected in grades 9 and 10. In summary, all but two key figures imply required competences and capabilities too demanding for students in the targeted age group of 15 – 16 years. The overview below clearly illustrates this regardless of the slight progress in language depicted by the shift of the red-colored zone:

		+	+	+	+	+	+
EVP A1							
EVP B2							
EVP C1							
BNC 50							
AWL							
Nominal Elements							
Verbal Elements							
Flesch-Reading-Ease							
		A1	A2	B1	B2	C1	C2
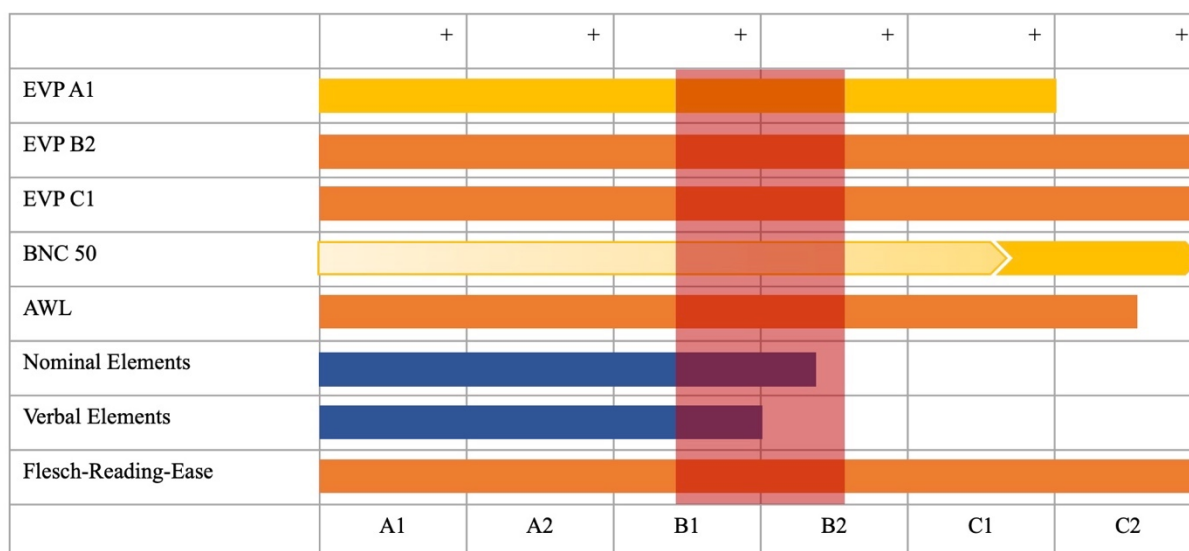
Figure 7.7 - Ratings per key figure category Exploring History 2, chapter 3, compared to curricular standards

The lexical diversity measure MTLD placed the texts at an academic level and thus most likely beyond the scope of teaching in school. With a computed average MTLD-value of 91.42, the value was still above the range determined for English textbooks for the age group of 15 to 16-year-old students,

which was between 81 to 89 for grade 9 (see Figure 7.8). As this chapter is most likely taught in grade 9, the comparison of the MTLD-values of the English and History textbook analysis illustrates again a slight overtaxing:

Textbook					
English G21, A5 (range)					
Exploring History 2, Chapter 3					
MTLD-value	50-60	60-70	70-80	80-90	90-100

Figure 7.8 - MTLD-values compared for English G21, A5 and Exploring History 2, chapter 3

The overall CEFR levels computed per subchapter – divided into texts and sources again – confirms the stance on this chapter displaying an apparent skill mismatch between the required and level of skills presumed available. Contrary to the initial assumption, the sources in this subchapter were rated as less complex language-wise compared to the texts provided by the publisher. Still, both categories were rated as being above level B1 and also B2, since the overall assessments by *Text Inspector* averaged out at level C1+.

Subchapter	+	+	+	+	+	+
3.1 Text						
3.1 Source						
3.2 Text						
3.2 Source						
3.3 Text						
3.3 Source						
3.4 Text						
3.4 Source						
3.5 Text						
3.5 Source						
3.6 Text						
3.6 Source						
	A1	A2	B1	B2	C1	C2
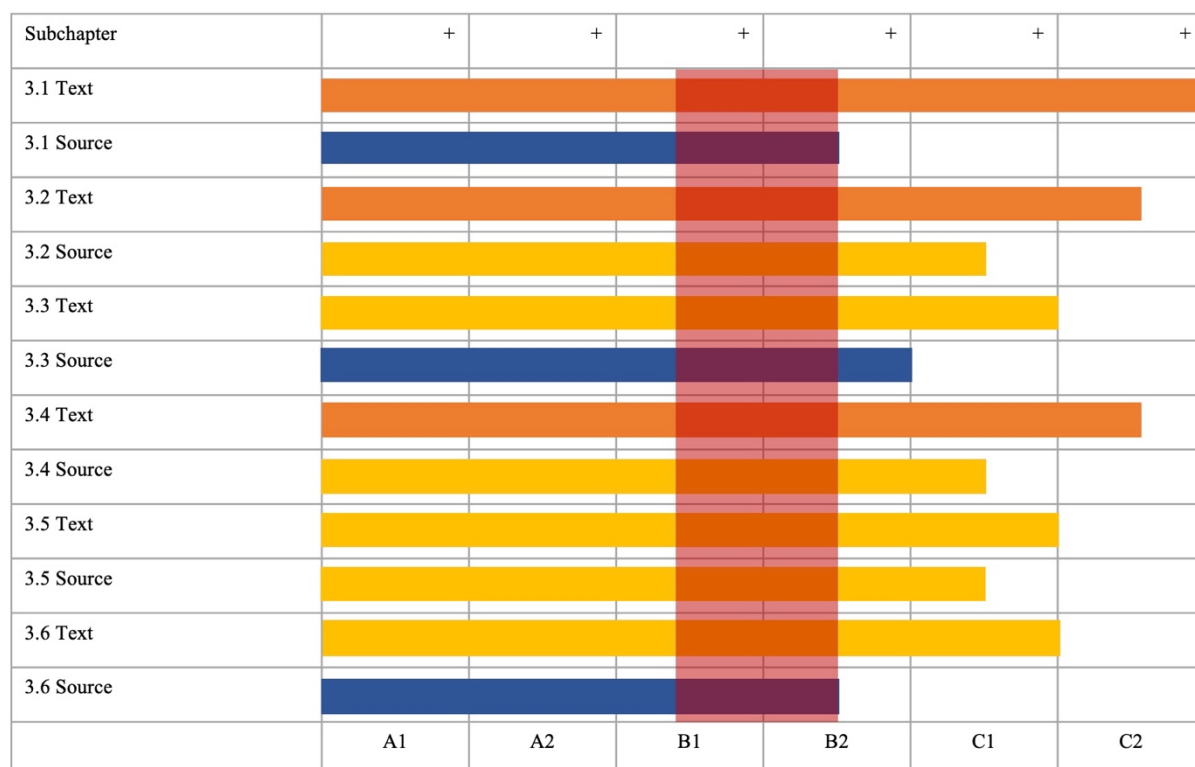
Figure 7.9 - Overall CEFR ratings per subchapter Exploring History 2, chapter 3 compared to curricular standards

Concluding with the fifth chapter of *Exploring History 2*, a chapter presumably taught towards the middle or end of year 10, progress in students' language skills can be assumed from B1 towards a transition of B1 to B2, with bilingual students most likely displaying a considerable number of skills

within the B2-band. The featured texts required increased CEFR levels for lexical sophistication based on the findings for the English Vocabulary Profile. For EVP A1, with a third of the types rated to be within that band, level C2 was assigned. The same applies to the other two categories within the EVP analysis; both the B2-band as well as the C1-band were rated as C1. Although those three ratings are still above the targeted CEFR level B2, students may progress towards the gap between what students may be able to understand, on a vocabulary basis, and what the texts require them to understand. With this the degree of potential overtaxing may decrease. With appropriate scaffolding, students may even be able to work with this chapter, given that the progress towards CEFR level B2 is genuine. By examining the frequency measures based on BNC and the derived level of complexity, the calculated median position for chapter 5 was determined to be CEFR level C1+ and C1, respectively. This placement exceeds the maximum curricular standard set for the targeted learning group by one level. However, as was the case for the other three chapters and their BNC measurement, the percentage of types used less frequently and thus ranked below this median position, implying higher complexity, was at approximately 95%. That is to say that almost all types were more demanding than the word in the median position. This high percentage implies that the chapter uses demanding vocabulary, potentially overtaxing students regardless of their advanced language skills. The comparison with the Academic Word Lists revealed a slight increase or overtaxing of demand rated at B2+. Regarding information or propositional density, this chapter's demand was entirely in line with the assumed competences of students 15 – 16 years of age. The CEFR level determined for the density of nominal elements and verbal elements was B1+. The Flesch-Reading-Ease index required students again to display high level reading skills at C1+, thus being too demanding for the intended target group.

	+	+	+	+	+	+
EVP A1						
EVP B2						
EVP C1						
BNC						
AWL						
Nominal Elements						
Verbal Elements						
Flesch-Reading-Ease						
	A1	A2	B1	B2	C1	C2
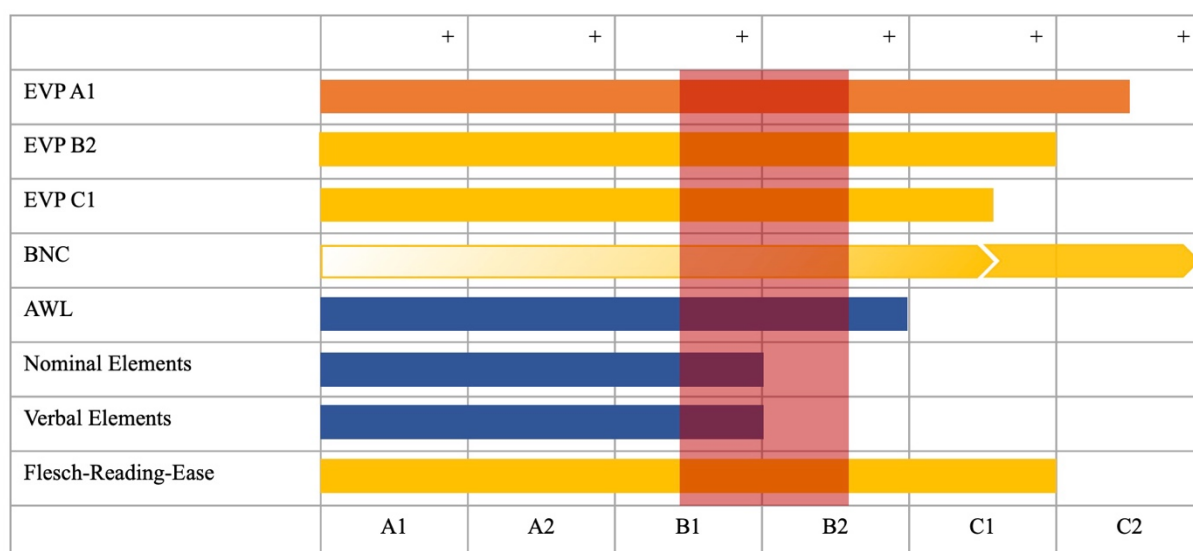
Figure 7.10 - Ratings per key factor category Exploring History 2, chapter 5, compared to curricular standards

The average MTLD value of 81.09 clearly indicated that the texts display a diversity to also be found in texts used in the university context. This could lead to the conclusion that the texts in this chapter are too demanding. However, compared to the computed MTLD-values for *English G21, A6,*

this value can be considered appropriate. *English G21, A6* is the textbook used in grade 10 (see Table 6.2), which corresponds to the targeted group of learners *for Exploring History 2*, chapter 5 since both address students in grade 10. Despite the high diversity detected in the History material, the range for *English G21, A6* from 98 – 99 is above the computed average value for *Exploring History 2, chapter 5*.

Textbook					
English A6 (range)					
Exploring History 2, Chapter 5					
MTLD-value	50-60	60-70	70-80	80-90	90-100

Figure 7.11 - MTLD-values compared for English G21, A6 and Exploring History 2, chapter 5

When having arrived at grade 10, the most common grammatical aspects have been dealt with in English teaching (cf. Ministerium für Bildung, Wissenschaft und Weiterbildung 2000: 91f.), and accordingly, the phenomena featured in the texts were adequate for students.

Still, despite the advancement of language skills towards level B2, the overall CEFR ratings for the subchapters are unaltered above the level expectable following the curricular standards (see Figure 7.12). The targeted group of students for this specific subchapter most likely has reached the end of the B1-band or the beginning stages of the B2-band by the time this chapter is potentially taught. Subchapters requiring levels B2+ to C2+ could jeopardize successful participation in those classes. Similarly, the CEFR level required on average is too demanding at level B2+. The following figure illustrates the findings as well as their assessment:

Subchapter	+	+	+	+	+	+
5.1 Text						
5.1 Source						
5.2 Text						
5.2 Source						
5.3 Text						
5.3 Source						
5.4 Text						
5.4 Source						
5.5 Text						
	A1	A2	B1	B2	C1	C2

Figure 7.12 - Overall CEFR ratings per subchapter Exploring History 2, chapter 5 compared to curricular standards

Although students undoubtedly develop their language skills and progress into level B2, this chapter was again too demanding. Other than the above-discussed chapters, this chapter's content may be

manageable with meaningful and targeted support by a skilled teacher. This potential ability to handle the texts has less to do with a possible decrease in demand caused by less complex material but rather with the level of language skills targeted in students of that age group. Those findings support the assumption that bilingual History is probably better taught at higher grades.

The analysis of all four chapters revealed that each chapter is too demanding, even those of *Exploring History 2*. Surprisingly, the chapters do not seem to consider the progress as ascribed by the curriculum. Neither is a development or increase in complexity as per the progress in grades and competences traceable. Most CEFR levels assigned to the key figures exceed, at times by far, the levels that students are supposedly able to master at the given grade or age. These findings allow for three possible conclusions:

1. The textbooks were designed with the intention to achieve a high level of text quality leading to complex texts. This intention in itself is not problematic as long as the textbooks provide sufficient scaffolding, which unfortunately is not the case with *Exploring History* as only select vocabulary is scaffolded, mostly as translations to German.

2. The textbooks seem to not being designed with the targeted groups of learners in mind. Likewise, editors seem to not be considering the fact that History can be challenging already in the original language of schooling, i.e., German (cf. Lamsfuß-Schenk 2015).

3. With the textbook material as it is, History as a subject is likely to be even more overtaxing than already in German. This may lead to a loss of interest and motivation in students and possible setbacks regarding knowledge and competence gain.

The assessment made in the European Center of Modern Languages study, placing the linguistic demand for History at CEFR level B1/B2 (cf. Moe et al. 2015) could – at least for this textbook and its reception – not be confirmed. Consequently, it seems as if History may not live up to its characterization as the prototypical bilingual subject to be taught in early stages already (cf. Wunderlich 2018). At least not with the material analyzed and teachers without training in language teaching, or preferably a specialized training in CLIL teaching.

7.5.3 Geography

With Geography being slightly less text-dependent than History, the required level of complexity is almost equally decisive when it comes to successful participation in bilingual Geography lessons. Consequently, the textbook material should only– in the sense of the Zone of Proximal Development – be marginally beyond the level of skills students have achieved or are expected to have achieved at the time of teaching. As described in chapter 7.2, the analysis results show that the textbook material of *Diercke Geography Volume 1*, chapter 1 exceeds the targeted level of A2/B1 in grade 7 and B1 in grade 8. Starting with the individual key figures, the lexical sophistication displayed in the textbook material was too demanding, requiring level C2 for the share of words categorized to be within the EVP A1-band. The same applied to the EVP B2-band, which still required vocabulary knowledge with an average

CEFR level of C1+, while the share of C1 categorized words was comparably lower at levels B1+ and B2. However, most words analyzed for their requirement level were above the targeted CEFR level for students at that stage. The same can be interpreted in the results obtained from the analysis of frequency counts. The computed average median position implied a mid-tier level of use which led to CEFR level B2+ to be assigned. Again, this level and the connected demand would be beyond students' capabilities if left unassisted. What is more, is the fact that approximately 75% of all words were likely to be more complex than the allocated level. With the results for the third key figure for lexical sophistication, i.e., the comparison of the texts to the Academic Word Lists, this surmise is further backed. The average percentage of types featured on those lists implied a CEFR level of C1, hence supporting the texts' stance, or their lexical sophistication, to be too demanding for students aged 12 to 13. The information density applied in these texts was, however, target group appropriate, if not a little too simple, at level A2 and A2+, respectively. The results for the Flesch-Reading-Ease index placed the texts at right CEFR level B2, which at the beginning of grade 7 is too demanding. In conclusion, with the lexical sophistication being overtaxing, this chapter's demand may stand in the way of knowledge gain and learner motivation. The graphical display of results in comparison to the targeted CEFR level shows the mismatch in skills.
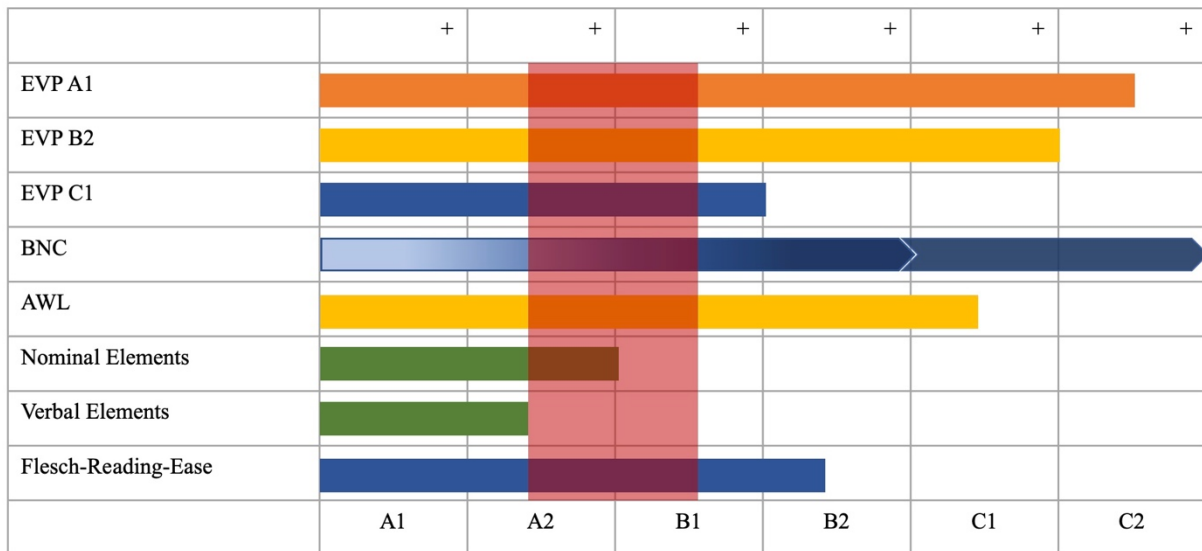
Figure 7.13 - Ratings per key factor category Diercke Geography 1, chapter 1, compared to curricular standards

The key figure of MTLD additionally supports this assessment to measure lexical diversity placing the texts at the beginning of an academic level. The average value for *Diercke Geography Volume 1*, chapter 1 at 72.9 (see chapter 7.2), only slightly exceeds the range of MTLD-value determined for the comparable English texts of *English G21, A3/A4* ranging from 65 – 70.

Textbook					
English A3/A4 (range)					
Diercke Geography 1, chapter 1					
MTLD-value	50-60	60-70	70-80	80-90	90-100

Figure 7.14 - MTLD-values compared for English G21, A3/A4 and Diercke Geography, Volume 1, chapter 1

This picture does not change decisively when looking at the subchapters as such with their overall CEFR levels. Only one out of five subchapters' CEFR level is proximate to the CEFR level ascribed by the curriculum. For grade 7, the expected level would be A2, with distinct elements transitioning to B1. The bars in Figure 7.15 extend already to B1, which is only expected at grade 8. So, even with the vocabulary help provided in the textbook and incorporated in the analysis (see chapter 6.4), the subchapters are still too demanding for students to deal with especially when teachers are not aware of the level required and thus, do not provide scaffolding or other means of support during lessons. The average of the overall CEFR levels ties in with that assessment at level B2.

Subchapter	+	+	+	+	+	+
1.1						
1.2						
1.3						
1.4						
1.5						
	A1	A2	B1	B2	C1	C2

Figure 7.15 - Overall CEFR ratings per subchapter Diercke Geography Volume 1, chapter 1 compared to curricular standards

The second chapter taken and analyzed from *Diercke Geography Volume 1*, targeting grade 8 rather than grade 7 students, was still too demanding for students concerning the levels required by the applied lexical sophistication. All three EVP bands exceeded level B1, regarded as the level to be maximally expected with the targeted age group. While EVP A1 and B2 required level C1, EVP C1 was slightly less demanding at B2. Similarly, the average median position computed for the frequency counts based on the BNC implied too demanding since it required CEFR level C1. Additionally, the majority of types were even above that level since only 5.43% were assessed to be less complex or more frequently used. Although a single subchapter of chapter 2 was rated at A2+ and thus appropriate if not too simple for the targeted learner group, only two addressed level B1, while all other subchapters

displayed frequency counts and corresponding ratings clearly and by far exceeding level B1 (see Table 7.6). Likewise, the CEFR level assigned to the counts for the Academic Word Lists implied the use of too complex and demanding vocabulary (CEFR level C1+) regardless of the scaffolds provided. The packaging of information in chapter 2 was target-group-appropriate with the average CEFR level for nominal and verbal elements at A2. The Flesch-Reading-Ease index, however, was again above the expected level of B1, namely at C1 and thus, overtaxing. With the overall CEFR levels ranging from B1+ to C2 for chapter 2, *Diercke Geography, Volume 1*, chapter 2 is too demanding, although single aspects were appropriate for the targeted age group. The illustration of all key figures in comparison to the targeted CEFR level supports this assessment:
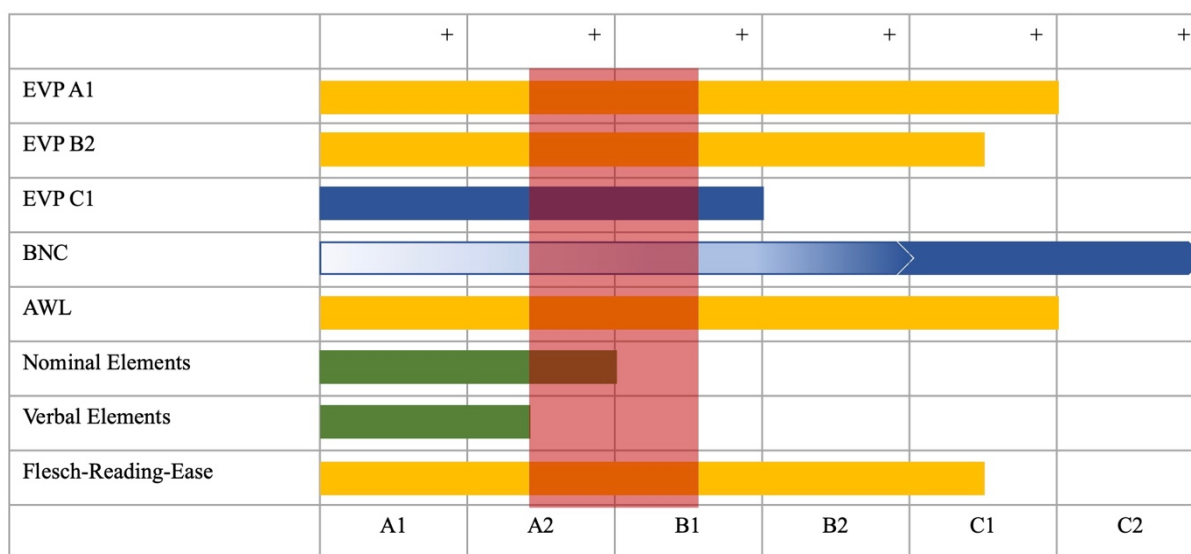
Figure 7.16 - Ratings per key factor category Diercke Geography Volume 1, chapter 2, compared to curricular standards

The average value for lexical diversity, MTLD, however, was within the range found in the equivalent English textbooks *English G21, A3/A4*. With an average MTLD-value of 68.49 for *Diercke Geography,* the underlying diversity and its demand meets the diversity students are supposed to handle, judging from the English textbook.

Figure 7.17 - MTLD-values compared for English G21, A3/A4 and Diercke Geography, Volume 1, chapter 2

Looking at the individual subchapters' CEFR level, it becomes apparent that chapter 2 is still too demanding, although one-third of nine subchapters were almost target group-appropriate at CEFR level

B1+. These subchapters can be mastered if students receive targeted and qualified support by the teacher. The remaining subchapters ranged between levels B2 and C2. B2-rated subchapters could also be manageable for students in year 8 in a bilingual track if the teacher provides sufficient support and language work, and effectively reduces complexity further through, e.g., skilled didactic reduction. This need for support, however, has to be known to and prepared diligently by the teacher. Still, this chapter remains too demanding, which becomes visible by comparing CEFR levels assigned per subchapter and the targeted level given by the curricular standards, as well as by the average CEFR for this subchapter at level B2+.

Subchapter		+	+	+	+	+	+
2.1							
2.2							
2.3							
2.4							
2.5							
2.6							
2.7							
2.8							
2.9							
	A1	A2	B1	B2	C1	C2	
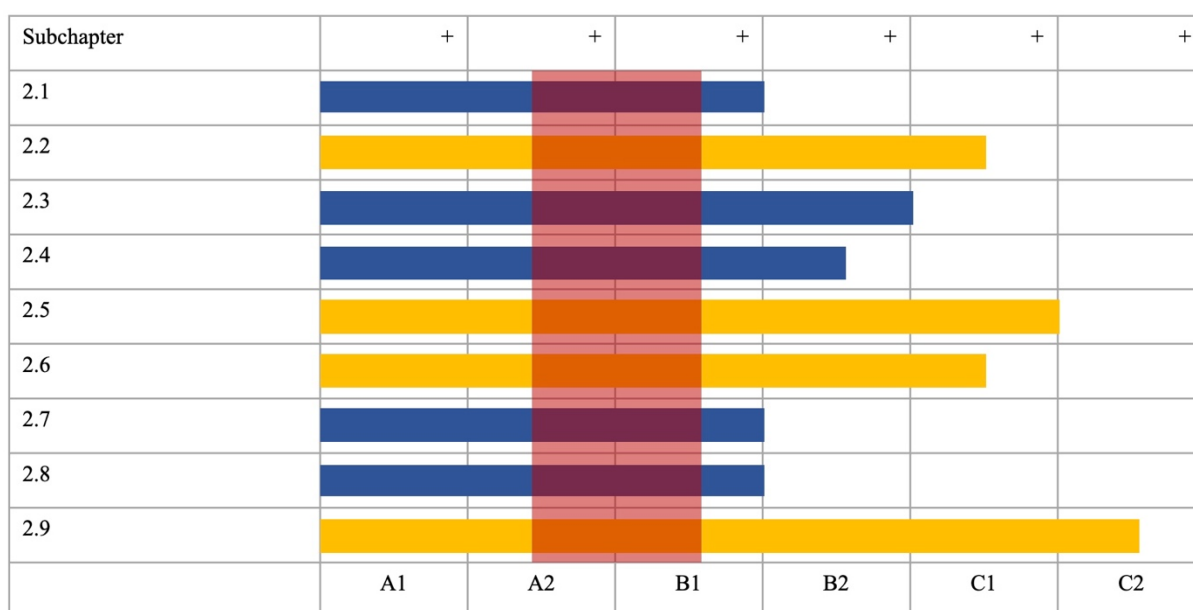
Figure 7.18 - Overall CEFR ratings per subchapter Diercke Geography Volume 1, chapter 2 compared to curricular standards

Diercke Geography Volume 2 is for students in grades 9 and 10. The analyzed chapter 2 is supposed to reflect upon students and their assumed skill level and competences in grade 9. Hence, the value or set CEFR level to compare all results to the curricular standards is an advanced stage within the B1-band illustrated with the moved red-colored zone on the figures showing the findings (see below, e.g., Figure 7.19). The chapter's demands, however, seem to have increased as well since the 2nd volume is for a higher grade. This fact may explain the high CEFR levels required with vocabulary according to the analysis and its results for EVP. Both lower EVP bands, A1 and B2, were assessed at levels C2 or C2+, respectively. Only the C1-band was at C1+. Still, all three are too demanding in grade 9. Likewise, the CEFR level allocated to the average median position at C1+ with a majority of types above that level with more complex words being used less frequently points at exceeding demand. This key figure again shows that the text is overtaxing. Furthermore, the comparably high percentage rate of types found on the Academic Word Lists and its corresponding CEFR level being C2+ indicates that the required lexical sophistication exceeds that of grade 9 students. The texts mentioned in connection with the assessments of *Diercke Geography Volume 1* and their propositional density were mostly rated to be target-group-appropriate. In chapter 2, however, the nominal elements or their average count per

sentence corresponded to the demands met within CEFR level B2, while the verbal elements were at level A2+. So, the density for nominal elements is too demanding. The density for verbal elements was on par with most of the above-discussed ratings for this key figure and, thus, a little too simple for the targeted group. A distinct increase in demand was observable in the Flesch-Reading-Ease index. For the first volume, this figure was only one level above the targeted CEFR level of B1; volume 2, chapter 2, arrived at an average value of 40.58, corresponding to level C2+ which is decisively above the targeted level.
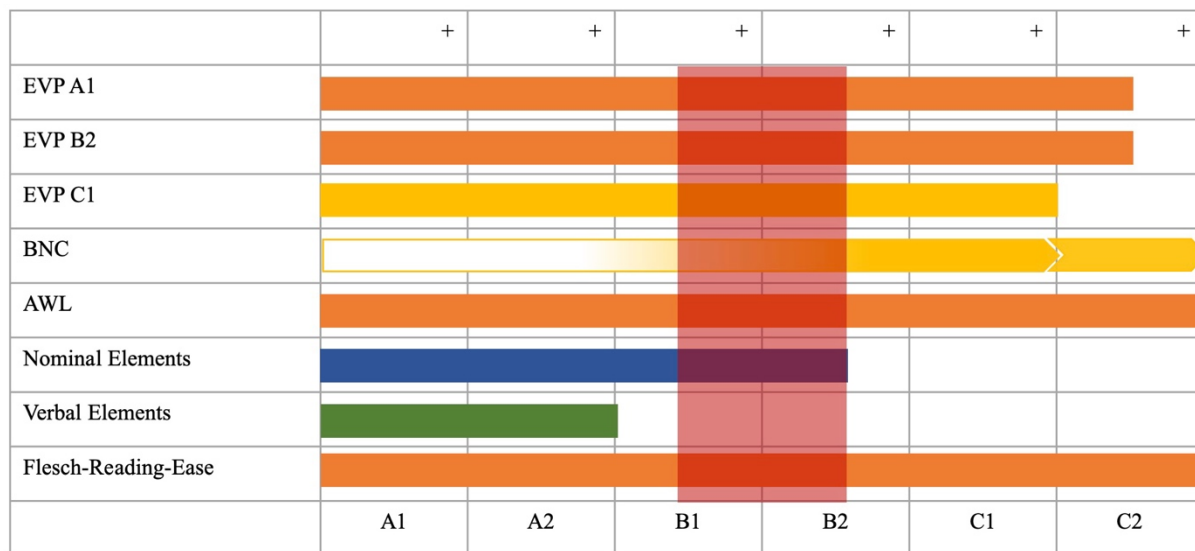
Figure 7.19 - Ratings per key factor category Diercke Geography Volume 2, chapter 2, compared to curricular standards

The measurements of lexical diversity expressed as MTLD imply that the applied diversity within chapter 2 is appropriate for the targeted group of students. The computed value for *Diercke Geography* Volume 2, chapter 2 of 81.72, is within the determined range for MTLD in English G21, A5, which was between 81 and 89 (see chapter 7.2).
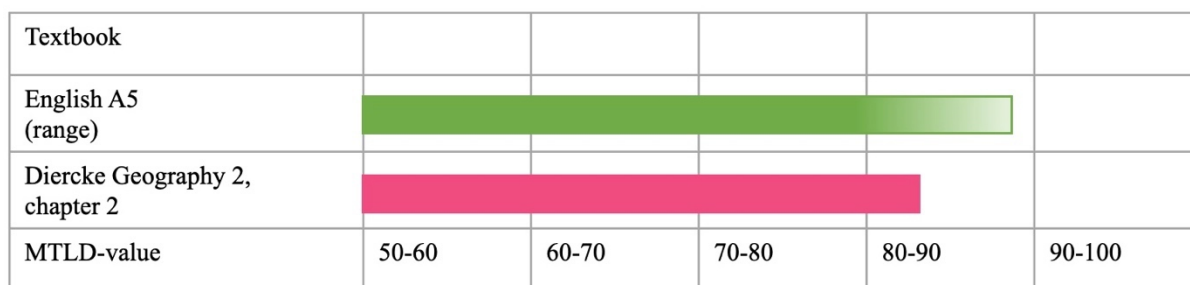
Figure 7.20 - MTLD-values compared for English G21, A5 and Diercke Geography, Volume 2, chapter 2

Figure 7.21 also shows that *Diercke Geography Volume 2*, chapter 2 is too demanding, even if assuming that students of bilingual tracks are more advanced than students in regular English classes. The overall CEFR levels assigned confirm this assessment with the lowest CEFR level assigned for an individual subchapter being B2+. As implied above, this relatively demanding level B2+ could be worked with if

students receive the required amount of language support. Since only two out of eight subchapters were at level B2+, with the remaining six being rated in the C-band, the assessment of the text as overtaxing is further supported. The calculated average of all these individual overall assessments by *Text Inspector* cannot attenuate this conclusion being at level C1+.

Subchapter	+	+	+	+	+	+
2.1						
2.2						
2.3						
2.4						
2.5						
2.6						
2.7						
2.8						
	A1	A2	B1	B2	C1	C2
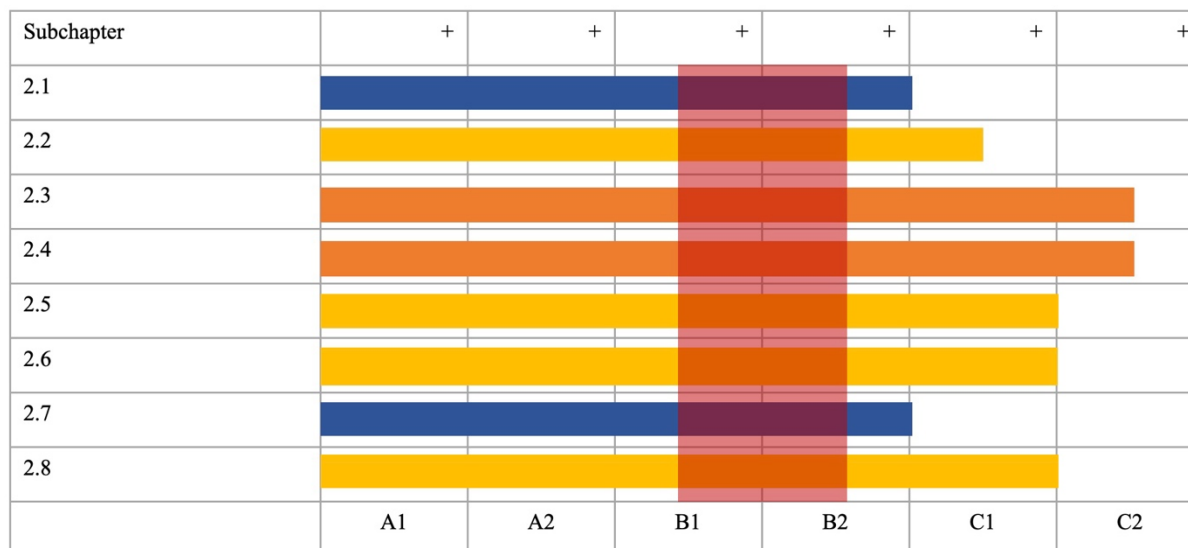
Figure 7.21 - Overall CEFR ratings per subchapter Diercke Geography Volume 2, chapter 2 compared to curricular standards

Per the curriculum, the topic dealt with in chapter 7, *Europe Changes*, is most likely taught in grade 10, accounting for the level of complexity on the content side (cf. Ministerium für Bildung, Wissenschaft, Weiterbildung und Kultur 2021: 22). Accordingly, another increase in language complexity and linguistic demand can be expected. Likewise, students further progress in their language skills and are expected to arrive at least at the end of the B1 band. In parts, competences should develop towards and into the B2 band. The illustration of the results below reflects this potential development by the red zone. The obtained results met the expectation of increased linguistic requirements, and their corresponding CEFR levels were determined for the EVP-related key figures. For EVP A1 and C1, the average level C2+ was calculated, while for EVP B2, the level was even higher and ended up being D1, representing academic vocabulary. Regarding the frequency counts and the computed average median position, the analysis revealed an average assessment at level C1, with the majority of types used less frequently and thus more challenging and potentially to be rated at a higher level. Again, this high level is well above the set expectation of B1 transitioning onto B2 and thus, can be deemed overtaxing. This finding also tied in with the results obtained for the quality of types concerning the Academic Word Lists, of which the share of types calculated implied CEFR level C2 again. Vocabulary or rather the sophistication thereof with the CEFR levels assigned imposes a high demand on students, even in an advanced grade, which applies to grade 10. Concerning propositional density, chapter 7 was in line with chapter 2. While the density of nominal elements per sentence was slightly higher than with verbal elements, the allocated CEFR levels B2 and A2 are within the range of expected competences for students aged 15 to 16 whereas scaffolding with nominal elements might be necessary.

Although this key figure indicated coherence between demand and potential, the Flesch-Reading-Ease index indicated again that chapter 7, just like chapter 2, is too demanding at level C2+ bearing the danger of hurting both student motivation and knowledge gain.
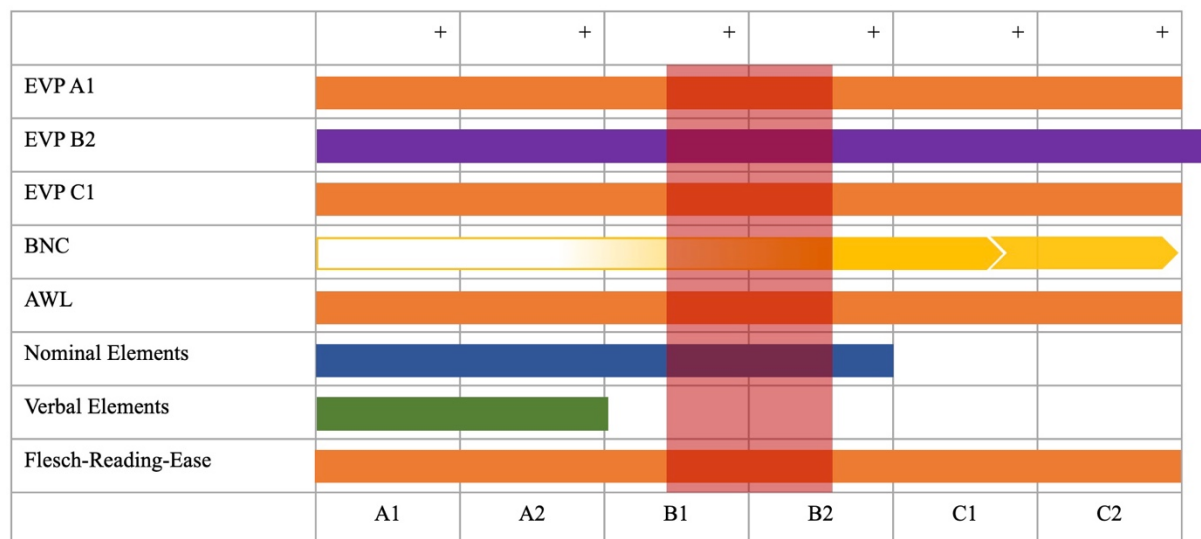
Figure 7.22 - Ratings per key factor category Diercke Geography Volume 2, chapter 7, compared to curricular standards

The lexical diversity, however, was – just like the previous chapters – target-group-appropriate or even below the range indicated by *English G21 A6* (see 7.4.1). With an average MTLD value of 81.3 for chapter 7, the diversity displayed within this chapter can be expected to be mastered by students in grade 10.

Textbook					
English G21, A6 (range)					
Diercke Geography 2, chapter 7					
MTLD-value	50-60	60-70	70-80	80-90	90-100

Figure 7.23 - MTLD-values compared for English G21, A5 and Diercke Geography, Volume 2, chapter 7

With a targeted CEFR level at B1 in transition to B2, the results for the individual key figures are astoundingly exceeding the target. Those high CEFR levels are overtaxing for students or – from a teacher's perspective – extremely challenging with the required scaffolding and support to be prepared. Either way, these results put students' successful participation at risk.

With only the key figures on propositional density being within the targeted CEFR level, it is not surprising that all subchapters are equally overtaxing in their overall assessment of the CEFR level as is the average CEFR level for this subchapter at level C1+.

Subchapter	+	+	+	+	+	+
7.1						
7.2						
7.3						
7.4						
7.5						
7.6						
7.7						
7.8						
	A1	A2	B1	B2	C1	C2

Figure 7.24 - Overall CEFR ratings per subchapter Diercke Geography Volume 2, chapter 7 compared to curricular standards

Apart from the challenges caused by lexical sophistication, lexical diversity and propositional density, none of the analyzed chapters featured grammatical aspects beyond the expected competences of the targeted groups of students. The immense requirements, especially with *Diercke Geography Volume 2*, exceed that of *Exploring History* and hence can be considered as threatening to successful learning as the History textbook. Again, the Geography textbook has seemingly been designed to convey subject-related information with a high proportion of subject-specific language. Other than the History textbook, the two Geography textbooks seem to consider student progress, although the required language level is much too demanding. A teacher would have to know about that and, accordingly, would have to reduce linguistic complexity to ensure sufficient understanding to enable students to process the content. As with History, this would require specialized teacher training, preferably content-subject and language teaching combined.

7.5.4 Biology

History and Geography are both language-driven subjects, categorized as subjects of the humanities in the German school context; Biology is a science subject from the STEM canon (cf. Bildungsserver RLP, Fächer & Themen 2022), which is comparably frequently offered in bilingual teaching (see chapter 12.1). Typically for a STEM subject, Biology employs language mainly for communication of and about the content. The general aim is to enable students to use precise language and, in that, use a rather scientific language over everyday language. The latter is often subject in, e.g., History and, as seen above, can cause difficulties when it comes to processing the language. With the characterization of

language in Biology (see chapter 4.2), the assumption that language here is very demanding seems feasible, especially since Biology discusses cognitively complex processes.

The results obtained by the analysis (see chapter 7.3) showed that for the material designed to be used in grade 7, i.e., *Discover Biology Volume 1*, chapter 2, the required CEFR levels for the assessment based on the English Vocabulary Profile were one level above the targeted CEFR level for all three bands. For EVP A1, EVP B2, and EVP C1, level B2(+) was assigned. Although the subject matter may have allowed for other assumptions, the frequency counts based on the BNC indicated that the CEFR level for the average median position was at B1, which matches the targeted CEFR level for that grade. If this subchapter was taught early in grade 7, with students being expected to be at level A2 or transitioning to B1, this might be challenging but would undoubtedly be within the Zone of Proximal Development and hence, entirely appropriate. The percentage share of types being rated above that median position, meaning that those types rank higher and thus, being used more frequently, is at roughly 15%. So, the remainder of types used less frequently and therefore being more complex is comparably smaller than in History and Geography. Still, most types are likely above level B1 (as indicated by the color gradient in Figure 7.25). The percentages of types found on the Academic Word Lists implied that the chapter analyzed employed only smaller shares of those types. This resulted in CEFR level B1+, which is within the scope of the targeted levels as stated by the curricular standards. The propositional density, i.e., the packaging of information, arrived at level B1 for both nominal and verbal elements. In comparison to the subjects and chapters mentioned above, this level was slightly higher but still within the targeted range; thus, both key figures can be deemed appropriate. With a score of 68.5, the Flesch-Reading-Ease index indicated that the texts employed were at level B2, which, according to the adjusted scale (see chapter 6.4.1), is aimed at in grades 10 to 13. Hence, this key figure hints at possible overtaxing, although to a less extent than was discussed for History and Geography. Figure 7.25, contrasting the results to the targeted CEFR level per the curricular standards, supports the overall impression of this chapter's linguistic demand being slightly above the assumed competences of students in the respective age group.

		+	+	+	+	+	+
EVP A1							
EVP B2							
EVP C1							
BNC							
AWL							
Nominal Elements							
Verbal Elements							
Flesch-Reading-Ease							
	A1	A2	B1	B2	C1	C2	

Figure 7.25 - Ratings per key factor category Discover Biology 1, chapter 2, compared to curricular standards

Lexical diversity was measured to be comparable to that of advanced adult learners of English as a foreign language, yet the MTLD-value of 68 for *Discover Biology*, chapter 2 is exactly in the range determined for the targeted age group of grades 7 or 8 (see Figure 7.26).

Textbook					
English A3/A4 (range)					
Discover Biology, chapter 2					
MTLD value	50-60	60-70	70-80	80-90	90-100
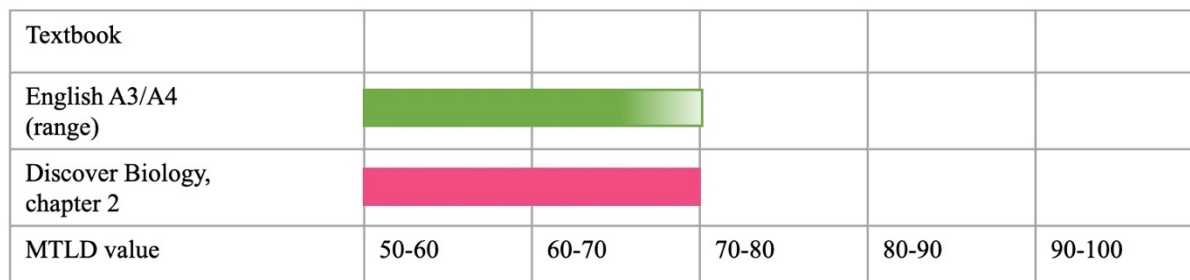
Figure 7.26 - MTLD-values compared for English G21, A3/A4 and Discover Biology 1, chapter 2

The results for the individual key figures already showed that the analyzed chapter of *Discover Biology* is appropriate for the target group, with singular aspects being a little too demanding. However, this excess of requirements appears to be manageable for both students, being slightly challenging, and teachers having to provide support for those aspects. Overall, this chapter does not bear the danger of disrupting the language acquisition process and gain in content knowledge caused by overtaxing language usage.

The overall ratings of the subchapters arrived at a comparable result with most subchapters rated B1 and thus, are fitting for the intended learning group (see Figure 7.27). Although the average CEFR level based on these individual overall assessments at level B2 seems to contradict at this point, expertly revising and scaffolding the three chapters exceeding the targeted CEFR level with B2 or even C1 could level out the demand.

Figure 7.27 - Overall CEFR ratings per subchapter Discover Biology 1, chapter 2 compared to curricular standards

The linguistic demand possibly increases with the next chapter analyzed, chapter 3. This assumption accounts for the expected progress of students towards the end of grade 7 or at the beginning of grade 8. The assigned ratings for the percentage of types allocated with the individual EVP bands confirm this. For each band A1 and B2, the CEFR level arrived at C1, clearly indicating an increase in complexity and, thus, in demand. At the time of potentially teaching this chapter, early in grade 8, level C1 exceeds the targeted CEFR level at B1 (cf. Ministerium für Bildung, Wissenschaft, Weiterbildung und Kultur 2014a: 147ff.). Same goes for the C1 band. Here the CEFR level arrived at B2+ which is also above the targeted level at the given grade. Hence, the vocabulary employed is too demanding, which may be explained by the frequent use of Latin-or Greek-based terms such as "oxygenated" or "aorta". Regarding the frequency of the featured types, the average median position computed indicated CEFR level B1, which was just in the appropriate scope of linguistic requirements. The percentage rate of types found on the AWLs indicated increased demand at level B2+; again, too demanding for the targeted group of learners. With propositional density, the requirements were below or just at the expected level of students in grade 8. Nominal elements arrived at level B1 while verbal elements arrived at an average CEFR level of A2+, which is slightly lower than the expected level at B1. When it came to reading and the readability assessment, the text was above the expected competence level, rated B2+ for a Flesch-Reading-Ease index at 64.

Since students would arrive at that value during grades 10 to 13 this chapter – at least with regard to this key figure – has to be rated as too demanding. The excess, however, is less than with History or Geography.

		+	+	+	+	+	+
EVP A1							
EVP B2							
EVP C1							
BNC							
AWL							
Nominal Elements							
Verbal Elements							
Flesch-Reading-Ease							
	A1	A2	B1	B2	C1	C2	
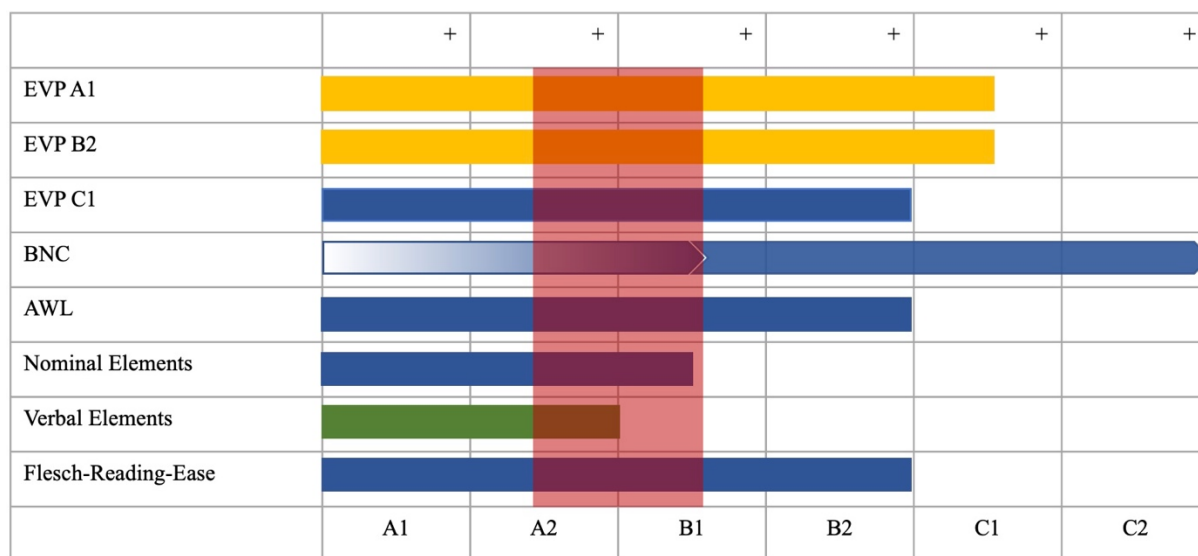
Figure 7.28 - Ratings per key factor category Discover Biology, chapter 3, compared to curricular standards

The average lexical diversity, calculated at 49.02 for MTLD, points to a diversity already given at a very early stage in language learning, according to the results obtained for *English G21* (see Figure 7.29). Regarding this figure, the chapter's lexical diversity can be considered no challenge for the targeted learning group. The chapter's content may explain this comparably low figure in lexical diversity: since the chapter elaborates on blood and blood circulation, the variety of vocabulary is rather limited. The noun "blood", e.g., is used 220 times (with a total of 5000 words), not counting all the compounds containing "blood-".

Textbook						
English A3/A4 (range)						
Discover Biology, chapter 3						
MTLD-value	40-50	50-60	60-70	70-80	80-90	90-100
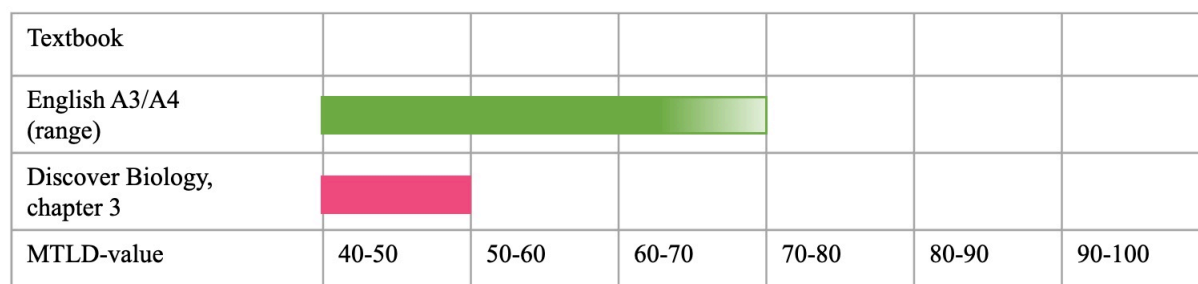
Figure 7.29 - MTLD-values compared for English G21, A3/A4 and Discover Biology, chapter 3

Since students in that grade would know most used grammatical phenomena teachers would have to focus on vocabulary and their level as assessed in this analysis. Numerous technical terms, such as those mentioned above, are already scaffolded or translated. Furthermore, extensive graphic support scaffolds the textual parts of the chapter for students to understand the described processes. Despite the high demand for lexical sophistication, especially for the EVP part, the overall CEFR levels per subchapter were equally distributed between B1 and B2(+). The average CEFR level for this chapter based on the overall subchapter assessments arrived at level B1+. Accordingly, the chapter in total is

challenging with occasional excessive demands that a skilled teacher, who knows about the challenges and provides support, can help manage. This way, the gain in both content and language knowledge can be secured, and the chances of successful participation for students are comparably high, especially in contrast to History and Geography. This will be further discussed in chapter 7.5.6.

Subchapter	+	+	+	+	+	+
3.1						
3.2						
3.3						
3.4						
3.5						
3.6						
3.7						
3.8						
3.9						
3.10						
	A1	A2	B1	B2	C1	C2
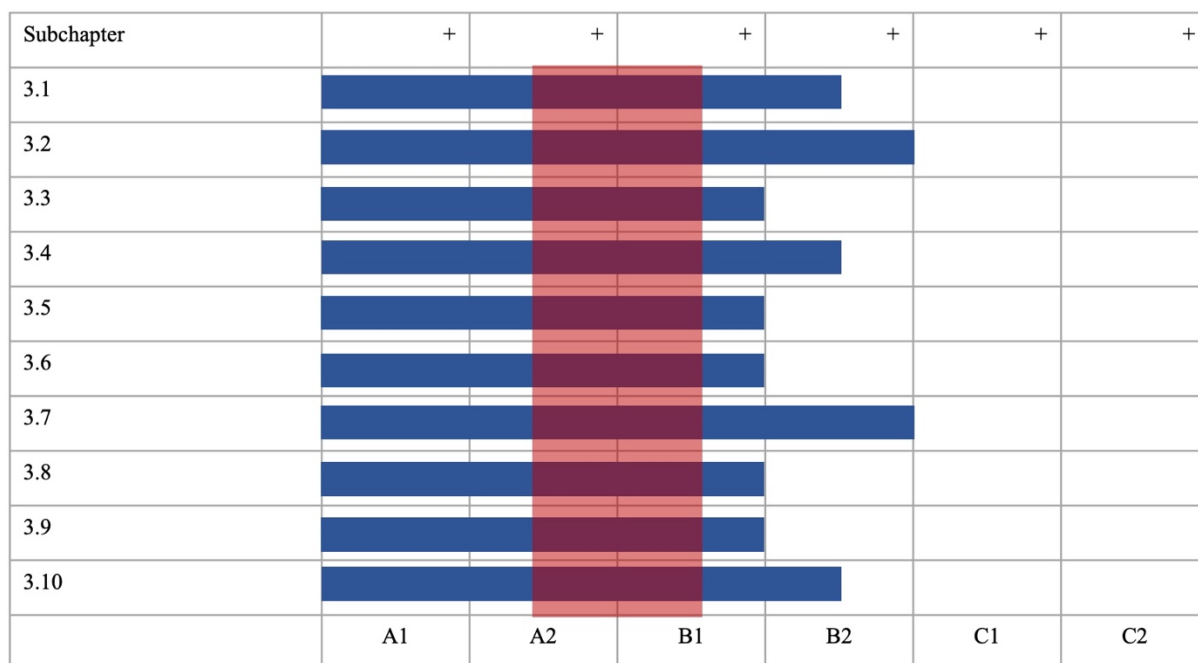
Figure 7.30 - Overall CEFR ratings per subchapter Discover Biology 1, chapter 3 compared to curricular standards

With chapters 4 and 5, a further increase in linguistic demand would have been plausible since both chapters cover subjects that are rather taught in grades 9 and 10 (cf. Ministerium für Bildung, Wissenschaft, Weiterbildung und Kultur 2014a: 147ff.). For those two upper grades, the targeted CEFR level is B1 with an onset of B2 towards the end of grade 10 (cf. KMK 2003: 11f.). Therefore, CEFR levels for select key figures in the upper B2 band at B2+ can still be considered as too demanding given the fact that students most likely arrive at that level, if at all, towards the end of grade 10 while this chapter would be taught in grade 9.

Chapter 4 was as demanding as the previous chapters and with that, still too demanding in some aspects. The vocabulary used, however, was rated on point for the targeted group of learners, with average EVP ratings of B1+ for EVP A1, B2 for EVP B2 and EVP C1. Apart from EVP A1, the assigned CEFR levels exceed the targeted level of a further developed B1. A closer look at the book's subchapter 4.1 with its high rating (C2+, see Table 7.11, p. 85) revealed that especially terms like *immune system*, rated by EVP as C2 (cf. Mathews & Olmesdahl 2010: 58), have caused this high complexity. Although *Discover Biology* did not offer a translation for those terms, most were very close to their German equivalent, i.e., *Immunsystem.* Consequently, working with those terms should not pose a problem for students. Nevertheless, the texts require a thorough analysis before working with them in class. Despite the technical terms or much rather terms of the medical or nutritionist field, the frequency counts arrived at an average median position of 515, implying an overall frequent use of the types featured in the

chapter. With the CEFR level being B1+ and approximately 85% of types being used less frequently and, thus, more complex, the lexical sophistication is still overtaxing for students at the beginning of grade 9. The CEFR level assigned to the average percentage share of types featured on the Academic Word Lists supports the assessment. With roughly 5% and a corresponding level B2, the demand here is still a level too high for the targeted level in grade 9. The information density was at level A2+ for nominal elements and B1 for verbal elements; with that, both key figures are appropriate for students at the targeted level of skills. The Flesch-Reading-Ease index was again above the targeted level with an average CEFR level of B2. Like chapter 3, chapter 4 features numerous pictures supporting the texts and hence, students could probably manage a slightly increased demand stemming from the readability in grade 9 (see Figure 7.31).

		+	+	+	+	+	+
EVP A1							
EVP B2							
EVP C1							
BNC							
AWL							
Nominal Elements							
Verbal Elements							
Flesch-Reading-Ease							
	A1	A2	B1	B2	C1	C2	
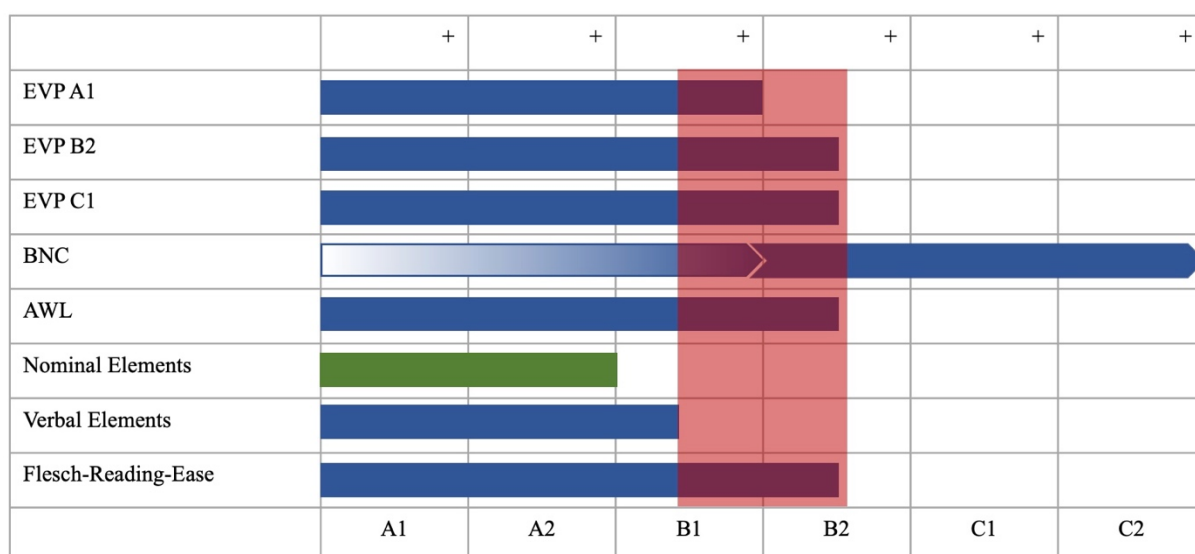
Figure 7.31 - Overall CEFR ratings per key factor category Discover Biology, chapter 4, compared to curricular standards

Lexical Diversity measurements revealed that the computed value for MTLD at 82.20 was just within the range of the MTLD value determined from the English G21 volume designed for grade 9. Just like with chapter 3, the variation of vocabulary, i.e., the diversity, was rather limited, which explains the topic-related complexity of types in opposition to the appropriate diversity measures, although the text as such was comparable to academic quality.

Textbook						
English A5 (range)						
Discover Biology, chapter 4						
MTLD-value	40-50	50-60	60-70	70-80	80-90	90-100
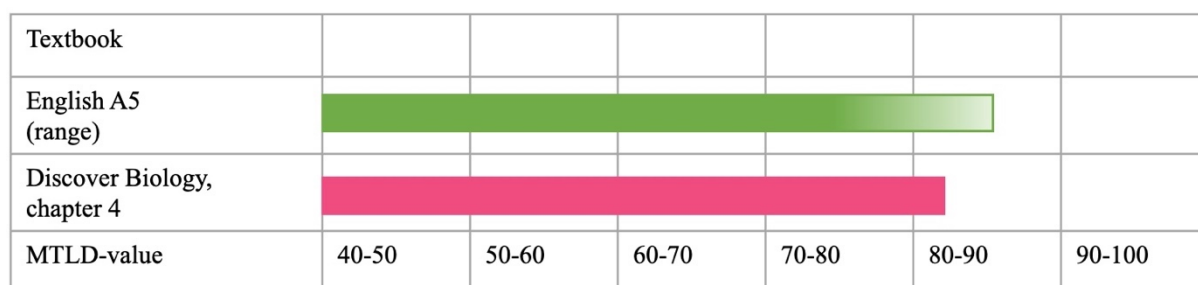
Figure 7.32 - MTLD-values compared for English G21, A3/A4 and Discover Biology, chapter 4

In comparison to the chapters analyzed before, chapter 4 is entirely overtaxing in one subchapter and minimally overtaxing in another 4 subchapters at level B2. However, requirements rated one level above the targeted CEFR level may even be deemed appropriate since the Zone of Proximal Development describes just that: things that learners can manage with the guidance of trained instructors, who in this study's context should be specially trained teachers. The overall CEFR levels computed for the individual subchapters imply the same: only one subchapter was B2+, four more at level B2 and with that, a single level higher than the targeted CEFR level. The average overall CEFR level also arrived at level B2 and with that, even if at the maximum value, can still be deemed manageable if the required support is identified, prepared, and provided. Consequently, the potential level of skill achieved by students within grade 9, supported by a skilled teacher, meets the demand this chapter imposes on students.

Subchapter	+	+	+	+	+	+
4.1						
4.2						
4.3						
4.4						
4.5						
4.6						
4.7						
	A1	A2	B1	B2	C1	C2

Figure 7.33 - Overall CEFR ratings per subchapter Discover Biology, chapter 4 compared to curricular standards

The last chapter analyzed within this study for Biology was chapter 5. This chapter's topic could be covered towards the end of grade 9 or at the beginning of grade 10. Since this chapter can be associated with the curriculum's *Themenfeld 8* (cf. Ministerium für Bildung, Wissenschaft, Weiterbildung und Kultur 2014a: 147ff.), it is reasonable to assume that this topic will be dealt with in advanced grades such as the abovementioned. Accordingly, a further increase in demand would be possible. Students should progress towards the end of the B1-band and onto the onset of B2. This expectation would cater to the average CEFR levels computed for the EVP key figures. With EVP A1 and EVP B2 arriving at C1 and EVP C1 average at B2+, an increased demand was noticeable. While EVP A1 and EVP B2 may be too demanding, EVP C1 vocabulary is already in the range of manageable vocabulary with teacher support following the idea of the Zone of Proximal Development. The CEFR level for the average median position of the frequency counts within the BNC arrived at A2+, indicating the lexical sophistication to remain below the targeted CEFR level. As mentioned above, the considerable percentage shares of the types with lower frequency counts must be considered as those

are more complex than the initially yielded CEFR level. Roughly 75% fell into that category, meaning that the median position does not clearly indicate the actual complexity of lexical sophistication. The median position resulting from the assignment of lower CEFR level, however, hints at the fact that the terms used in the text are generally of lower complexity. Thus, students likely process those without greater difficulty, even those types rated in higher CEFR levels. The findings and averages determined for the share of types featured on the Academic Word Lists resulted in a CEFR level for this key figure of B2+, placing this again on the brink of being too demanding if students were left unsupported. If, however, scaffolding and teacher support were in place, these types could enrich students' vocabulary even though the assessment indicates a level of complexity students may not be able to process at the time. For both key figures concerned with propositional density, the assigned CEFR level was at B1 placing the percentage share of nominal and verbal elements per sentence in the fitting band for the targeted learner group. The Flesch-Reading-Ease index placed this chapter again slightly above the targeted CEFR level at level B2 and an average value of rounded 64. As level B2 is already within students' reach, the curricular standards project an onset of level B2 towards grade 10; students could manage this level of demand in reading with support.

		+	+	+	+	+	+
EVP A1							
EVP B2							
EVP C1							
BNC							
AWL							
Nominal Elements							
Verbal Elements							
Flesch-Reading-Ease							
		A1	A2	B1	B2	C1	C2
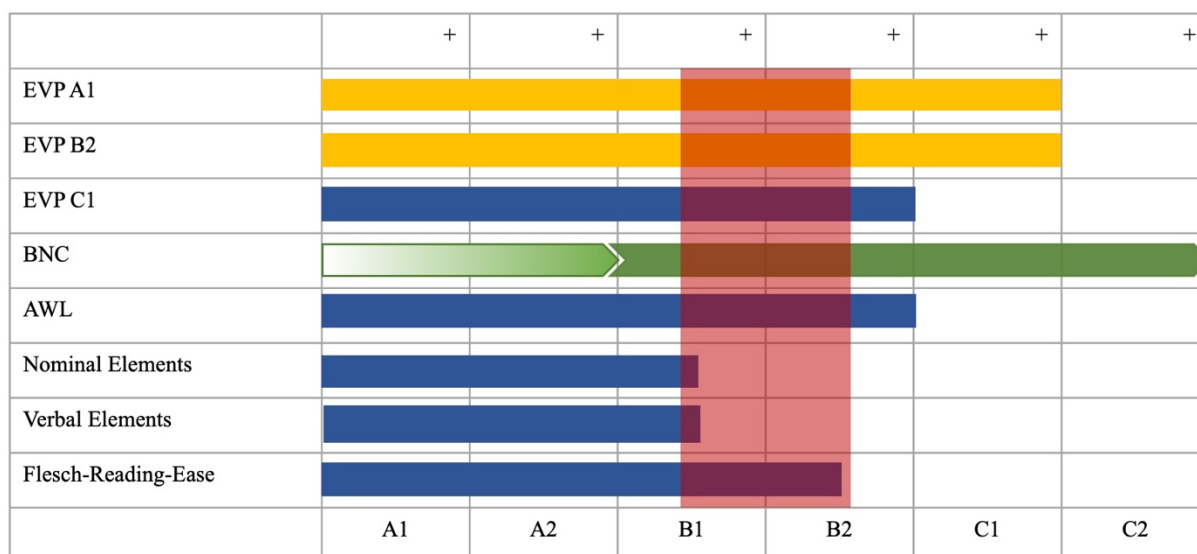
Figure 7.34 - Ratings per key factor category Discover Biology, chapter 5, compared to curricular standards

The average value calculated for lexical diversity arrived at 68, an MTLD value that would already be possible to work with in grade 8 (see Figure 7.35) and consequently fits the targeted learner group in grades 9 or 10 (see Table 7.15). Just like the afore-discussed chapters, chapter 5 displays a seemingly reduced lexical diversity, especially compared to chapters of History and Geography. The thematic focus can explain this and the constant recurrence of aspects introduced and discussed.

The chapters in *Discover Biology* follow the idea of forming a foundation per chapter and building differentiated, advanced knowledge focused on the foundation. Therefore, there is a frequent repetition of vocabulary, resulting in a lower diversity measure, MTLD.

Textbook						
English A5/A6 (range)						
Discover Biology, chapter 5						
MTLD-value	40-50	50-60	60-70	70-80	80-90	90-100

Figure 7.35 - MTLD-values compared for English G21, A5/6 and Discover Biology, chapter 5

The overall CEFR levels per subchapter (Figure 7.36) indicate that all but two subchapters are within the targeted interval of CEFR levels B1 and B2(+). Likewise, the average across all subchapters, providing an assessment for the chapter as such, at level B1+ places this selection of text well within the zone indicating the targeted levels. Accordingly, this chapter appears appropriate and manageable with occasional support. To provide students with the necessary help, trained teachers would have to analyze the material just like this study did using an automated analysis which yields information on the CEFR levels. Furthermore, teachers would have to be able to make sense of the returned values and levels and understand the implications. For chapter 5, e.g., teachers would have to look closely at the vocabulary used since in said two subchapters, the EVP ratings were considerably higher than in the other subchapters and the targeted CEFR level.

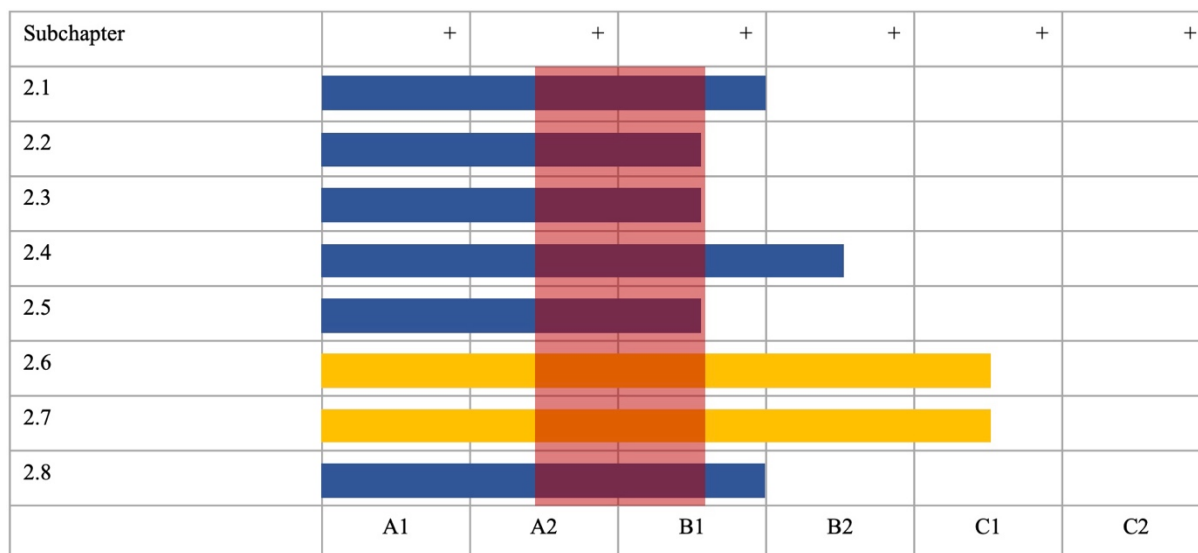| Subchapter	+	+	+	+	+	+
5.1						
5.2						
5.3						
5.4						
5.5						
5.6						
5.7						
5.8						
5.9						
5.10						
	A1	A2	B1	B2	C1	C2

Figure 7.36 - Overall CEFR ratings per subchapter Discover Biology, chapter 5 compared to curricular standards

The texts in chapters 4 and 5 exclusively use present tenses, either the simple present or the present continuous. According to the English curriculum, these tenses do not increase the linguistic requirements. Likewise, if used at all, conditional clauses type 1 are found to hypothesize on possible results of experiments. Remarkably, the chapters presumed to be taught in the advanced grades 9 and 10 are, on average, less demanding with CEFR level B1 than the first two chapters analyzed. Still, all four chapters are comparably less demanding than those analyzed for Geography and History.

Moreover, all four chapters are not overtaxing to such an extent that knowledge gain and language acquisition could be reduced or hindered due to an overly high language demand in the material. Support, however, is required throughout all chapters but within feasible dimensions. Seemingly, Biology imposes less linguistic demand than Geography and History. Chapter 7.5.6 compares all four subjects.

7.5.5   Mathematics

Other than the subjects analyzed and discussed above, Mathematics material was not from a single comprehensive textbook but a collection sampled from various textbooks used in the United Kingdom (see chapter 6.2.4). Consequently, the selected age group and topic structure the discussion on the implications of the key figures. Starting with grade 7, Algebra 1, the ratings for lexical sophistication, precisely the EVP categories, were at B1 for EVP A1, B2 for EVP B2 and B1+ for EVP C1. Correspondingly, EVP A1 and C1 are within the interval set by the curricular standard as the targeted CEFR level for students in grade 7. The EVP B2 types or their percentage share indicated that 7.29% of all types were too demanding. Appropriate support could help to manage types of this smaller share. Comparing the types featured to the BNC showed that the texts employ frequently used types. The median position computed for Algebra 1 indicated a CEFR level of A2+, implying a level of complexity just fitting for the targeted group of learners and the set curricular standards. Nevertheless, the fact that roughly 75% of all types were of less frequent use and thus, can be deemed more demanding has to be kept in mind when assessing the overall complexity and the required level of language skills for this topic. The percentage share of types featured on the Academic Word List, not surprisingly so, was comparably high at level B2, although definitions substituted most technical terms to simplify the text. Presumably, the AWLs feature comparably more types of Mathematics than they do from the other subjects. The propositional density of Algebra 1 was very low, with CEFR levels of A2 for nominal and only A1+ for verbal elements. This can be explained again with the character of the tasks analyzed. They consist primarily of two to three sentences which contain the required information and task in the form of an imperative sentence. With the brevity of the sentences, the number of verbal and nominal elements is very limited and hence, is the linguistic complexity, e.g., "If 17 textbooks cost €150.45, how much will 28 cost?" (Parsons 2000: 142, task 1). Consequently, the tasks do not exceed the level of skill students in grade 7 would possibly display. As a last key figure indicating the level of demand stemming

from the text, the Flesch-Reading-Ease index indicated that the collection of tasks in Algebra 1 fits at level B1. Although students in grade 7 may be at A2 transitioning onto the B1 band, students can master this demand level if a trained and skilled teacher provides the required language support, especially when considering that the text in Mathematics is usually read per task not as a whole. This graphical representation illustrates the level of linguistic requirements of Algebra 1 in comparison to the CEFR level set by the curricular standards indicated by the red-colored zone:

	+	+	+	+	+	+
EVP A1						
EVP B2						
EVP C1						
BNC						
AWL						
Nominal Elements						
Verbal Elements						
Flesch-Reading-Ease						
	A1	A2	B1	B2	C1	C2
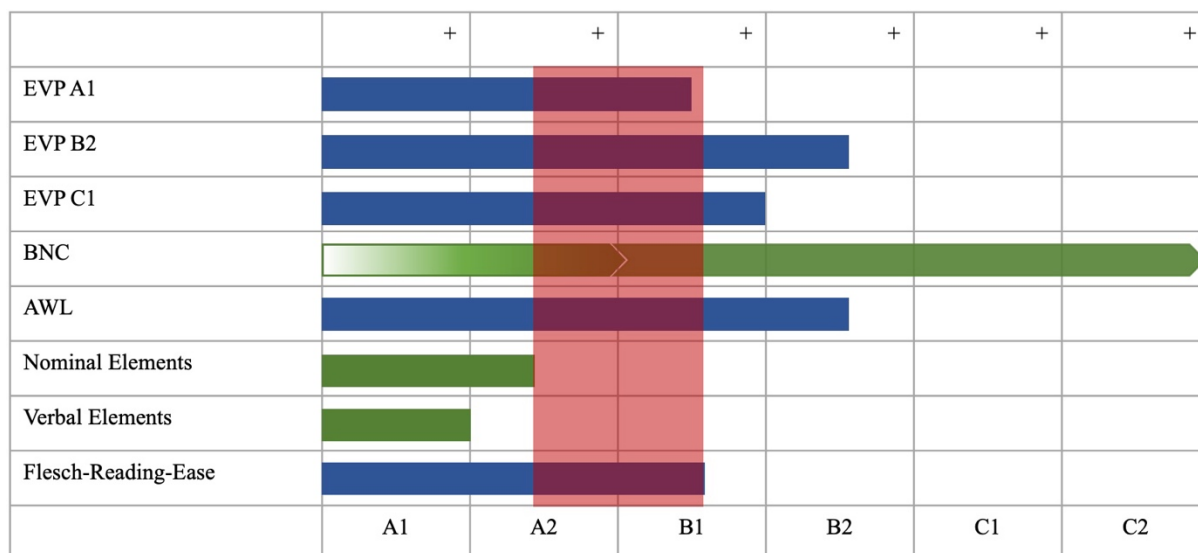
Figure 7.37 - Ratings per key factor category Algebra 1, compared to curricular standards

Lexical Diversity was very low, with MTLD arriving at 33, a value that was not even calculated for any of the reference texts in Cornelsen's *English G21*. The character of the compiled samples may cause this. In contrast to all other texts analyzed and discussed so far, the corpus for Mathematics was a sample of selected tasks and exercises for the given topics. Naturally, those focused on the task at hand and were given in the thematic context. Consequently, the vocabulary used was repetitive and of a comparably limited selection.

Textbook							
English A3/A4 (range)							
Algebra 1							
MTLD-value	30-40	40-50	50-60	60-70	70-80	80-90	90-100
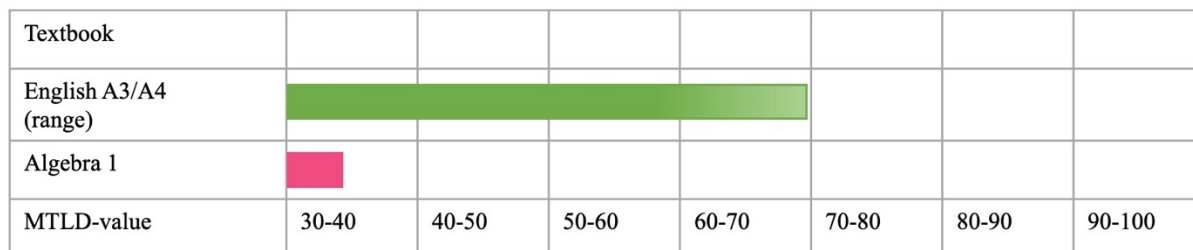
Figure 7.38 - MTLD-values compared for English G21, A3/4 and Mathematics, Algebra 1

Since the Mathematics material was not divided into subchapters, comparing those was impossible. Overall, the CEFR level for Algebra 1 arrived at CEFR level B1; hence, this material collection is suitable, with some support, for teaching in the lower age group of students aged 12 to 13.

The results for Basic Geometry (Geometry 1) were quite similar to Algebra 1. For lexical sophistication, the EVP ratings arrived at level A2, with more than 50% of all types in the A-band. The percentage of types within the EVP B2-band caused the CEFR level at B1+ while the percentage of types from EVP C1 was slightly higher than for Algebra 1 with a rating of C1. Apart from the latter, the lexical sophistication displayed in the collection of tasks concerned with Basic Geometry requires just what students are supposed to be able to master in the given age group, especially when Basic Geometry is taught at a later point within the school year. The median position within the BNC and its allocated CEFR level at A2 also imply that the requirements with Basic Geometry are suitable for students in the targeted age group. However, roughly 70% of all types were in higher positions as the word in the median position, hinting at a higher complexity which still, to a large extent, may be manageable for students in an advanced grade 7. This is especially likely since the words which may be used less frequently are technical terms, which would have to be explained or simplified in any language of schooling. The percentage of types found on the Academic Word Lists was comparably low, so this key figure arrived at level A2+. Besides Algebra 1, Basic Geometry, moreover the substituted text as described in chapter 6.4 does not seem to feature as many terms from the AWLs. The values determined for propositional density implied less dense information packaging for Basic Geometry than for Algebra 1, arriving at A1+ for both categories. This low density may partly be rooted in the fact that in Geometry, most tasks are accompanied and supported by illustrations and images to explain and clarify the task, question, or topic at hand. Accordingly, those illustrations serve as scaffolds and parts of the tasks or the problem to be solved. Especially with tasks that work with an illustration, the textual part, including the imperative sentences, are even shorter than with Algebra 1: "Look at the diagram. Write down the letters that describe an angle of 88°." (Baker 2008: 155, task 2). Working with this kind of task and the connected linguistic requirements should pose no problem for students in grade 7; hence, the selection of tasks for Basic Geometry fits the targeted group of students. The results for the Flesch-Reading-Ease index, which arrived at a slightly lower value for Geometry than for Algebra but was still placed within the range for B1, support this impression. With that, the demand was again in line with the curricular standards for the given grades.

With an overall CEFR level of B1 for Basic Geometry, the conclusion that this thematic part is completely appropriate – language-wise – for students in their first year of bilingual instruction is comprehensible.

		+	+	+	+	+	+
EVP A1							
EVP B2							
EVP C1							
BNC							
AWL							
Nominal Elements							
Verbal Elements							
Flesch-Reading-Ease							
	A1	A2	B1	B2	C1	C2	
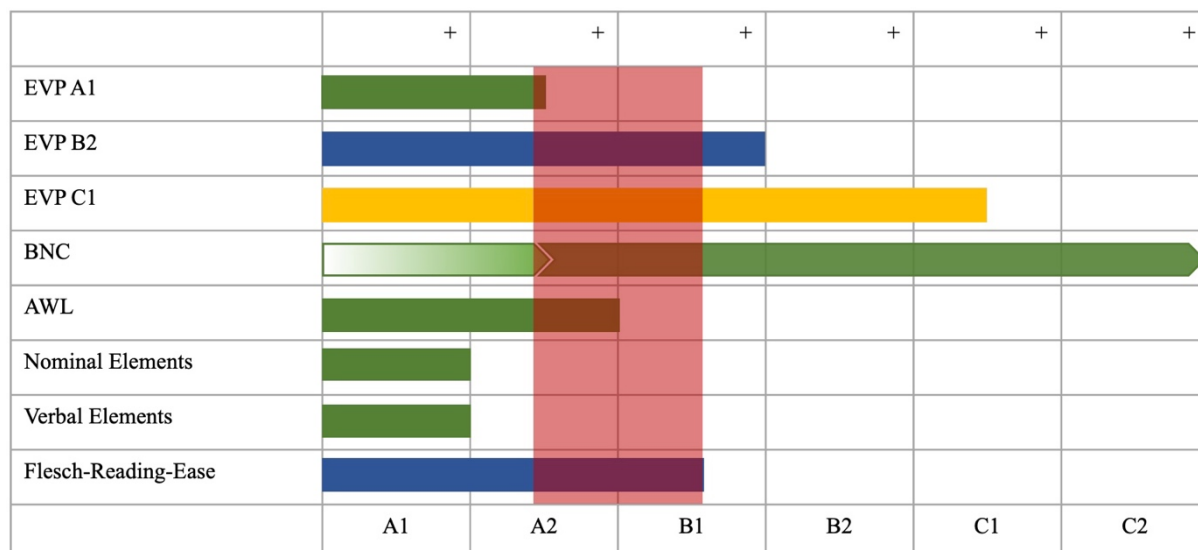
Figure 7.39 - Ratings per key factor category Basic Geometry, compared to curricular standards

Although for Basic Geometry, the measured value for lexical diversity, MTLD, was slightly higher than for Algebra 1, the value still was below all measurements taken for English G21 (see Figure 7.40). Consequently, the applied diversity of vocabulary within the selected tasks is not only appropriate for students in grade 7 but also a little to no challenge linguistically.

Textbook							
English A3/A4 (range)							
Geometry 1							
MTLD-value	30-40	40-50	50-60	60-70	70-80	80-90	90-100

Figure 7.40 - MTLD-values compared for English G21, A3/4 and Mathematics, Basic Geometry

Grammatically, neither topic requires competences beyond the level students have achieved at the presumed point of teaching (cf. Ministerium für Bildung, Wissenschaft und Weiterbildung 2000: 45). As seen in the two examples given above, most tasks featured the simple present, used passive constructions frequently, and rarely the continuous or other tenses. Occasionally, tasks in Algebra 1 used the conditional, mainly conditional 1. All mentioned grammatical aspects are introduced and could be repeated and entrenched with Mathematics in grade 7. Accordingly, it is doubtful those aspects overtax students when learning Mathematics in the targeted foreign language.

With progress in language competences to be expected when moving on to the higher grades, an increase in linguistic demand imposed by Mathematics material is possible. The tasks in grade 9 material

were again analyzed for the thematic fields of Algebra and Geometry. The requirements resulting from lexical sophistication observed with Algebra 2 did increase for the EVP-related assessment. For EVP A1, the assigned CEFR level was B1, while the B2-, and C1-band arrived at B2+. Level B2+ marks an increase from Algebra 1 and which could be slightly too demanding for students. With teacher support and scaffolding, however, students could manage this level. Assessing the median position within the BNC frequency counts, the material was rated at A2 with approximately 66% above the allocated position, meaning more demanding, as those types are likely more complex. Still, the requirement of this key figure allows for the determination that the material fits or is even a little too simple for grade 9. Likewise, the percentage of types found on the Academic Word Lists did not result in a level posing a challenge to students in the targeted age group as the CEFR level here arrived at level B1+ and with that just in the interval projected by the curricular standards. Since the character of tasks analyzed did not change, with short, informative or imperative sentences as tasks, the count for nominal and verbal elements per sentence remained comparably low, resulting in A2 for nominal and A1+ for verbal elements. Both assessments do not overtax the capability of managing propositional density for students in grade 9. For the Flesch-Reading-Ease index, the material for Algebra 2 was rated at CEFR level B2, slightly above the targeted level of an advanced B1. Again, this could be eased by competent support while teaching. Overall, the demand increased only in select key figure categories, if it increased at all, and not across all key figures. Indeed, the complexity of the material, arriving at an overall assessment of B1+ across all key figures, does not seem to reflect the progress students are thought to make when attending grade 9 in a bilingual program. This can be seen in the following bar chart as well.
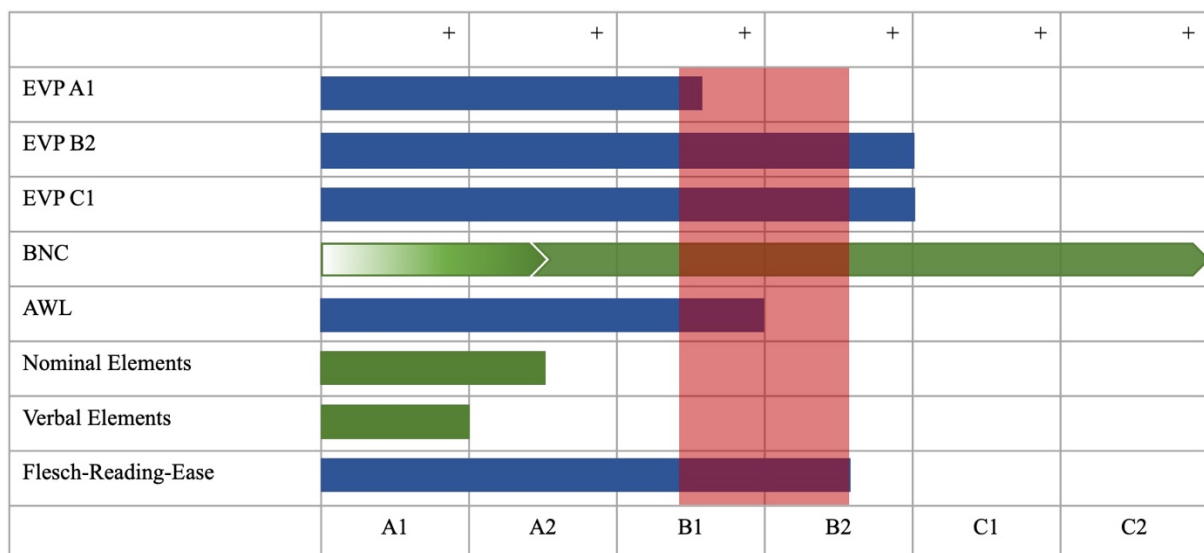
Figure 7.41 - Ratings per key factor category Algebra 2, compared to curricular standards

The low results for the MTLD measurements indicate that in this regard, the material for Algebra 2 is not overtaxing, just the opposite, fitting if not a little too easy for the targeted age group.

Textbook							
English A5 (range)							
Advanced Algebra							
MTLD-value	30-40	40-50	50-60	60-70	70-80	80-90	90-100

Figure 7.42 - MTLD-values compared for English G21, A5 and Mathematics, Algebra 2

With Algebra 2 being suitable for students having accomplished the set level of competences as given by the curricular standards, Advanced Geometry is just as fitting. For lexical sophistication measurements based on EVP, the CEFR levels arrived at B2 for a smaller number of types within EVP A1, whereas for EVP B2, the assigned level was B1. Students could handle both these levels, support provided. Only for EVP C1, the assigned CEFR level of C1 is overtaxing. However, this rating affects only 2% of the featured types and can probably be managed as well if students receive scaffolding and teachers provide more extensive language work on that end. For both the median position within BNC and the share of types featured on the Academic Word Lists, the same CEFR levels have been assigned as with Advanced Algebra, A2 for the former and B1+ for the latter key figure. Accordingly, lexical sophistication appears to be within the interval of CEFR levels defined by the curricular standards for the targeted learners. The assessment of propositional density arrived at comparable values and, thus, at similar CEFR levels as they did for Algebra 2. The CEFR level ended up being A2 for nominal elements, while for verbal elements, the CEFR level assigned was A1+. Just as with both Algebra selections as well as Basic Geometry, the general character of tasks and their semantics have not changed. If anything, the tasks for Advanced Geometry do not feature as many illustrations as the tasks for Basic Geometry did. The reduced number of illustrations could explain the slight increase in nominal elements per sentence since the tasks feature more descriptions of the objects discussed, e.g.,

> ABCDE is a regular pentagon. It is drawn circle center O. SAT is a tangent drawn to circle at A. Calculate the size of angle BOA.

> (Parsons 2000: 28, task 1).

The Flesch-Reading-Ease index for Advanced Geometry placed the material at a slightly lower level of complexity arriving at level B1 and with that, again entirely in the interval set by the curricular standards.

		+	+	+	+	+	+
EVP A1							
EVP B2							
EVP C1							
BNC							
AWL							
Nominal Elements							
Verbal Elements							
Flesch-Reading-Ease							
	A1	A2	B1	B2	C1	C2	
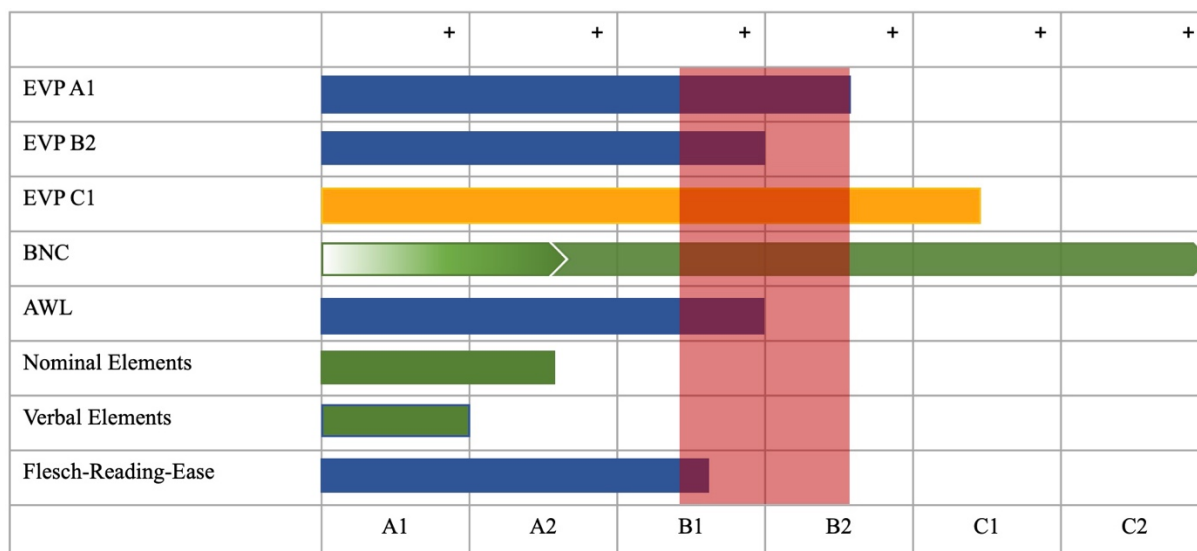
Figure 7.43 - Ratings per key factor category Advanced Geometry, compared to curricular standards

Lexical Diversity measurements, however, arrived at a higher value for Advanced Geometry but still clearly below the value students are supposed to handle when in an advanced grade 9. The analysis yielded a value of roughly 42, while the English textbook English G21 requires students to handle diversity measures between 81 and 89 in grade 9.

Textbook							
English A5 (range)							
Advanced Geometry							
MTLD-value	30-40	40-50	50-60	60-70	70-80	80-90	90-100
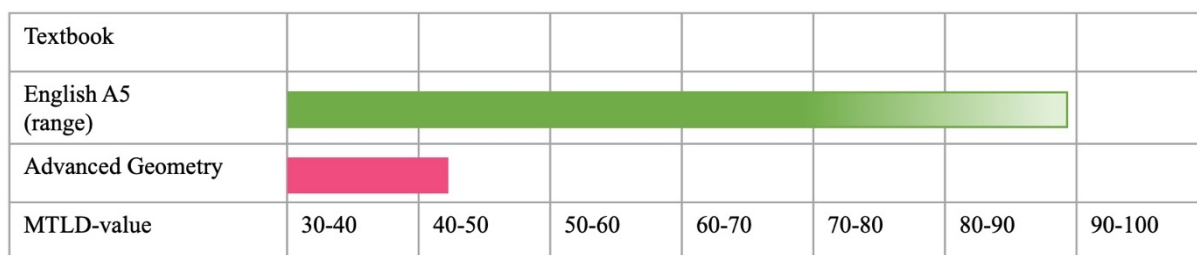
Figure 7.44 - MTLD-values compared for English G21, A5 and Mathematics, Advanced Geometry

Grammatically, the material selected for a potential grade 9 featured comparable aspects to the material for grade 7. The simple present was mainly used here to describe aspects of mathematical rules, such as the Pythagoras Theorem (cf. Casson 2001: 186, 11), which could be a topic of language work. The only surprising aspect was the rather colloquial use of the verb "get" in combination with the continuous. Nevertheless, the selected material is not overtaxing, much rather the opposite. Mathematics appears to be the least challenging subject, language-wise, and would thus be a fit for a first subject taught in a bilingual strand. This question, however, will be discussed in more detail in the following subchapter.

7.5.6 Subjects in comparison

Before going into the individual key figures, their results yielded from the analyses, their differences throughout the subjects analyzed, and the implications of those findings, it is mandatory to take a step back and discuss the differences between the selected subjects and the material as such. Also, the quality and the intention of the material per subject have to be focused on to make a valid comparison. At the end, a first suggestion of a sequence of subjects derived from this discussion is possible to answer RQ I. Furthermore, the discussion also answers those parts of RQ II and IV on the required proficiency by textbooks and the comparison of required and targeted CEFR levels, which aim at receptive skills.

Chapter 4.1 outlined the differences, in general, between the types of text, their intention, and their manner of making information accessible to students for the four subjects in focus of this study. Looking at the specific selected textbooks and comparing them, the differences in the intention of those textbooks and the way they were put together and edited become apparent. The History textbooks *Exploring History 1 + 2* aim to give a broad but comprehensive overview of an epoch or parts of an epoch, such as the analyzed chapters on *Absolutism and Enlightenment* (cf. Kröger & Lohmann 2007: 6-20) and *Industrialization* (cf. Kröger & Lohmann 2007: 54-76). Within those chapters, subchapters provide information and source material in line with the categories set by the curriculum (cf. Ministerium für Bildung, Wissenschaft, Weiterbildung und Kultur 2021: 84) such as authority, society, economy, and world view. Those four categories mark the guiding themes to be covered in every epoch. The information given in those subchapters is additive, meaning that the subchapters do not build on one another but rather add to a complete picture at which students are supposed to arrive by the end of the teaching unit having obtained a complete picture of the epoch, given the constraints of limited contact hours in the subject. Consequently, the information conveyed in one subchapter is not seen as prior knowledge for the following subchapter but as a part of a whole, leading to an understanding of historical intricacies and challenges within a given time.

The approach chosen in the Geography textbooks *Diercke Geography Volumes 1 + 2* is quite similar to that of the History textbook. The topics here cover geographical aspects such as those which the chapters dealt with, e.g., *Living in different climate zones* (cf. Löbmann 2007: 33-68) or *Europe Changes* (cf. Ahrend 2013: 104-120). Within these thematic fields, the textbooks are designed to comprehensively overview everything, every phenomenon, or geographic aspect tied to the topic at hand. Other than History, the curriculum for Geography does not feature categories but follows the idea of addressing the development of different competences as a guiding principle (cf. Ministerium für Bildung, Wissenschaft, Weiterbildung und Kultur 2021: 35). That is not to say that Geography teaching is not mandated with fostering the development of said competences; the role those competences play within History teaching is just a different one than in Geography. Accordingly, the chapters and subchapters in *Diercke Geography* are also more additive, complementing a comprehensive overview of a given topic. Accordingly, the chapter on *Living in different climate zones* covers all climate zones

to be found and, with those as a sub-division, introduces the characteristics of agriculture within each climate zone. So, with each subchapter, more information is added about both aspects, climate zones and agriculture. However, the subchapters cannot be fully regarded as prior knowledge in the course of working. Naturally, all subchapters are interrelated because they deal with the various climate zones, but other than that, they cover variations of the agriculture topic.

The material analyzed for Biology follows a different approach than History and Geography. *Discover Biology* also strives to give a comprehensive overview of the distinct topics, but other than the textbooks discussed above, the information within each subchapter can be regarded as a prerequisite to the following subchapter. That is to say that the subchapters cannot be dealt with as stand-alone but are consecutive while in History, economy during Industrialization, or in Geography, agriculture in the temperate zone could be dealt with as a stand-alone topic. To understand, e.g., the distribution of oxygen in the human body, it is essential to know about blood as a mean of transport, the function of the heart, and the circulation within the body. Consequently, the subchapters depend on one another in the Biology textbook. Furthermore, other than transporting or conveying the bigger picture on a topic, in Biology, the subchapters instead zoom in on a particular topic and all the details there are to it.

As was stated before, the quality of the collection of Mathematics material is very different from the material obtained for the other three subjects analyzed (see chapter 6.2.4). For the lack of a bilingual Mathematics textbook developed for and published in Germany, a collection of tasks was put together. This collection, however, is very similar to the German set-up of textbooks for Mathematics (cf. Körner 2016, 2016a). Unlike the abovementioned subjects, Mathematics textbooks rarely offer explanatory texts or aim at a comprehensive overview of specific topics. Mathematics textbooks rely on the teacher to elaborate on the topic during the introductory teaching sessions and provide students with tasks or exercises to practice the newly obtained skills. Accordingly, the quality and character of the texts featured in Mathematics textbooks differ fundamentally from those texts in the other three subjects' textbooks.

Looking more closely at differences in information access and processing with the selected material, the general differentiation as outlined in chapter 4.1 could be reconstructed. *Exploring History 1*, e.g. featured an excerpt of Rousseau's *Contract Social* on the question of Enlightenment, which contains no direct information on the topic and rather Rousseau's personal or political take on it (cf. Kröger & Lohmann 2007: 21). This source can be deemed as giving information indirectly and only through interpretation. The informative texts provided along with the sources contain direct information, which are enhanced by illustrations and caricatures further support the interpretation of contemporary views on the topic and are placed in the textbook to be thought provoking. Included maps, e.g., are given to provide students directly with information and support to process it. The chapter on *Living in different climate zones* of *Diercke Geography 1*, e.g., provides students with direct information both in the informative texts and the maps, graphs, or illustration. Everything is designed to provide students with explanatory information on the different climate zones and support them in the processing and

understanding of the differences in climate (cf. Löbmann 2014: 9f.). The same applies to how *Discover Biology 1* presents and supplements information for students. In addition, in this textbook, information is also given as instructive text, e.g., with description on how to set up and carry out an experiment, e.g. the instructions on "Making a smoke apparatus" (Mathews & Olmesdahl 2010: 37). Lastly, the texts in the Mathematics corpus compiled for this study, are solely instructive texts with graphical presentations intending to support and explain the text's imperative information, e.g., the depiction of a regular octagon to support students with calculating a specific angle (cf. CGP 2014: 28, Q1). Table 7.16 shows above-described findings in summary. It places the texts of each analyzed subject with the quality and character of texts in general as elaborated on in chapter 4.1 and specifically the findings yielded from the analysis support the preliminary placement of the subjects in Figure 4.1. The colors per subject are used throughout the remainder of this chapter.

		information access		information processing		
		direct	indirect	interpretative/ thought-provoking	explanatory / supportive	instructive/ descriptive
History	text	x		x	x	
	textual sources		x	x		
	illustrations & caricatures		x	x		
	graphs & maps	x		x	x	
Geography	text	x			x	
	graphs	x			x	
	maps	x			x	
	illustrations	x			x	
Biology	text	x			x	x
	illustrations				x	x
Mathematics	text					x
	illustration				x	

Table 7.16 - Subjects and text type related presentation and processing of information

The detected differences as described and illustrated serve the purpose of further understanding the vast differences discovered through the analyses of text complexity and the deduced requirements those texts impose on students in the various age groups in chapters 7.1 - 7.4. Having discussed the subjects and the obtained results and their implication so far only for themselves, a comparison of those findings throughout and across the four subjects is due.

The comparison of the four subjects takes the single key figures into focus again. Therefore, based on the calculated average CEFR levels per key figure for the single selected subchapters, an average is calculated again for the complete text selection per textbook. That is to say that for History, e.g., the averages for EVP A1 per subchapter are combined to arrive at an average value for the chapters on *Absolutism and Enlightenment* combined with *Industrialization*. This way, average demands were found for the single key figures for grades 7 and 8 and thus can be compared among all four subjects.

The corresponding CEFR levels to the average values derived from and discussed in chapters 7.1 to 7.4 are displayed in a matrix to support the rankings and findings. Each subject is represented by an individual symbol and color (see legend) and placed with the CEFR level derived from the values above. The red transparent rectangle marks the zone representing the curricular standards. The actual position of each symbol per key figure does not contain any further indication of gradual differences within the CEFR level, as it is caused by placing more than one symbol into one portion of the grid. In case CEFR levels were calculated to be elevated within the distinct band, a plus is added to the symbol.

Figure 7.45 - Average CEFR levels per key figure for grades 7 & 8, all subjects and nominal standard compared

If the discussion of the single subjects and their individual results already allowed for the idea of the vast differences between the linguistic demand of the selected texts of the four subjects, the illustration above confirms this. At least half of all key figures should be within the range of the targeted levels of the curricular standard to deem the analyzed texts appropriate. If less key figures are within the range, the texts are regarded as too demanding. In those cases, teachers would have to be aware of the challenges and would most likely have to reduce complexity considerably beforehand. That is, if teachers are specially trained and sensitized for such aspects.

According to the limit outlined above, out of eight key figures, Mathematics had only one key figure rated as more complex than the curricular standard ranging from advanced A2 to B1 (see Table 7.15). All other seven key figures were within the interval set by the curricular standards or even below, indicating less complexity and demand. Likewise, the overall assessment of Biology showed that this subject is more complex than Mathematics but less so in comparison to Geography or History. Out of the eight key figures for Biology, three were within the interval of curricular standards, while another three were one CEFR level above the targeted level at B2. The remaining two, EVP A1 and B2, were at level B2+ and thus, are overtaxing on a vocabulary level. Accordingly, with only three out of eight key figures in the appropriate CEFR level, almost two-thirds of the text is too demanding. Looking at the CEFR levels assigned to the key figure values for Geography, the impression of the subject being linguistically quite demanding is confirmed. Of the eight key figure categories, only two are within the interval indicating the curricular standards, while one arrived at an elevated level of the targeted level,

and was thus slightly too demanding. Another two key figures are even at level B2+, another two are at level C1 and one even at level C1+. Accordingly, for Geography, again, the majority is too demanding and even more demanding, as more key figure values resulted in the C-band level than for Biology. While Biology has no key figure in the C-band, Geography has three. History, however, has four key figures placed within the C-band at C1+ and only one placed with B2 and B2+, respectively. This implies overtaxing requirements and demand. Two key figures, and thus one less than with Geography, were assessed to be within the zone of the targeted CEFR level for the age group.

Since the analysis also yielded a ninth key figure, the values calculated for lexical diversity, MTLD, a comparison of these for the individual subjects is necessary. Those figures display the same trend with Mathematics being the least demanding, followed by Biology, Geography, and History. While the value for Mathematics implied very low requirements for lexical diversity, Biology already arrived at a value comparable to the diversity observable in averagely skilled adult ESL (see Figure 6.4, chapter 6.3.2). Geography and History displayed increasing diversity, with both values comparable to academic texts, with History being almost at the upper end of Malvern's scale. Compared to the MTLD-values calculated as a reference from *English G21, A3/A4* (see chapter 6.4.1, Table 6.2), here displayed as a purple trapezoid through which a line is drawn to make the reference as the curricular standard clear, shows that Mathematics is least demanding. Biology's MTLD values are lower than these for English, whereas the value for Geography is only slightly higher. Accordingly, both can be deemed appropriate with regards to lexical diversity. History, however, requires mastering a lexical diversity up to 25 points higher on the scale than English and thus, remains the most demanding and very likely overtaxing subject.
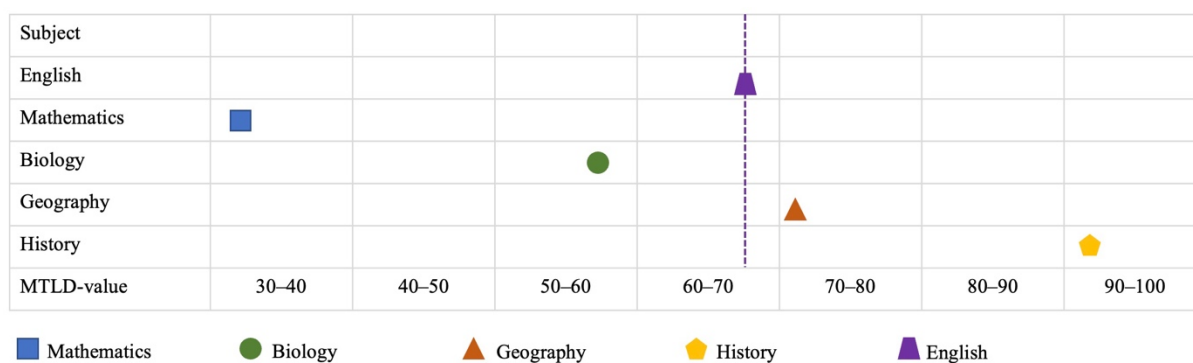
Figure 7.46 - Average MTLD values for grades 7 & 8, comparison all subjects

Based on the comparisons above, a first idea for a sequence of subjects can be derived, putting Mathematics in the first position, implying that this subject would be most appropriate for teaching in lower grades like grades 7 and 8. Biology could be suitable for an advanced grade 8. The subject could be introduced as a bilingual subject in the second half of the school year. Compared to the curricular standards set at A2/B1 (cf. KMK 2003), History clearly is too demanding for students in the targeted age group. Geography, and unexpectedly also Biology, are outside the targeted range of CEFR levels; however, the individual findings discussed in chapter 7.5.4 for Biology should lead to the conclusion

that with a specially trained teacher and their support, Biology could be taught in a later grade 7 or early grade 8, respectively. However, Geography, should not be introduced in these early grades as its demand already arrived at C-level for three key figures while another two were at level B2+. This makes overtaxing of students, and as a consequence, deficiencies in both content and language learning very likely. Accordingly, History and Geography do not seem fitting for the lower grades as the linguistic demands imposed by the textbooks appear too high.

The comparison of the results for the subjects and their material designated for the second age group, students of 15 to 16 years or within grades 9 or 10, confirms the initial sequence derived from the comparison for the younger target group. Based on the results obtained in chapter 7 and the implications outlined in the previous subchapters of chapter 7.5, it is likely that a similar result and sequence will be observed for the material designed for the higher grades in bilingual teaching. The comparison and discussion ideally commence with the overview of average values for the key figures per subject. As was the case for the lower grades, this overview comprises the average values for all chapters or topics analyzed of the distinct subjects.

Comparing the results above to the curricular standards is most convenient if all values are again transferred onto the CEFR levels. Accordingly, the average CEFR levels obtained through the calculation based on the values outlined in chapter 6.4.2 are displayed in a matrix again. This matrix allows for the comparison of the individual subjects as such and also with the targeted CEFR level as defined by the curricular standards. For grades 9 and 10 in bilingual education, this targeted level or much rather these targeted levels range from B1 to B2 (see Table 7.15).
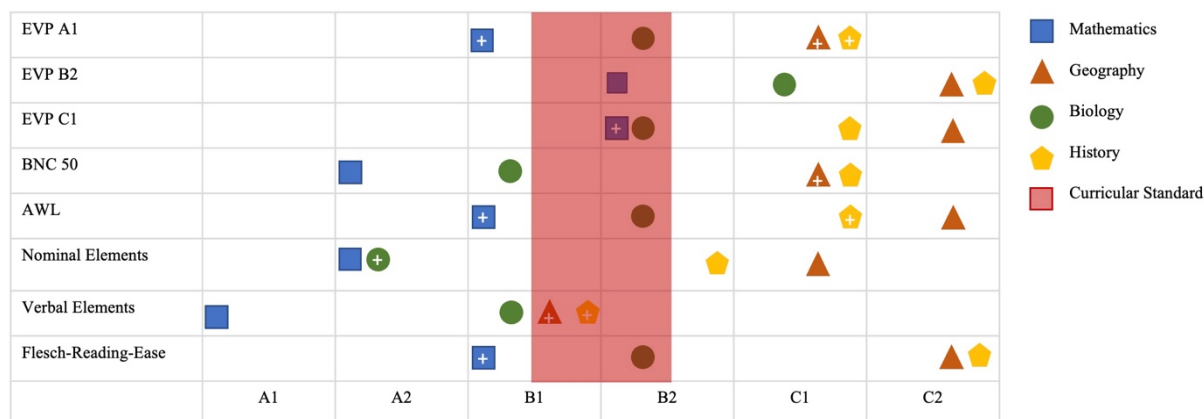
Figure 7.47 - Average CEFR levels per key figure for grades 9 & 10, all subjects and nominal standard compared

Supported by the illustration, it becomes visible that both History and Geography material was still too demanding in most categories. Based on the targeted levels as outlined in Table 7.15, students in grade 9 are expected to master language requirements wholly rooted in the B1-band, while towards the end of grade 10, students, especially in a bilingual track, should be able to master language requirements stemming from B1 and, in parts, B2. The red-colored zone displays this range. Consequently, the CEFR levels assigned to the categories for lexical sophistication indicated that for History and Geography, the

demand is too high to the extent that it exceeds what could be attenuated by teacher support and/or scaffolding. Learning success and knowledge gain are endangered. The possible effects of the significantly higher readability levels, actually targeted at university level, further add to the assessment of the material being inappropriate for the targeted group of learners. The categories concerned with information density were the only two in which History and Geography seem to match the targeted skill level.

According to the limit of categories being outside the targeted CEFR level range (see Figure 7.45, p. 136), Biology, in comparison, is challenging with most categories located in the B2 band but possibly manageable with sufficient support through a skilled teacher as this level exceeds the targeted levels by at most one level. All the more so, since for Biology, three categories were located in the zone of targeted CEFR levels. However, six out of eight categories for Mathematics were within or even below that zone. Only two figures were at levels outside of that zone, of which one, EVP B2, could be managed if language support is provided.

The ninth key figure, the measurement for lexical diversity based on MTLD, requires a different approach for comparison. Again, comparing the MTLD-values for the subjects to those computed for the English textbook material obtained from *English G21, A5/A6* is ideal. All values are displayed should be manageable in comparison to the values obtained for *English G21, A5/A6*. Here, the reference value was 92.07 as an average (see Table 6.2 - Overview of Text Inspector results for English G21 A2 - A6). This value was above all other subjects' MTLD values, so the lexical diversity of the texts analyzed was appropriate for the targeted age group. History, however, again had the highest value for MTLD at 86.26, followed closely by Geography with a diversity value of 81.51. Biology then comes third at a value of 75.13. The Mathematics value puts a considerable distance between the third and fourth position, with MTLD being 37.47.
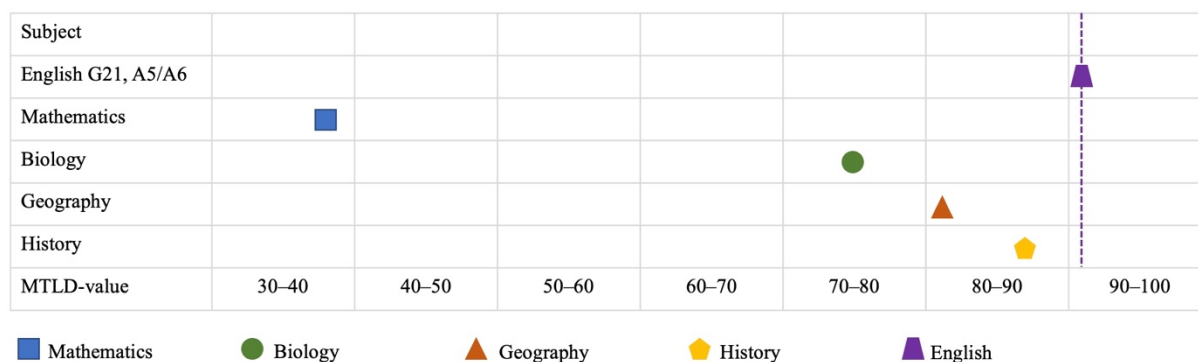
Figure 7.48 - Average MTLD values for grades 9 & 10, comparison all subjects

The completed comparison confirmed the initially identified sequence from least demanding to most demanding for the analyzed material of grades 9 and 10 (see also Figure 7.49). Mathematics material was the least demanding for both age groups, with CEFR levels generally within or even below the targeted zone of competences. With an average level of required receptive language skills for Mathematics between A2 and B1 (RQ II) and the targeted CEFR levels between A2 and B2 (RQ IV),

bilingual Mathematics does not overtax students and thus, potentially facilitates successful participation. According to this, Mathematics is an adequate subject to start with when offering a bilingual program at school (RQ I). Biology could be introduced as a second bilingual subject (RQ I). The average CEFR levels of the analyzed textbook material were between B1 and B2+ for both age groups (RQ I). However, it must be noted that the lexical sophistication detected in the material is too demanding if left unsupported. Even though the majority of the overall CEFR levels were within the targeted range of A2 to B2 (see Table 7.9 to Table 7.12) for all four grades (7 – 10), selected key figures were evidently too demanding, putting successful participation at risk (RQ IV). Lastly, History and Geography appeared to be prone to be overwhelming based on the findings yielded from the text analysis. With a required level mostly at B2(+) for the lower grades and C1 for the higher grades (RQ II), History textbook material was significantly rated above the targeted level. Successful participation would become extremely difficult, if not impossible. Alternately, the teacher would have to lower the level of expectation for assessments (RQ IV) or student contribution in the target language.

The conclusion on the results for Geography is similar. With CEFR level of B1(+) for the lower grades and C1 for the higher grades (RQ II), Geography at first seems suitable to be introduced as a third subject to a bilingual strand. However, a considerable number of key figures was significantly too demanding. Single key figures being rated at the additional CEFR level D1 for academic language (see Figure 7.22 and Table 7.6 to Table 7.8) and also, e.g., 16 out of 45 assessed key figures at level C1 to C2+ (see Table 7.5) allows for the assumption that the material would have to be rewritten to make it manageable and thus, beneficial for students. In both subjects' cases, History and Geography, the aspect of successful participation is doubtful or depends on either qualified teacher support or the revision of material beforehand. Since both subjects require a high degree of lexical sophistication, both should be

subject				
targeted CEFR-level	A2/B1	B1	B1	B1/B2
grade	7	8	9	10

■ Mathematics   ■ Biology   ■ Geography   ■ History
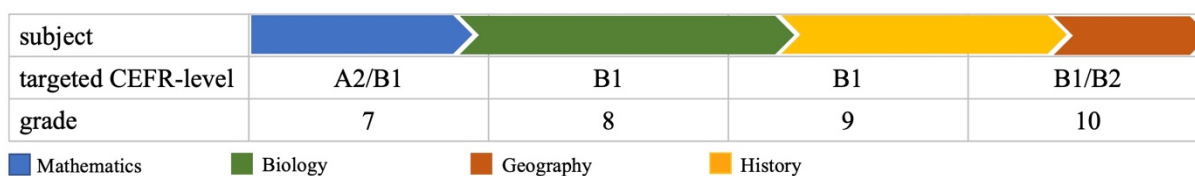
Figure 7.49 - Initial sequence of subjects based on results of text analysis

introduced to bilingual strands at a very late point in time (RQ I), maybe starting with History, as the advanced material was slightly less demanding than the material for Geography.

The following chapter will further evaluate all four subjects from the perspective of mainly productive skills. So far, it is appropriate to assume up front that the identified sequence of increasing linguistic demand and complexity persists with the productive part of teaching, i.e., the communicative part of working with tasks and discussing results and opinions since the complexity of the texts analyzed so far will partly reflect on the requirements imposed on students when it comes to working with these texts.

8 Language proficiency requirements – tasks

After analyzing the textbook material, or rather the receptive part of the textbook material, this part of the study analyses the productive parts of bilingual teaching. It looks at both receptive and productive skills required with the tasks and exercises students have to carry out based on the textbook material provided. The textbook analysis revealed that the complexity and demand differ significantly for the subjects, leading to a possible sequence of subjects potentially making successful participation in bilingual classes possible (see Figure 7.49). Mathematics material proved as least

Figure 8.1 - Research Design for Task Analysis

demanding – linguistically – followed by Biology, with History and Geography equally demanding and overtaxing. This order is based on the findings by *Text Inspector* and the interpretation thereof. It is most reasonable and beneficial to introduce Mathematics as a first subject in bilingual strands is most reasonable, possibly in grade 7. The other three subjects require increasingly advanced language skills in terms of the levels given in the CEFR and thus, would best be introduced at a later stage, e.g., Biology in the second half of grade 8. However, since teaching in schools comprises perception and predominantly production, this part must be thoroughly analyzed. Since this study is based on the material alone, all assessments regarding productive skills, both spoken and written, are made based on assumptions. Chapter 8.2 elaborates further on this approach.

8.1 Design

The approach to determining the demand imposed by tasks is very different from the approach chosen to analyze the textbook material. For this part of the study, no tool equivalent to *Text Inspector* was available. Consequently, the answer to the third research question (RQ III), along with RQ IV and, lastly, RQ I, has to be found differently. The objective here is to find the required level of skill to successfully and satisfyingly answer, work, or discuss the tasks given within the textbooks. Hence, the analysis for this part of the study features a tripartite approach as each task includes a receptive part: the (re-)reading of the text or parts of the text concerned with the potential answers and information. The productive parts include the oral production to answer, and the written production to round off a task or exercise. Accordingly, each task addresses three different types of skills.

Reading in this context includes reading the task as such but mainly reading the relevant text or parts of the text. For this part, *The CEFR Illustrative Descriptor Scales* (cf. Council of Europe 2018: 62f.) provides level-based descriptors to assess the reading skills.
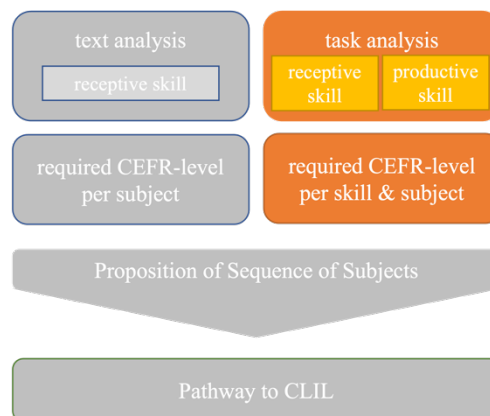
The companion differentiates between the purpose of reading, as indicated by the two variants

- *reading for orientation:* as the relevant skill with tasks implying that students only *scan* the text, i.e., "to looking quickly through a text to searching for something specific" or to *skim,* "reading a text diagonally at speed to decide whether to read […] properly" (ibid. 2018: 61),

- *reading for information and argument:* refers to detailed reading mainly applied in the academic context both at school and university (ibid. 2018: 63). This skill requires reading and understanding the text at a very detailed level to extract and work with the information given. Therefore, the CEFR levels provide a scale for the depth and intensity of understanding ranging from "simple phrases and important information " at level A1 to "very long, complex [and] specialized text [on] unfamiliar topics" at level C2 (ibid. 2018: 63).

Different genres further differentiate the manner of reading. Those genres include *reading correspondence,* such as emails, letters, or formal correspondence (ibid. 2018: 61) and *reading as a leisure activity*, which includes all fictional writing in contrast to *reading for information/argument* which is basically concerned with non-fictional writing. An exemplary description for each level is given in the *CEFR Illustrative Descriptor Scales* and was used for the analysis (see chapter 8.2 for further information).

Usually during lessons, the second step to working tasks is the spoken part. Spoken discourse clarifies questions or details up front and helps to gather initial answers and findings, or discusses or comments on the matter. The CEFR differentiates between *overall spoken production* as an overarching, general category (cf. Council of Europe 2018: 69ff) and three subskills differentiated by the character of spoken production. The three subskills of *spoken production* fulfil three "macro-functions (interpersonal, transactional, evaluative)" (ibid.) along which the categories are organized:

- *describing experience:* mainly concerned with narratives (of own experience) and, when it comes to teaching, descriptions of illustrations, painting, charts, or caricatures (cf. Council of Europe 2018: 70). Bilingual textbooks usually provide the vocabulary for those specific tasks in the form of chunks, technical terms, or legends given with the illustration. The CEFR levels for this category reflect upon the "aspects described, from simple everyday information […] to detailed descriptions of complex subjects" and the "complexity of discourse, from simple words and expressions to […] integrating subthemes and developing particular points […]" (ibid.).

- *sustained monologue: giving information:* focuses less on the exchange between interlocutors but on the contribution of a single person to a matter (ibid.: 71). Hence, this category fits the usual teaching and discourse approaches: asking a question and students individually answering. The levels differentiate along the line of two main aspects, the "type of information, from a simple description of an object […] to complex professional

or academic procedures" and further the "level of precision, from simple descriptions […] to communicating detailed information reliably […]" (ibid.)

- *sustained monologue: putting a case.* moves away from the descriptive character and towards the ability to take part in and lead an argument. The scale of CEFR levels reflects three main aspects, which are topics from personal interests and opinions to "thematic issues and complex issues" outside of the personal scope" (ibid.: 72), the "manner of arguing, from making simple direct comparisons to […] taking into account the interlocutor's perspective […]" (ibid.). The third aspect reflected in the scales is the "manner of formulation, from presenting an idea in simple terms to […] formulating points in well-structured speech" (ibid.).

Assessing the assumed required level of skill should mainly happen within those three spoken categories catering to the macro-functions of interpersonal, transactional, evaluative speech. Only if none of the descriptors of the three subskill fits the character of spoken production at hand, overall spoken production as a general category can be consulted. Within this category, the scales for that category are similar to those employed with the other three categories, yet not as differentiated for the aspects described before. The quality of language is not part of the evaluation in this category as the CEFR companion features "scales for communicative language competence". The qualitative aspect, i.e., the required vocabulary level, was already covered in the text analysis and deemed sufficient as students would most likely use what is given within the textbook material. The findings by Llinares and Wittaker support this approach since they found that students "hardly ever used" elaboration, a central part of developing content subject competence but imitate language patterns whenever possible rather than producing language themselves (cf. 2007: 90). Therefore, the analysis is not based on actual student production but rather on the assumed required level of language to answer, work, or discuss a task.

The third step of working a task or exercise is usually the written part. It serves the purpose of securing newly gained knowledge, recording results, or transferring obtained knowledge to a similar context and applying knowledge in a new yet related context. The abovementioned macro functions (interpersonal, transactional, evaluative) partly merge in written production, which differentiates again in a general, overarching category and two subskill categories:

- *written reports and essays*: mainly concerned with a combination of "transactional and evaluative writing" (Council of Europe 2018: 77). Its scales are formed along the lines of content, "from familiar subjects of interest and routine factual information to […] complex academic and professional topics […]" (ibid.), and the type of texts, "from short reports and posters to complex texts which present a case […]" (ibid.). Lastly, the complexity of discourse is decisive when it comes to the assessment of CEFR levels in connection to written production. It ranges "from linking sentences with simple connectors, to smoothly flowing expositions […]" (ibid.).

- c*reative writing* leaves more room for individuality as it "involves more personal, imaginative expression […]" (ibid.: 76). The main aspects the CEFR level scale follows are the aspects described, "from simple, everyday information to […] engaging stories and descriptions of experience" (ibid.) along with the types of texts, "from diary entries [….] to well-structured and developed descriptions […]" (ibid.). Another facet influencing the assessment of *creative writing* is the complexity of discourse, which ranges "from simple words and phrases to […] following established conventions" when composing text (ibid.).

Alternately, if neither *written reports and essays* nor *creative writing* matches what is asked for in the individual task within the analysis, the general category can be applied:

- *overall written production:* similar to *overall spoken production,* this category is a rather general, overarching category which lacks specific differentiation and applies more generic aspects to assess writing along the scales of the CEFR.

Other than with spoken production, the level descriptors for written production assess the quality of language as a central aspect. The differentiation follows the "use of language, from basic vocabulary and simple sentences, to an assured natural style appropriate to both the genre adopted and the reader" (ibid.).

This study on the productive linguistic requirements assesses each task individually to find the required level of skill respecting, i. a., the skills and their differentiated categories at hand. It is important to note that the analysis is not based on actual student contribution but on the assumption of what content/ production on which level would be required to ensure successful participation in the bilingual subject at hand. The overriding question in this context is to find out what would be sufficient to answer, work, or discuss the task at hand, which also provides an answer to RQ III on the (average) level of proficiency required by tasks and discourse functions within the textbook material. The exact procedure applied to determine the required skill level is elaborated on in 8.2 below.

8.2 Methodology

All four subjects and their teaching material entail many tasks and a wide selection of discourse functions. Each discourse function requires a different level of skills when it comes to reading, speaking, and writing. It is possible that the same discourse function requires different skill levels with different tasks, i.e., speaking could be more demanding than writing and vice versa. Depending on the level of complexity of the material determined in the first study on the required receptive skills, the assessment of discourse functions in distinct tasks could vary. For example, a discourse function such as DESCRIBE[13] is aimed at a straightforwardly outlined process composed of relatively short sentences and simple phrases requiring a lower level of receptive and productive skills than the description of a

---

[13] Verbs in capitalized letters refer to the discourse function, the verb in its imperative form used to specify the expected activity given with a task or exercise.

multi-facetted illustration or caricature, which invites interpretation. This fact had to be kept in mind during preparation for the study.

8.2.1 Preparatory steps

The analysis of the tasks for their linguistic demand required substantial preliminary work. As a first step, all tasks and exercises were extracted from the textbooks and the teaching material per subject. Based on that, an overview of all discourse functions used per subject was derived. All verbs in the imperative form such as LIST, DESCRIBE, or ANALYZE, usually located in the initial position of the sentence, qualified as discourse function. These imperative verb forms comprised then a comprehensive list of all distinct discourse functions. These discourse functions were in the target language, while the lists by the KMK for German schools abroad were in German. Accordingly, a mapping was required to match those English discourse functions with the German ones. Different approaches were necessary to find equivalents for all discourse functions.

For History, the discourse functions were given along with definitions in the target language by the textbooks *Exploring History 1 & 2* (cf. Kröger & Lohmann 2007: half-title). These were reconciled with the German equivalent to ensure that the definitions were similar (see chapter 12.6.1). For Geography, however, the textbook did not provide a comparable list. Hence, a list of discourse functions for bilingual teaching, published by the Ministry of Education of North Rhine-Westphalia (NRW) was consulted and matched with the list compiled from the teaching material (cf. Ministerium für Schule und Bildung des Landes Nordrhein-Westfalen 2015a) (see chapter 12.6.2). The Ministry of Education NRW also published a list of discourse functions used in bilingual Biology classes providing the discourse function and its definition – however, the definitions were in German (cf. Ministerium für Schule und Bildung des Landes Nordrhein-Westfalen 2015). This part was complemented with a list of discourse functions frequently used in science classes (Physics, Biology, Chemistry) for German schools in foreign countries compiled by the KMK (cf. KMK 2013b) (see chapter 12.6.3). Coming up with a list of discourse functions for Mathematics required a different and extensive approach. A collection comparable to the ones for the other three bilingual subjects was unavailable. Moreover, a mere translation bore the risk of misinterpreting the actual meaning and intention of the discourse function and was therefore ruled out as an approach. Finally, the discourse functions extracted from the teaching material were matched with either a list of discourse functions published by administrations of English-speaking countries, such as the *Glossary of Verbs Associated with the New York State Next Generation Mathematics Learning Standards* (cf. New York State Education Department 2017), or the actual curricula provided by the UK government on the key stages (cf. Department for Education 2013b), the US government (cf. Dossey, Halvorsen & McCrone 2012), or the syllabus for Mathematics K10 as published by the Department of Education of New South Wales (cf. NSW Standards Authority 2019) (see chapter 12.6.4).

With the resulting comprehensive lists of discourse functions and their subject-related definitions, another preliminary step was taken to ensure that all details influencing the assessment of linguistic requirements were taken into consideration. This step entailed a thorough reflection on each discourse function per subject to allocate all possible CEFR levels to said discourse functions. Each discourse function has two sides of requirements/demand towards students: first, the information has to be received, hence receptive skills are addressed. The discourse function DESCRIBE, e.g., as simple as it may seem to describe something, requires students to identify and understand the aspect to be described. Furthermore, descriptions can be based on an object *in rem*, an illustration, or a written source. After receptive processing, students need to put their findings either in oral production, hence productive skills of speaking are addressed. Likewise, students may also have to put their findings in writing, which requires the productive skill of writing. Depending on the quality/complexity of the material to which the task and the corresponding operating function refers, different levels of the respective skills are addressed. The mapping of operating functions and CEFR levels refers to working with the textbook or written material. Therefore, listening as a receptive skill is not respected; also, because research in schools was not possible due to the global pandemic, CoVid 19, at the time.

For every discourse function and skill, the various levels and their can-do-statements were examined as to whether the statement might be feasible for the discourse function. This resulted in extensive mapping tables for each subject. Table 8.1 below illustrates an example of the mapping process, taken from *Exploring History 1 & 2*, to clarify the above-described approach. Here, the discourse function SHOW was mapped onto potentially fitting categories and can-do-statements from the *CEFR Illustrative Descriptor Scales* (cf. Council of Europe 2018). Following the definition provided by *Exploring History 1 & 2* (cf. Kröger & Lohmann 2007: half-title), the descriptors for the different skills were analyzed for the most appropriate match with a can-do-statement. These were listed and annotated with the corresponding CEFR level.

AFB1	Show	Give information to be found in other material (sources, maps, etc.)	Reading for information/argument	Can get an idea of the content of simpler informational material and short simple descriptions, especially if there is visual support.	A1
				Can identify specific information in simpler written material he/she encounters such as letters, brochures and short newspaper articles describing events.	A2
			Sustained Monologue: giving information	Can explain the main points in an idea or problem with reasonable precision.	B1
				Can communicate detailed information reliably.	B2
			Overall Written Production	Can write a series of simple phrases and sentences linked with simple connectors like 'and,' 'but' and 'because'.	A2
				Can write straightforward connected texts on a range of familiar subjects within his/her field of interest, by linking a series of shorter discrete elements into a linear sequence.	B1

Table 8.1 - Example of mapping of discourse functions to CEFR levels based on Can-Do-Statements for History

The discourse function SHOW, allocated with AFB 1 for History, in its typical application in History, i.e., "give information to be found in other material", usually requires reading skills at a lower CEFR level as students have to gather only specific information from other material. Therefore, A1 and A2 were allocated. SHOW in History teaching requires students to point out single aspects, other than in Mathematics, in which SHOW is thought more in the sense of proving something. In History, however, as per its definition, SHOW is mainly focused on naming or pointing out distinct features or aspects to be found in material outside of the main text. Consequently, the reading technique of scanning or skimming would suffice to retrieve the required information, especially when this information is based on illustrations or objects. The spoken production part is more challenging than the reading since students are required to communicate their findings with a certain level of precision and detail. Based on the material students work with, the required skill level for the spoken part, *sustained monologue: giving information,* could be B1 or even B2. Finally, writing would be a little less demanding as the information obtained and communicated orally only has to be recorded. Again, depending on the material worked with, the levels can reach from A2, recording simple sentences, to B1, writing a straightforward, coherent text. This preliminary step was indispensable since not only differences in the definition and interpretation of discourse functions among the subjects could be detected, but also awareness and realization of defining factors for the various CEFR levels were created. All mapping tables are attached in the chapter (see chapter 12.6). Finally, this mapping and allocating of discourse functions to all possible CEFR levels resulted in a process flow that served as the coding manual for analyzing tasks and their linguistic requirements.

This process flow (see excerpt of Figure 8.2) is a systematized guide to analyzing the tasks. The full process flow is attached to this thesis at the end and can be folded out, alternately, it is also given in the chapter (see chapter 12.7). The process flow or decision tree, consisted of four main threads or steps to be run through for each task and discourse function. Based on Kuckartz (see chapter 5.2), the categories and criteria to structure the process were deducted from the CEFR-based skills reading, speaking, and writing. The first step identified the nature of the task and the discourse function used. This step was to determine which activities were required, and which definition for the discourse function based on the subject and the respective mapping lists applied. A catalog of criteria extracted from the can-do-statements of the Companion Volume with New Descriptors helped to allocate the appropriate CEFR level to each skill required to work the task for the next three steps. For reading, the criteria were the length of the text, the complexity of phrases, the level of depth to which the text had to be understood, the quality of information presentation differentiated between direct and straightforward as opposed to indirect information, the recipient's familiarity with the topic, and the type of discourse. Initially, an inspection of the task at hand determined the required activities. For speaking, the various levels of criteria were identified along the lines of the required quantity of oral production (single words and phrases to fluent, cohesive communication), level of complexity of language or lexical sophistication, level of detail and elaboration, and quality of the subjects talked about (familiar to

unknown, subject-related fields). The required writing skills were evaluated with the same collection of criteria. Similar to spoken production, written production was assessed along the lines of length of written production (single phrases/terms to lengthy well-structured text), level of familiarity with the subject, level of detail, lexical sophistication, and the genre of the expected written product ranging from writing about the own opinion to an argumentative text. In cases the material to be worked with did not comprise text, this step was left out, and the analysis started with evaluating the spoken production. Based on this catalog of criteria per skill, each task and its allocated activities were evaluated to arrive at a CEFR level. The coding manual is shown completely in the appendix (see chapter 12.7).
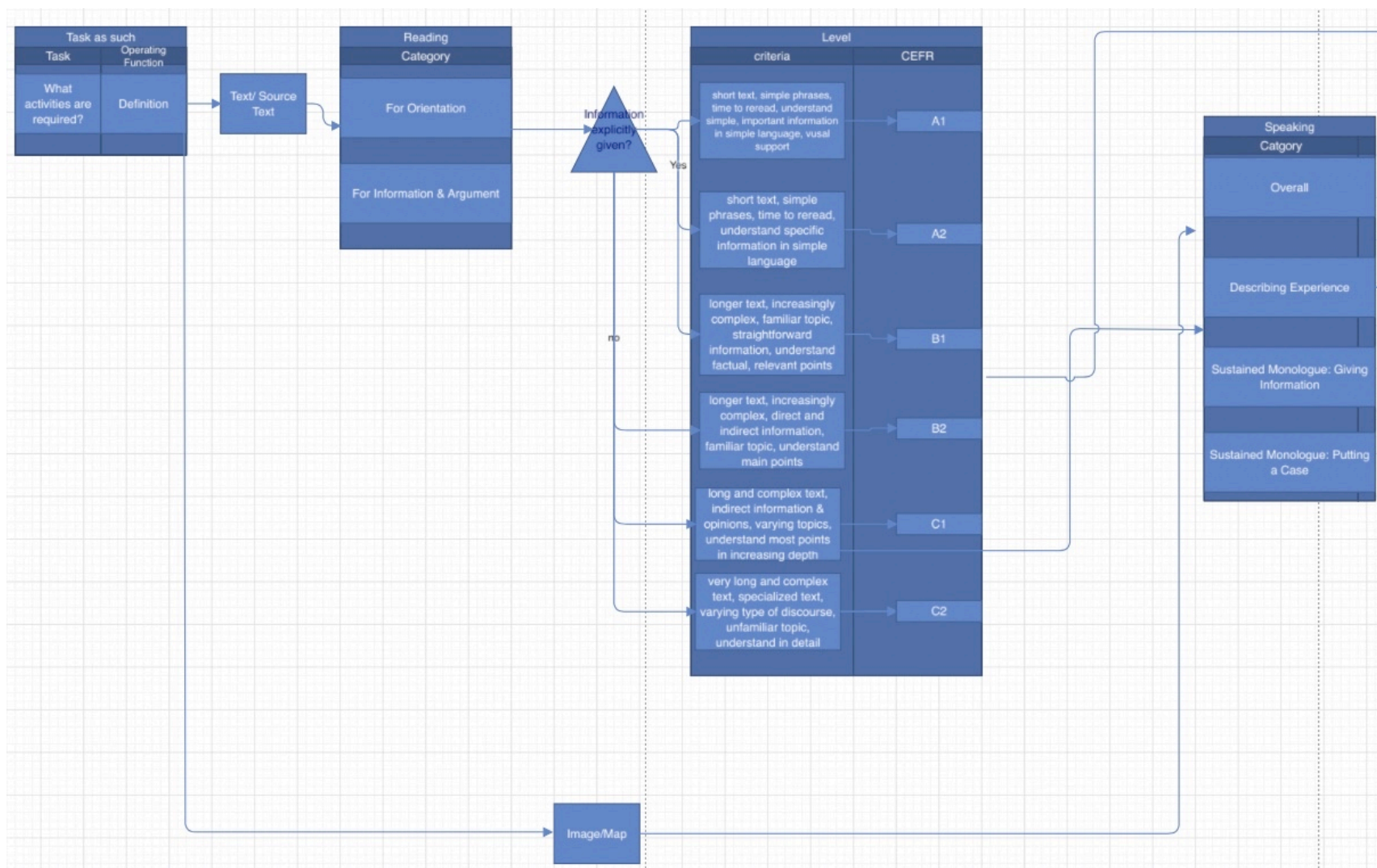
Figure 8.2 - Process Flow for Task Analysis, excerpt (without part for written production)

8.2.2   Conducting assessment

Based on the process and code manual described above (see Figure 8.2), the (expected) learner language, the level at which learners should have arrived in their language development (cf. Marx & Mehlhorn 2016: 298), was analyzed using the scales per skill and thus, applying the methodology of valence and intensity – evaluating language or text along the lines of standardized scales (see chapter 5.2). Since "to do well" used as the main criterion in Moe et al. (cf. 2015: 9) is a rather fuzzy concept, this study needed a guideline of criteria and their minimally required level to consider possible spoken or written answers as carried out successfully and satisfyingly. In other words, each discourse function and its assessment were oriented towards an individual horizon of expectations. The validity of the assessment was ensured through peer coding as described in chapter 5.3. The following description of a task analysis exemplifies the applied process.

The task was taken from *Exploring History 1*, chapter 1 on *Absolutism and Enlightenment.* The text required to work the task was concerned with *The Three Estates* in French society at the time of Louis XIV. The task was "Explain the structure of French society and find out what the chances of upward mobility were." (Kröger & Lohmann 2007: 11, task 1). As a first step, the two discourse functions were retrieved and defined in accordance to the abovementioned mapping table. In History, EXPLAIN is defined as "Tell something in a way that helps to understand it better" (ibid.: half-title) and is allocated with AFB 3 (see chapter 4.2). FIND OUT is defined as "Discover a fact or piece of information which is not explicitly mentioned" (ibid.) and is a discourse function of AFB 2. Based on those definitions, the next step was to decide as to whether *reading for orientation* or *reading for information and argument* applied. To assign the text and its reading demand to standardized categories (see chapter 5.2), i.e., the CEFR levels, the four paragraphs on the topic had to be analyzed. The various criteria focused on length, level of complexity and lexical sophistication, the level of depth to which the text had to be understood, the quality of information presentation, the recipient's familiarity with the topic, and the type of discourse. Given the level of demand implied with the AFB*,* and the definition of the discourse functions, a rather detailed level of understanding of this longer text (approximately 400 words) was required. Accordingly, the can-do-statement to be found with *reading for information and argument* at level B1 fitted stating that the reader "can recognize significant points […], can understand most factual information" (Council of Europe 2018: 63). Basically, the can-do-statements are worded in such a way that this allocation was a straight forward decision for most cases. With the level of lexical sophistication rated at CEFR level B2 during the previous study (see Table 7.1), this also influenced the evaluation of the required reading skill. With direct and straightforward information in the text, reading for this part of the task was rated at B1 since the reader "can understand straightforward, factual texts […]" (ibid.). Information for the second discourse function, FIND OUT, was not directly given, which required the reader to "[…] obtain information, ideas and opinions" (ibid.) describing level C1 skills. Given that this subchapter was only the second on the topic justifies the assumption that students were not very familiar with the topic of French society as such. However, students may recollect the structure

of society from previous History lessons during the Middle Ages. Consequently, the evaluation was based on the presumption of a certain familiarity with the general topic and the genre of textbook material, leading to the allocation of level B1 again. The can-do-statement "Can understand straightforward, factual texts on subjects relating to their interests or studies" (Council of Europe 2018: 63) was agreed upon through peer coding in a group of experts (see chapter 5.3) as the most appropriate fit for the part concerned with the explanation. For the part concerned with finding out about the upward mobility to which no direct information was given and which required students to work with prior knowledge and information obtained by interpreting the text, the CEFR level C1 was agreed upon by the involved coders to fit best. In order to be able to answer this task satisfyingly, students would have had to "understand in detail […] complex text […] identifying finer points of detail […]" (ibid.).

"Explain the structure of French society and find out what the chances of upward mobility were." (Exploring History 1, subchapter 1.2, task 1, p. 11)				
Discourse function	Definition	Skill	Can-Do-Statement	CEFR-Level
Explain	Tell something in a way that helps to understand it better	Reading for information and argument	Can understand straightforward, factual texts on subjects relating to their interests or studies	B1
Find Out	Discover a fact or piece of information which is not explicitly mentioned	Reading for information and argument	Can understand in detail […] complex text […] identifying finer points of detail […]	C1

Table 8.2 - Exemplary evaluation of reading skills in task 1, Exploring History 1, p.11

After evaluating the reading requirements imposed by the tasks and their discourse functions, the spoken production was evaluated. The requirements were assessed against the expected learner language, meaning what level of language students would have to display to sufficiently and satisfyingly answer the tasks. For both discourse functions, the category *sustained monologue: giving information* applied. To describe the structure of French society, students would have to give information in a linear sequence with an increased level of detail and variation, and the information given would have to be increasingly precise. Although possibly talking about a remotely familiar topic, students would have to use vocabulary of an increased level of lexical sophistication. Consequently, the can-do-statement found with level B1 for this category would be the best fit. It states that speakers "can explain the main points in an idea or problem with reasonable precision" (Council of Europe 2018: 63). As for the second discourse function of this task, the level of complexity increased since, on the one hand, the information was not directly given and could thus not be simply reproduced. Moreover, on the other hand, results would have to be presented as an interpretation of the actual information given, requiring a certain extent of reasoning when answering that part of the task.

Hence, the descriptor with its can-do-statement given with level B2 seems appropriate, requiring speakers to "[…] communicate complex/detailed information reliably." (ibid.).

"Explain the structure of French society and find out what the chances of upward mobility were." (Exploring History 1, subchapter 1.2, task 1, p. 11)				
Discourse function	Definition	Skill	Can-Do-Statement	CEFR-Level
Explain	Tell something in a way that helps to understand it better	sustained monologue: giving information	can explain the main points in an idea or problem with reasonable precision	B1
Find Out	Discover a fact or piece of information which is not explicitly mentioned	sustained monologue: giving information	can communicate complex/detailed information reliably	B2

Table 8.3 - Exemplary evaluation of speaking skills, task 1, Exploring History 1, p. 11

Usually, recording results, findings, or arguments gathered throughout a lesson occurs. Hence, the evaluation of written production was the third step in the analysis for most of the 400 tasks analyzed, a smaller number of tasks did not call for a written part and hence, were not included in this step. Based on the typical kind of how results would be recorded for each discourse function, the expected written products were evaluated using the criteria compiled from the CEFR companion. With the choice of categories given with written production, both discourse functions fell in the category of *written reports and essays* since no creative aspect was involved. To sufficiently report on the description of the French society, written products would "summarize, report […] about accumulated factual information […]" which describes skills at level B1 (Council of Europe 2018: 77). The report on the findings concerned with upward mobility would require writing skills at level B2, as students would have to "[…] synthesize information […] from a number of sources." (ibid.).

"Explain the structure of French society and find out what the chances of upward mobility were." (Exploring History 1, subchapter 1.2, task 1, p. 11)				
Discourse Function	Definition	Skill	Can-Do-Statement	CEFR-level
Explain	Tell something in a way that helps to understand it better	written reports and essays	Can summarize, report […] about accumulated factual information […]	B1
Find Out	Discover a fact or piece of information which is not explicitly mentioned	written reports and essays	Can synthesize information […] from a number of sources.	B2

Table 8.4 - Exemplary evaluation of writing skills, task 1, Exploring History 1, p. 11

The results of the individual assessments were further used to compare the demand between discourse functions and the four subjects. This process is described in the following chapter.

8.2.3   Averages and relative frequencies

In order to compare distinct discourse functions and their demands across the subject aiming at confirming the sequence of subjects found in chapter 7.5.6, it was necessary to calculate average CEFR levels and derive the maximum relative frequency of distinct CEFR levels assigned with all individual assessments per discourse function.

Each of the 400 tasks featuring 927 discourse functions were assessed and evaluated following the three steps (see chapter 8.2.2) in an attempt to apply the methodology of "understand[ing] the [skill-related] situations through the eyes of the participants [students and teachers]" (Cohen, Manion & Morrison 2011: 32 abridged). After assigning a CEFR level to each skill per discourse function, those levels were again translated to numeric values using the same scale as the textbook analysis (see chapter 6.4.2). This substitution of point values allowed for calculating average CEFR levels per skill. To find a CEFR level for the distinct discourse function across the subject, the maximum relative frequency was identified (see chapter 8.2.3). Calculations like these were relevant to arrive at a comparable result for each entity since, in another step, those CEFR level results were compared to the curricular standards. This was necessary to determine whether the level required in a specific subject and with individual discourse functions is appropriate for students at the given grade and stage of language learning. This part consisted of four steps. As a first step, the averages per skill and discourse function were calculated, marked in Figure 8.3 as

- 1a (for reading),
- 1b (for speaking),
- and 1c (for writing).

discourse function	skill			max. relative frequency per discourse function (2)
	avg. per discourse function and reading (1a)	avg. per discourse function and speaking (1b)	avg. per discourse function and writing (1c)	
EXPLAIN	2 →			4
DESCRIBE	3			
…				
average per skill across discourse functions per subject (3)	4 →			max. relative frequency across skills and discourse functions per subject (4)

Figure 8.3 - Steps of calculating averages and computing relative frequencies with the task analysis

Additionally, to arrive at an assessment of the discourse function as such, the maximum relative frequency was detected including all assessments across the three skills per discourse function. The CEFR level with the highest relative frequency indicates the demand for the discourse function since this specific level was needed most often in connection to the discourse function (the process is

represented by the arrow no. 2 and the result is marked as 2 in Table 8.6 and all following tables). Furthermore, as the third step (arrow no. 3), the average CEFR level of a single skill across all discourse functions used in the selected material per subject was calculated, i.e., the average of all CEFR levels assigned for, e.g., reading within the corpus of tasks per subject was computed. So, regardless the discourse function, any assessment for the individual skill was included in this step (3). Lastly, to obtain an assessment of all discourse functions and derive an overall demand for the material and its tasks as a whole, the relative frequencies for all assessments was calculated (arrow no.4). Again, the CEFR level with the highest relative frequency indicates the overall level of demand for the material and its tasks. In summary, the averages are as follows:

- average per skill and discourse function, that is all reading connected to, e.g., DISCUSS (1a-c)
- average across all discourse functions with a single skill, e.g., for reading across all discourse functions (3),

furthermore, relative frequencies are calculated for

- the level of demand across all skills for distinct discourse functions (2)
- the level of demand across all discourse functions and skills based on the material per subject and grade, e.g., Mathematics grade 7 (4).

These four steps allow for comparisons on a skill-discourse function level, the discourse function as such, the skills as such, and for the subject's demand in connection to tasks.

Another comparison across the subjects was necessary to develop a sequence of subjects, similar to the sequence yielded from the textbook analysis. This part required a selection of discourse functions to be compared to one another. Overlapping discourse functions were selected for this and are further described below in 8.4. This comparison follows the idea of a gap analysis, comparing a preset standard, the curricular standards, to the level of expected skills with the help of triangulation (cf. Schramm & Knorr 2016) and identify differences between the targeted standard and the required level of skill to work the tasks given with the teaching material. This triangulation employs the combination of methods, i.e., the initial assessment of the tasks for their three skills, the calculation of average values and identification of the maximum relative frequency based on the assigned CEFR level, and the comparison of the determined levels to the set curricular standards. Only after having run through all of the steps described above, an answer to RQ III, IV, and lastly RQ I is possible for this part of the study.

It must be noted, though, that tasks may not require all three skills. Selected tasks only called for reading and speaking, while others may only have required speaking and writing, especially when paintings were involved.

8.3   Results

To describe the analysis results conducted for all 400 tasks individually would undoubtedly exceed the limits of this thesis. Hence the description of the results obtained accumulates the individual findings per task on the basis of the distinct discourse function. An overview of the results, that is the assignment of CEFR levels per skill and discourse function can be found in chapter 12.8.

8.3.1   History

The quantitative collection of the discourse functions in *Exploring History 1* started with analyzing 97 applications of discourse functions. The analyses were conducted by the author as well as a group of experts for the subjects selected as described in chapter 5.3.

The comprehensive overview below (see Table 8.5) provides a list of all discourse functions employed and their definitions to help understand the demand each discourse function imposes on students.

The tasks used 11 different discourse functions, of which three were from AFB 1, five from AFB 2, and three from AFB 3. Out of the 97 discourse functions employed, the smallest part was from AFB 1, with 24 uses of those discourse functions. AFB 2 discourse functions were applied 44 times, while 14 discourse functions were from AFB 3, imposing the highest level of difficulty. This collection also serves as the starting point for evaluating each skill engaged in the respective discourse function. Based on the definition and the preliminary evaluation (see chapter 8.2.1 and chapter 12.6), each skill has to be considered following the above-described list of criteria per CEFR level and following the coding manual.

Discourse Function	Definition (Kröger & Lohmann 2007: half-title)	AFB	Frequency
LIST	Write down information in a given order without commenting on it	1	6
DESCRIBE	Give details about what something or someone is like	1	12
SHOW	Give information to be found in other material (sources, maps, etc.)	1	6
FIND OUT	Discover a fact or piece of information which is not explicitly mentioned	2	9
COMPARE	Put facts next to each other and decide what they have in common or what differentiates them – which factors are the same, similar, different or opposite	2	9
ANALYZE	Study or examine something in detail in order to understand and explain it	2	10
ASSESS	Carefully consider a situation, person or problem to make a judgement	2	13
EXPLAIN	Tell something in a way that helps to understand it better	2	18

DISCUSS	Talk about something in detail, coming to a reasoned opinion	3	9
ARGUE	Give reasons why you believe that something is wrong/tight or true/false	3	3
GIVE AN OPINION	State a reasoned opinion about a (historical) fact in this context	3	2

Table 8.5 - List of discourse functions, definitions used in Exploring History 1

<u>Exploring History 1</u>

For the discourse function LIST, the required skill or subskill of reading would be *reading for orientation*, in specific "scanning for specific information" or "identifying important information" (Council of Europe 2018: 62). Based on the definition of LIST, this activity requires nothing more than to read for single, specific terms. To filter texts for such specific information, students in History have to read longer text with increasing complexity. Presumably, the topic would be familiar. The texts to be read for the tasks employing LIST were composed of either straightforward or indirect information, of which the factual, relevant points have to be understood. Since usually only single expressions are required to carry out LIST successfully, only single terms have to be understood, hinting at the required level of understanding being comparably cursory. The scale for *reading for orientation* features a descriptor as a can-do-statement fitting the required level of skill "Can scan longer texts in order to locate desired information, and gather information from different parts of a text, or from different texts in order to fulfil a specific task." (ibid.) which describes level B1. In cases in which the information is not directly or straightforwardly given, *reading for information and argument* would be the more appropriate skill to apply since here the can-do-statement "can obtain information, ideas and opinions from highly specialized sources within their field" (Council of Europe 2018: 63) seems fitting. This can-do-statement then corresponds to level B2. In addition to the information that has to be obtained from the text, the genre may also influence the evaluation. Cases that require students to obtain information from source texts, often for statements of opinions or arguments, the required skill is better described with the latter can-do-statement as *reading for orientation* does not feature a fitting descriptor. By definition, LIST is usually a written task only. Accordingly, LIST was only evaluated on the level of difficulty for written production. Depending on what kind of list is required, either the can-do-statement taken from overall written production at level A1 suffices, saying that students "Can write simple, isolated phrases and/or sentences" (ibid. 2018: 75) or, if a list requires complete sentences, level A2 of written reports and essays is appropriate as students "Can write simple texts on familiar subjects of interest, linking sentences with connectors like 'and', 'because' or 'then'." (ibid. 2018: 77). In rare cases, the required level of skill in writing can be even above A2. There was a task, e.g., that required students to "Compare the two maps closely and list the changes" (Kröger & Lohmann 2007: 65, task 6). For this task, students would have to write elaborate, connected sentences with increased detail. Students would further have to summarize their findings in their own words, i.e., in simple[r] language. In those cases,

level B1 would be appropriate as students "can write straightforward connected text […] by linking a series of shorter discrete elements into a linear sequence" (Council of Europe 2018: 75) as stated with overall written production. Evaluating all six tasks that used the discourse function LIST in accordance to the requirements and criteria as described above, the average CEFR level for LIST in reading (see 1a, Table 8.6) arrived at B1. For the skills concerned with writing (1c), tasks that featured LIST required students to display writing skills of A2. The CEFR level assigned most often in relation to the total number of assessments in connection to LIST across all skills (2) arrived at B1 with 55% of all assessment being at that level.

LIST	Skill			max. relative frequency (2)
	Reading (1a)	Speaking (1b)	Writing (1c)	
	B1	-	A2	B1

Table 8.6 - CEFR level for LIST, per skill and max. relative frequency across skills, Exploring History 1

The second discourse function within AFB 1 is DESCRIBE. Students here "give details about what something and someone is like" (Kröger & Lohmann 2007: half-title). In laypersons' terms, this would refer to objects mainly, but in History, DESCRIBE also refers to texts in which procedures, experiences, or living conditions are the central issue. Accordingly, the skill required in particular is *reading for information and argument*, since understanding at a greater level of detail is usually required to DESCRIBE rather than listing things. Again, depending on the length of text, the level of complexity, the required level of understanding, and the manner of information provision, the respective CEFR levels are assigned. For DESCRIBE, levels A1, stating that students "can understand short texts on subjects of personal interest […]", to B1, "can understand descriptive notes […]" (Council of Europe 2018: 63), can be assigned. Both can also occur with objects or images to be described, provided with annotations or short, informative texts. With the spoken production in connection with DESCRIBE, mostly can-do-statements of *sustained monologue: giving information* fitted the task and thus, supplied the appropriate level of requirement. Since describing a picture or artifact can also be regarded as an experience, descriptors of the corresponding subskill *sustained monologue: describing experience* were taken into consideration as well. This subskill of spoken production features a can-do-statement that mostly fitted the required skill level in its B1-level statement "can give straightforward descriptions on a variety of subjects within their field of interest" (ibid.: 70). In cases in which a greater level of detail was required, the assigned CEFR level was raised to B2 as a result since this level's can-do-statement explicitly mentions "a high level of detail" (ibid.). The requirements connected to the written aspect of DESCRIBE are mostly covered with *overall written production* or *written reports and essays*, the assigned CEFR level depended on the amount of text, the level of detail, and the required quality of language. As descriptions usually require more than just a short string of sentences, level B1 and the corresponding can-do-statement "Can write straightforward connected texts on a range of familiar subjects within his/her field of interest, by linking a series of shorter discrete elements into a linear sequence." to be found with *overall written production*. Alternately, "can write a text on a thematic subject of personal

interest, using simple language […]" of *written reports and essays* seemed appropriate. Hence, the average CEFR level for the written part of DESCRIBE (see Table 8.7, 1c) arrived at level B1 for both *overall written production* as well as *written reports and essays.* For oral production with DESCRIBE*,* the average CEFR level arrived at B1 for both *sustained monologue: giving information* and *describing experience* (1b). For reading to be able to give any description at all, the average CEFR level was for both subskills A2 (1a). The discourse function arrived at CEFR level B1 as the max. relative frequency across all skills and subskills (2).

	Skill			
DESCRIBE	Reading (1a)	Speaking (1b)	Writing (1c)	max. relative frequency (2)
	A2	B1	B1	B1

Table 8.7 - CEFR level for DESCRIBE, per skill and max. relative frequency across skills, Exploring History 1

To act on the discourse function SHOW, students "give information to be found in other material" (Kröger & Lohmann 2007: half-title). Source texts, images, photographs, or maps can contain this information. Therefore, both *reading for orientation* and *reading for information and argument* may be the required subskills for this discourse functions. As was the case with reading skills above, the same criteria apply. Especially source texts can prove to be quite challenging since information may not be given directly, and additionally, students have to recognize when a source text represents a biased opinion or is purely factual. Consequently, the required skill level can range from level A2 for both variants to B2 for *reading for information and argument.* The latter applies especially to those tasks which require extensive reading of source text material. The required skills for oral production for this discourse function were mainly of the subskill of *sustained monologue: giving information* and can range from A1, the simple description of things in basic words and with the help of visual support (cf. Council of Europe 2018: 71), to B1, the straightforward reporting of factual information with an increased level of detail and reliability (ibid.). A single task in *Exploring History 1* required activities deviating from those associated with the definition for SHOW. The task "Show what Krupp's instructions on how to take pictures tell you about environmental pollution in the 19th century." (Kröger & Lohmann 2007: 69, task 2a) rather asked students to list and substantiate their findings. Therefore, the subskill for oral production, in this case, was *sustained monologue: putting a case*. Finally, the discourse function required *written reports and essays* at level B1, meaning straightforward connected text in linear sequence with increased detail. Furthermore, students would have to be able to use simple language to compare relevant aspects. After evaluating all six tasks featuring the discourse function SHOW, the average CEFR level (see Table 8.8, 1a) for *reading for information and argument* was B1. At this level, students "[…] understand straightforward, factual texts on subjects relating to their Interests or studies" (Council of Europe 2018: 63). Further, the average CEFR levels for successfully carrying out the oral or communicative part (1b) of this discourse function averaged at A2+. Apart from giving a simple description students would have to demonstrate increased fluency and "explain main points with reasonable precision" (ibid. 2018: 73). Lastly, the *written reports and essays* (1c) that

rounded of the tasks for this discourse functions were averaged at CEFR level B1 as the minimum level expected or required to "summarize, report or give their [students'] opinion about accumulated factual information […]" (ibid.: 77). The maximum value for relative frequency with SHOW was at CEFR level B1 with 57% of all uses across all skills were assessed at that level.

	Skill			
SHOW	Reading (1a)	Speaking (1b)	Writing (1c)	max. relative frequency (2)
	B1	A2 +	B1	B1

Table 8.8 - CEFR levels for SHOW, per skill and max. relative frequency across skills, Exploring History 1

After assessing all discourse functions located in AFB 1, the analysis continued in the same manner with the four discourse functions of AFB 2. The first discourse function listed in *Exploring History 1* is FIND OUT. This activity requires students to come to a deeper understanding of the text or material given since it asks to "Discover a fact or piece of information which is not explicitly mentioned" (Kröger & Lohmann 2007: half-title). Finding implicit information in a text requires students to work with increasingly longer and more complex texts, recognize both direct and indirect information, distinguish one from the other, and understand the main points. This requirement is best described in a can-do-statement in *reading for information and argument* at level B2. Students would be able to "[…] recognize different structures in discursive text, […] recognize when a text provides factual information and when it seeks to convince readers of something" (Council of Europe 2018: 63). The level of understanding would have to be to such an extent that students can manage more specialized sources and particular stances adopted by the authors of such texts (ibid.). However, looking at the specific tasks and the associated texts, the required information was mainly directly given in both chapters of *Exploring History 1*, erasing the aspect of indirect information. Only two out of 10 tasks employing FIND OUT actually required students to pick up on indirect information. Hence, the average level of required reading skills (see Table 8.9, 1a) arrived at level B1. To elaborate verbally on the results stemming from the activity FIND OUT, students would have to engage in a *sustained monologue: giving information*. The required level of skill was between A1 and B2, from single words and phrases to increased fluency, accuracy, reliability, and complexity with the analyzed tasks. On average, the required skill level for *sustained monologue: giving information* (1b) was B1, asking students to "[…] explain the main points in an idea or problem with reasonable precision" or to "[…] report straightforward factual information […] provided they can prepare beforehand" (Council of Europe 2018: 71). A single task based on a source text also required skills of *sustained monologue: putting a case* since students had to elaborate on a particular point of view conveyed by the author. Students had to display speaking skills at level B2 since they had to "[…] explain a viewpoint on a thematic issue […]" (ibid.: 72). Putting all of their findings in writing in *written reports and essays*, would in general require students to display writing skills at level B1. This level was also mainly demanded with *Exploring History 1* and the tasks concerned with FIND OUT as the activity. Consequently, the average CEFR level for *written reports and essays* (1c) arrived at B1+. To record results in writing, students

would have to "[…] develop[s] an argument […] with appropriate highlighting of […] relevant supporting detail" (ibid.: 77). All three skills and their respective subskills required students to display language competences at CEFR level B1 (2) to successfully work tasks concerned with the discourse function FIND OUT. The maximum relative frequency was at 52% of all assessments at B1.

	Skill			
FIND OUT	Reading (1a)	Speaking (1b)	Writing (1c)	max. relative frequency (2)
	B1	B1	B1+	B1

Table 8.9 - CEFR levels for FIND OUT per skill and max. relative frequency all skills, Exploring History 1

The second discourse function of AFB 2 is COMPARE, which is to be carried out in its literal sense "put facts next to each other and decide what they have in common or what differentiates them – which factors are the same, similar, different or opposite" (Kröger & Lohmann 2007: half-title). To obtain facts to compare, however, students have to demonstrate skills in *reading for information and argument* that allow for an increased level of understanding of a usually longer and more complex text. This kind of text provides straightforward information, of which students are able to understand factual and relevant points. Depending on the required level of detail, either CEFR level B1 or B2 would be called for when reading to obtain information. To discuss said comparisons, all four subskills of spoken production can apply. Some use cases of COMPARE may require *sustained monologue: giving information*, e.g., when describing differences between two aspects. In such cases, CEFR level B1 would be the minimum required level of skill, as students here supposedly "can report straightforward information on a familiar topic […]" (Council of Europe 2018: 71). Again, in cases in which more details are of essence, speaking skills at level B2 could be necessary. Moreover, teachers may ask students to substantiate their findings by arguments and thus, students have to employ communicative skills of arguing to be found with *sustained monologue: putting a case.* For this argument, a certain level of detail when describing and justifying the results of a comparison is necessary. Level B1 provides the most fitting can-do-statement, which states that students "can briefly give reasons and explanations for opinions, plans, and actions" or "can develop an argument well enough to be followed without difficulty" (ibid.: 72). As was the case above, this subskill could also require level B2 with more complex arguments to be supported by finer details. Especially comparisons concerned with images, objects, or paintings could also ask students to use skills of *sustained monologue: describing experience* as their findings would be visual observations. In such cases, students would have to demonstrate an increased level of detail, lexical sophistication, and precision when describing their experience. Among the can-do-statements listed with this subskill, level B1 fitted the requirements imposed as students "can give straightforward descriptions on a variety of familiar subjects within their field of interest" (ibid.: 70). Putting a comparison in writing can be more challenging. Correspondingly, CEFR level B2 would be the most appropriate when it comes to *written reports and essays* as students would have to "[…] synthesize information and arguments […]" or "[…] write an essay or report that develops an argument […] with appropriate highlighting of significant points and relevant supporting detail." (ibid.: 77). The

evaluation of the nine tasks featuring the discourse function COMPARE showed that for *reading information and argument* the average CEFR level required was at level B1 (see Table 8.10, 1a). On average, all four subskills of speaking (1b) arrived at the same level with only one task requiring speaking skills to be at level B2 as it dealt with a particularly complex set of source texts. The above-described preliminary considerations on possible CEFR levels per subskill were accurate in the case of *written reports and essays* for which the average level was calculated at level B2 (1c). The overall required level of language skills for this discourse function across all three categories (2) was B1 with 68% of all assessments marking the maximum relative frequency.

COMPARE	Skill			
	Reading (1a)	Speaking (1b)	Writing (1c)	max. relative frequency (2)
	B1	B1	B2	B1

Table 8.10 - CEFR levels for COMPARE per skills and max. relative frequency across skills, Exploring History 1

While COMPARE asks students already for an increased level of understanding and the ability to work with more complex texts and information, the discourse function ANALYZE increases those demands even further. By definition, students are tasked with "study[ing] or examin[ing] something in detail in order to understand and explain it" (Kröger & Lohmann 2007: half-title). Mostly, this calls for *reading for information and argument.* Depending on the complexity level, the genre and length of the texts, the manner of information presentation, and the familiarity of the topic, levels B1 to C1 could apply. While with level B1, students would have to be able to "[…] understand most factual information […]", level B2 already requires an increased level of understanding especially, when "[…] specialized sources within their field" (Council of Europe 2018: 63) are concerned. When texts of particular length and/or complexity have to be analyzed, level C1 skills would be called for since students are expected to "[…] understand lengthy, complex texts […] provided they have time to reread difficult sections" (ibid.). Asking for detailed explanations by definition raised the expectations towards oral production accordingly. Students would most likely have to display skills in the category of *sustained monologue: giving information* at level B2 or C1. The same applied to *sustained monologue: putting a case.* The decisive factor in both categories would be the level of detail needed to present the findings of the activity verbally sufficiently. While level B2 asks for, i.a., "[…] detailed information" communicate reliably (Council of Europe 2018: 71) and "[…] relevant supporting detail" (ibid.: 72), level C1 extends the skill to be able to give "[…] clearly detailed distinctions […]" (ibid.: 71) or "[…] clear, detailed descriptions" (ibid.: 69). Consequently, *written reports and arguments* would be equally demanding and thus, would most likely call for writing skills of levels B1 through C1, spanning from "[…] summariz[ing] accumulated factual information […]" through "[…] synthesizing information and arguments […]" to "[…] expand[ing] and support[ing] points of view at some length with subsidiary points, reasons and relevant examples" (Council of Europe 2018: 77).

The analysis of the ten tasks concerned with the discourse function ANALYZE proved all those considerations and preliminary allocations of CEFR levels with the subskills were appropriate. Only

one task displayed a comparably low demand for reading as it dealt with an overview of working hours during the 19th century presented as a table. Consequently, the average CEFR level for *reading for information and argument* (see Table 8.11, 1a) arrived at B1+ as the level of expected language skills. Verbalizing the findings of the respective analyses required an average minimum level of spoken production of B2 (1b) for *sustained monologue: giving information, overall spoken production,* and *sustained monologue: describing experience*. Lastly, *written reports and essays* recording the results of each of the ten tasks was averaged out at level B2 (1c), with only a smaller number of tasks requiring level B1 but none asked for such an increased level of detail that would have called for level C1 writing skills. This discourse function did not have one clear maximum relative frequency. CEFR levels B1 and B2 were both assigned for 40% of all assessments (2). In the light of the fact that the maximum relative frequencies for the remaining discourse functions increased to level B2, this discourse function and its equal maximum relative frequency at B1 may indicate a threshold.

ANALYZE	Skill			
	Reading (1a)	Speaking (1b)	Writing (1c)	max. relative frequency (2)
	B1+	B2	B2	B1 & B2

Table 8.11 - CEFR levels for ANALYZE per skill and max. relative frequency across skills, Exploring History 1

The fourth discourse function of AFB 2 is ASSESS, an activity aimed at schooling students in critical thinking as they "carefully consider a situation, person or problem to make a judgment" (Kröger & Lohmann 2007: half-title). To make a judgment, a deeper understanding of details in texts is essential, especially when texts are increasingly complex and provide direct and indirect information. Accordingly, CEFR levels B1 to C1 could be appropriate when assessing the required skills for *reading for information and argument*. Furthermore, the length and genre of the text determine the decision for the actual CEFR level. As assessments always call for reasons and substantiation, the category used most in spoken production is *sustained monologue: putting a case* alongside *sustained monologue: giving information*. Depending on the required level of detail and precision as well as the degree of reasoning, the CEFR levels in scope to sufficiently fulfill the requirements of this discourse function can range from B1 to C1. Usually, ASSESS also demands for a written summary of findings via *written reports and essays*. Again, the most fitting CEFR levels to successfully master this skill could be between B1 and C1, depending on the required composition of the texts, either in linear sequence or with a systematically developing structure and the extent to which corroborative arguments are necessary. Following the preliminary considerations, the actual evaluation revealed that the 13 tasks concerned with the activity of assessment required an average reading competence at CEFR level B2 (see Table 8.12, 1a), which may also be caused by the fact that seven out of those were concerned with assessments of statements or opinions given in source texts, e.g., an excerpt of Montesquieu's *On the spirit of the Laws* of 1748 (cf. Kröger & Lohmann 2007: 21, M3). Furthermore, most tasks did not ask for the assessment of one particular part of the texts or source but for all texts and sources in context, demanding reading skills capable of understanding increasingly complex texts and details. Likewise,

the verbal exchange necessary to understand such texts and arrive at an assessment proved to be as demanding as the reading. For both categories, *sustained monologue: giving information*, applied with four tasks, and *sustained monologue: putting a case*, applied with all 13 tasks, the average required CEFR level was B2 (1b). Finally, to put the findings into *written reports and essays* an average CEFR level of B2 (1c) was calculated again, expecting "[…] essay[s] and report[s] that develop[s] an argument systematically with appropriate highlighting of significant points and relevant supporting detail" (Council of Europe 2018: 77). Consequently, the maximum relative frequency across all tasks featuring ASSESS and all categories of the three skills also arrived level B2 (2).

	Skill			
ASSESS	Reading (1a)	Speaking (1b)	Writing (1c)	max. relative frequency (2)
	B2	B2	B2	B2

Table 8.12 - CEFR levels for ASSESS per skill and max. relative frequency across skills, Exploring History 1

EXPLAIN marked the end of AFB 2, and an increase in demand could be observed with averages increasing from mainly B1 to B2 in the process. This discourse function is an activity that calls for students to "tell something in a way that helps to understand it better" (Kröger & Lohmann 2007: half-title), which makes a deep understanding of a text, the main points, and finer details necessary. Hence, when *reading for information and argument* the levels B1 to C1 could be appropriate, the actual level depending on the complexity and length of the text as such. When asked to explain something so that a peer can understand the aspect, the demand for oral production is accordingly. Hence, *sustained monologue: giving information* has to respect this demand and has to be between B1 for "[…] to explain the main points in an idea or problem with reasonable precision" and C1 to "[…] communicate clearly detailed distinctions between ideas, concepts and things that closely resemble one another […]" (Council of Europe 2018: 71). Furthermore, reasoning and corroboration could be necessary, which addresses *sustained monologue: putting a case.* Presumably, this subskill would have to be at level B2, as "[…] a chain of reasoned argument" has to be developed (ibid.: 72). Written production would either have to be at CEFR level B2, ensuring "[…] detailed descriptions of a complex process" (ibid.: 77) or expecting "[…] clear detailed texts on a variety of subjects […] synthesizing and evaluating information […]" (ibid.: 75). Cases elaborating on increasingly refined details or "[…] subsidiary points, reasons and relevant examples" (ibid.:77) could also be rated at level C1. With the horizon of expectations for the tasks featuring EXPLAIN defined, the analysis looked at 18 tasks in *Exploring History 1.* Out of those 18, 13 tasks required reading, while five were concerned with information obtained from paintings or images. Accordingly, the average CEFR level for reading arrived at B2 (see Table 8.13, 1a). Formulating the actual explanation in the sense of *sustained monologue: giving information* was averaged out at level B2 (1b) since most tasks called for detailed and complex information to be "communicated reliably" (ibid.:71). Additionally for cases which required reasoning, the CEFR level was also B2 on average with *sustained monologue: putting a case.* Lastly, the *written reports and essays* arrived at an average demand of B1+ (1c). Determining the demand for the discourse function across all

three categories of skills, the calculation of the relative frequencies had its maximum at level B2 again (2).

	Skill			
EXPLAIN	Reading (1a)	Speaking (1b)	Writing (1c)	max. relative frequency (2)
	B2	B2	B1+	B2

Table 8.13 - CEFR levels for EXPLAIN per skill and max. relative frequency across skills, Exploring History 1

Continuing with AFB 3, which encompasses discourse functions with the highest cognitive demand (see chapter 4.2), a further increase in linguistic demand can be expected. The first discourse function in this part is DISCUSS, a classroom activity that requires students to "talk about something in detail, coming to a reasoned opinion" (Kröger & Lohmann 2007: half-title). However, to partake in a discussion as a competent interlocutor, the essential information must be obtained through reading beforehand. Thus, the skill of *reading for information and argument* would have to be between CEFR levels B1 and C1, depending on the level of detail, the quality of the text, and the length and complexity. The verbal part of this discourse function may be a little more demanding since the obtained information has to be processed and transferred into arguments and subsidiary points. Accordingly, *sustained monologue: putting a case* would have to be at level B2 as students would have to "[…] construct a chain of reasoned argument" (Council of Europe 2018: 72). Tasks asking for personal observations and opinions and hence, employ *sustained monologue: describing experience* would most likely be sufficiently carried out applying language skills at level B1. When summarizing or reporting on a discussion in *written reports and essays*, students would have to display either level B1 or levels B2 to C1 when asked to compose a written discussion. The latter requires an increased level of detail, the ability to develop a structure in writing systematically, and to give arguments to expand and support. *Exploring History 1* featured nine tasks that called for a discussion. Three of those did not entail a reading part as the discussion was concerned with paintings, e.g., the painting of Louis XIV by Hyacinthe Rigaud (cf. Kröger & Lohmann 2007: 16). The average CEFR level of the six tasks requiring *reading for information and argument* (see Table 8.14, 1a) arrived at B1+ due to an equal distribution of levels B1 and B2. In line with the preliminary considerations on demand imposed by the requirements in spoken production (see chapter 8.2.1), *sustained monologue: describing experience* was rated at level B1. *Sustained monologue: putting a case*, however, was distinctly more demanding arriving at level B2 Accordingly, speaking skills (1b) arrived at level B1+ on average. The written component for this discourse function arrived at an average level of B2 (1c), with one task rated at level C1, asking to discuss the intended message of the abovementioned painting. To corroborate arguments, students would have to "[…] expand and support points of view at some length with subsidiary points and relevant examples" (Council of Europe 2018: 77).

For the discourse function DISCUSS the maximum relative frequency was at 58% of all assigned CEFR level arrived at B2 (2).

	Skill			
DISCUSS	Reading (1a)	Speaking (1b)	Writing (1c)	max. relative frequency (2)
	B1+	B1+	B2	B2

Table 8.14 - CEFR levels for DISCUSS per skill and max. relative frequency across all skills, Exploring History 1

The following discourse function in AFB 3 is ARGUE, which is closely interrelated with *DISCUSS* but focuses even more on reasoning with the help of subject-related details. Students practice their skills by "Giv[ing] reasons why [they] believe that something is wrong/right or true/false" (Kröger & Lohmann 2007: half-title). Depending on the length and complexity of the text serving as base for such tasks, the level at which *reading for information and argument* should be mastered could range from B1 at which students "can understand short texts […] in which people give their points of view […]" (Council of Europe 2018: 63) to level C1 at which students would have to be able to understand "[…] lengthy, complex texts […] identifying finer points of detail including attitudes as well as stated opinions" (ibid.). The latter could apply with source texts such as *The Manifesto of the Communist Party* by Karl Marx (cf. Kröger & Lohmann 2007: 75). An argument in favor or against a point of view calls for the skill *sustained monologue: putting a case*. According to the preparatory mapping, the required skills could range from level B1, enabling students to "[…] give simple reasons to justify a viewpoint on a familiar topic" to level C1 "argu[ing] a case on a complex issue, formulating finer points precisely and employing emphasis effectively" (Council of Europe 2018: 72). When it comes to *written reports and essays,* the required CEFR levels may range from B1 at which students "can write a text on a thematic subject of personal interest, using simple language to list advantages and disadvantages, give and justify their opinion" to C1 with students being able to "expand and support points of view […] with subsidiary points and relevant examples" (ibid.: 77). *Exploring History 1* featured only three tasks with the discourse function ARGUE. Still, the analysis of the required reading skills (see Table 8.15, 1a) averaged at level B2. This level also applied to *sustained monologue: putting a case* (1b). Putting the argument in writing was evaluated to be equally demanding at level B2 (1c). The relative frequency was at a maximum of 86% of all assigned CEFR levels at level B2 (2) across all three categories of skills for this discourse function.

	Skill			
ARGUE	Reading (1a)	Speaking (1b)	Writing (1c)	max. relative frequency (2)
	B2	B2	B2	B2

Table 8.15 - CEFR levels for ARGUE per skill and max. relative frequency across all skills, Exploring History 1

The last discourse function, GIVE AN OPINION, initially sounded slightly less demanding than ARGUE, however, when looking at the definition used in History, this impression changes as students are to "state a reasoned opinion about a (historical) fact in this context" (Kröger & Lohmann 2007: half-title). Certainly, the discourse functions of AFB 3 overlap and share the aspect of reasoning and

justification. However, GIVE AN OPINION is by definition the only discourse function mentioning historical facts to be incorporated into the activity. This unique aspect most likely affects the linguistic demand that comes with this discourse function. When *reading for information and arguments*, students would have to be able to understand a more refined level of detail, distinguish between historical fact and opinionated writing, and would have to be able to deal with the more complex text as often such information is given in source texts. Regarding the latter as "[…] specialized articles within their field" (Council of Europe 2018: 63) would lead to level B2 as a requirement for reading in the context of GIVE AN OPINION. Furthermore, students could even be required to apply skills at level C1 if they were to "[…] identif[y] finer points of detail including attitudes and implied as well as stated opinions" (ibid.). To GIVE AN OPINION, students would have to display skills at level B2 for *sustained monologue: putting a case*. This level implies students "can construct a chain of reasoned argument" and "can explain a view point on a thematic issue giving the advantages and disadvantages of stated opinions" (ibid.: 72). Providing more detail when *putting a case* would increase the necessary level of skill to C1. Additionally, students would have to display skills of at least level B2 in the context of *written reports and arguments* as this level ensured that students "can write an essay or a report which develops an argument, giving reasons in support of or against a particular point of view […]" (ibid.: 77). Again, in cases that require a higher level of details and examples, level C1 could be necessary. The discourse function usually rounds of a set of tasks concerned with the same topic. Usually, tasks are echeloned from AFB 1 to AFB 3; accordingly, students would already have acquired a solid knowledge on the topic when asked to *give their opinion*. This assumed prior knowledge may be the reason why in *Exploring History 1* of the two tasks that asked for students' opinions, none featured a dedicated reading part. All information was obtained through reading for the tasks or subtasks beforehand. Accordingly, this analysis started with *sustained monologue: putting a case* arriving at an average of level B2 (see Table 8.16, 1b) since both tasks required students to "[…] explain a view point on a thematic issue […], and construct a chain of reasoned argument" (ibid.: 72). The necessary CEFR level to compose a *written report and essay* on their opinion was calculated to be at B2 (1c). Not surprisingly, the maximum relative frequency with all assessments for GIVE AN OPINION also arrived at level B2.

	Skill			
GIVE AN OPINON	Reading (1a)	Speaking (1b)	Writing (1c)	max. relative frequency (2)
	-	B2	B2	B2

Table 8.16 - CEFR levels for GIVE AN OPINION per skill and max. relative frequency across all skills, Exploring History 1

As was outlined in chapter 8.2.3 as steps three and four, one more average was calculated as well as a second maximum value for relative frequency was obtained to come to an impression on linguistic demand imposed by the tasks in *Exploring History 1*. The first average value was obtained across all discourse functions for the respective skill. Here, the average CEFR level demanded for reading arrived at B1+ across all 11 discourse functions, while speaking was slightly less demanding at level B1. Writing

then went up again to B1. The demand across all discourse functions and tasks, derived from the maximum relative frequency (4), was also B1 with 43% of all 248 assessments made within the chapter at that level.

Before moving on to the results obtained for the tasks featured in *Exploring History 2*, an overview of all average assessments per discourse function allows for a comparison among these and to the set, targeted CEFR level as given in the curricular standards (see chapter 2.2). Those are equally valid in this analysis on productive skills as they are for receptive skills. Likewise, a red-colored zone, similar to the illustrations used in chapter 7.5.2 on the results of the textbook analysis, marks the targeted CEFR levels. For a combined overview of all 11 discourse functions, the three main categories of skills, and the allocated CEFR levels, the discourse functions are placed on the y-axis while the CEFR levels are on the x-axis. A unique symbol represents each individual skill (see legend of Figure 8.4). The position of each symbol per discourse function on the grid expresses the results of the analysis. If the average value arrived at the elevated CEFR level, a plus sign was added to the symbol, i.e., for the average value calculated for the discourse function SHOW and speaking, the plus sign accounts for the level of A2+. However, the exact position, i.e., more to the left or right within the column for the CEFR level, does not imply increased complexity but is caused by placing more than one skill on the grid.
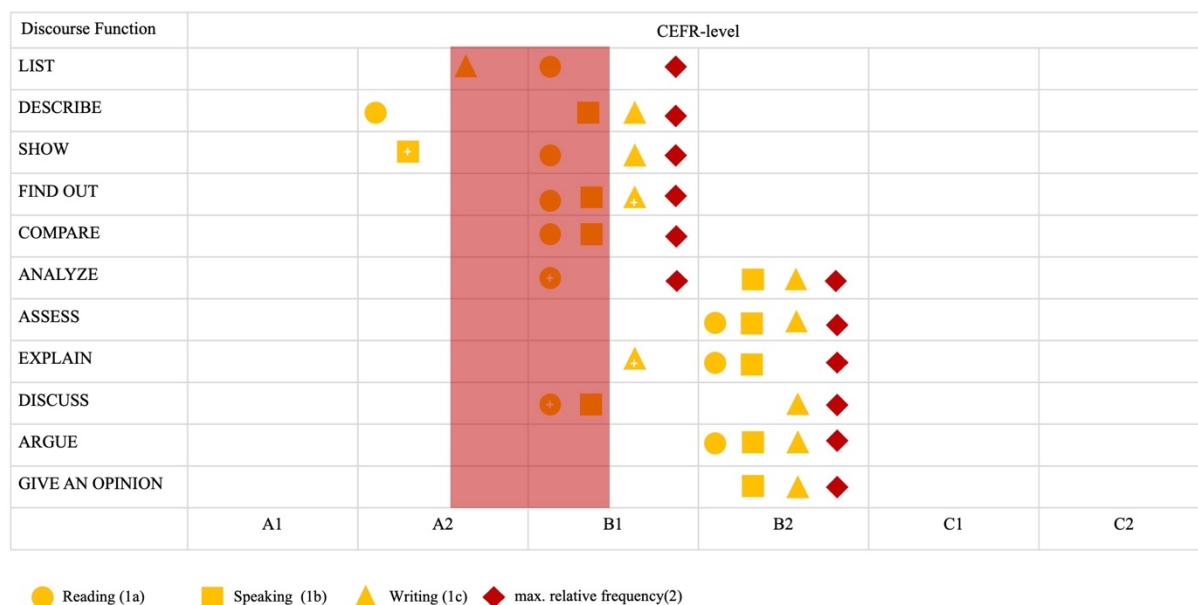
Figure 8.4 - *Average CEFR levels per discourse function and skill & max. relative frequency across skills, Exploring History 1*

Furthermore, the findings for the averages across the discourse functions and skills and the maximum relative frequency with all tasks in *Exploring History 1* are displayed in another grid below (see Figure 8.5).
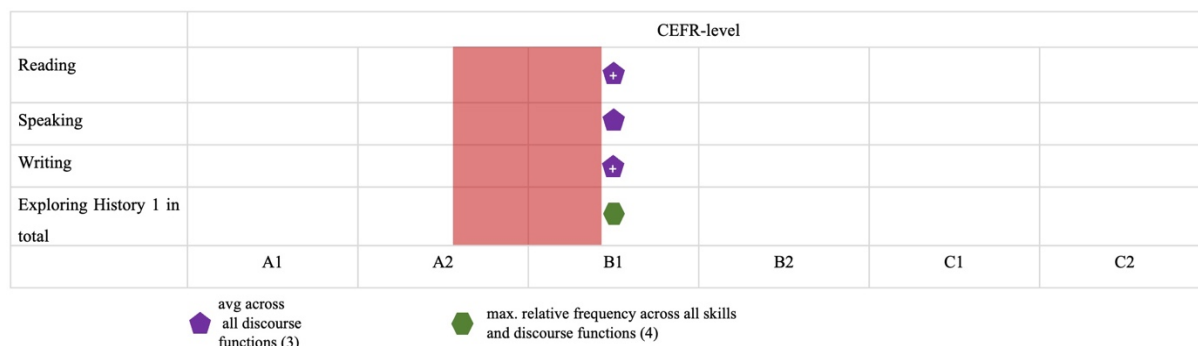
Figure 8.5 - Average CEFR levels per skill across discourse function and max. relative frequency of Exploring History 1

Exploring History 2

Continuing with *Exploring History 2,* designed for the higher age group of 15- to 16-year-olds, the same list of discourse functions was employed. Analyzing the number of discourse functions and their distribution across the three AFBs (see chapter 4.2)*,* there was a shift in the number of tasks from AFB 1 to AFB 2 in comparison to *Exploring History 1*. In total, 12 tasks employed discourse functions from AFB 1, 26 from AFB 2, and 13 from AFB 3. The total number of discourse functions in tasks analyzed was 51, significantly less than the 97 tasks in the first textbook. The concentration of tasks in AFB 2 and AFB 3 may serve as an explanation for this decrease as these are usually more complex and take longer to work through. Since *Exploring History 2* employs the same collection of discourse functions and definitions, the preliminary considerations on all possible CEFR levels applicable per skill and subskill are still valid. Therefore, they are left out of the following description of the results. Accordingly, only the average values and maximum relative frequencies calculated for all four steps (see 8.2.3) will be discussed.

The first discourse function LIST occurred twice. One task did not require reading as it was concerned with listing the provinces of former West and East Germany. The required map was not provided by *Exploring History 2* (cf. Kröger 2009: 91, 5) and thus, students would have had to obtain the information from other sources. The other task needed reading skills for *reading information and argument* at level B2, as the main points of the relevant but increasingly complex text had to be understood to answer the task (see Table 8.17, 1a). The spoken part came with lower demand. The first task asked for a simple listing of provinces and was thus rated at level A1. The second task required a sequence of sentences with simple connectors arriving at level A2. The average value for speaking was also A2 (1b). The written component of those two tasks, listing the provinces or the events of the Vietnam War chronologically, required skills of A1 and A2, respectively. The written production's

average value was A1+ (1c). The maximum relative frequency of CEFR levels across all skills for LIST was A2 (2).

	Skill			
LIST	Reading (1a)	Speaking (1b)	Writing (1c)	max. relative frequency (2)
	B2	A2	A1+	A2

Table 8.17 - CEFR levels for LIST per skill and max. relative frequency across all skills, Exploring History 2

The second discourse function in AFB 1 is DESCRIBE. Three tasks featured this discourse function in *Exploring History 2*. Here, the average CEFR level required for *reading for information and argument* was calculated to be at level B2 (see Table 8.18, 1a). The average level for *sustained monologue: giving information* arrived at level B1 (1b), as did the average for *written reports and essays* (1c). In contrast to the tasks that featured DESCRIBE in *Exploring History 1*, the tasks here required a significantly higher level of detail and understanding, e.g. in "[DESCRIBE] the features of the Basic Law which were adopted to avoid the problems of the Weimar Republic." (Kröger 2009: 91, 1). For this task, students had to read the text provided by the textbook, accompanied by a schematic illustration of the structure of the Federal Republic of Germany. The text contained direct and indirect information; not everything mentioned was relevant to the task. Accordingly, and for the increased lexical sophistication students would have to process here, this task was rated as more demanding than tasks in *Exploring History 1* featuring the discourse function DESCRIBE.

	Skill			
DESCRIBE	Reading (1a)	Speaking (1b)	Writing (1c)	max. relative frequency (2)
	B2	B1	B1	B1

Table 8.18 - CEFR levels for DESCRIBE per skill and max. relative frequency across all skills, Exploring History 2

The following discourse function, the last of AFB 1, SHOW, was featured in seven tasks. Most of those tasks concerned source texts asking students to obtain information from, i.a., speeches by politicians, excerpts from laws, international contracts, and constitutions, or contemporary authors' recounting in novels. Since source texts, especially of the nature given here, are per se more complex, the average CEFR level calculated for *reading for information and argument* (see Table 8.19, 1a) was at B2. The texts were rather complex in both lexis and syntax, and the information was often only indirectly given. Talking about, e.g., "[…] why the American public became more and more weary of the [Vietnam] war (text, M3, M4, M9)." (Kröger 2009: 139, 6), was slightly less challenging, leading to *sustained monologue: giving information* at level B1+ (1b). Especially tasks that required students to read more than one text to answer asked for a higher level of detail, a deeper understanding, and increased accuracy, since the information had to be pieced together correctly. This increased demand for oral production also affected *written reports and essays*. This skill's average arrived at level B2 (1c) caused by cases in which information had to be synthesized from various sources (cf. Council of Europe 2018: 77).

Overall, the maximum relative frequency for the discourse function SHOW arrived at level B1+ (2).

SHOW	Skill			
	Reading (1a)	Speaking (1b)	Writing (1c)	max. relative frequency (2)
	B2	B1+	B2	B1+

Table 8.19 - CEFR levels for SHOW per skill and max. relative frequency across all skills, Exploring History 2

The CEFR levels for discourse functions placed with AFB 1 already displayed a higher linguistic demand, so that with AFB 2 discourse functions, a further increase would likely be overtaxing for students in grades 9 and 10. However, for FIND OUT, the first function in AFB 2, this assumption did not hold true. For *reading for information and argument,* the average CEFR level arrived at level B1 (see Table 8.20, 1a). This discourse function, however, was only used twice in the selected collection of tasks. While the second concerned oral history examples, i.e., eyewitness reports recorded and transcribed (cf. Kröger 2009: 139, M7), the first task required students to research online for the required information. To be able to assess the potential demand on reading skills, a text on details of the *Marshall Plan* (cf. The Editors of the Encyclopedia Britannica 2021: art. The Marshall Plan)*, retrieved from the online version of the *Encyclopedia Britannica*, served as an example. Accordingly, *sustained monologue: giving information* was on average at CEFR level B1 (1b). Lastly, *written reports and essays* averaged also at level B1 (1c) although "FIND OUT from the internet how the Marshall Plan actually worked." (Kröger 2009: 85, 6) required an increased CEFR level at B2. The (potential) text would need an increased level of detail and thus, the information had to be synthesized from various sources as well as apparent connections would be of essence. The overall linguistic demand across all skills for find out was calculated to be at B1+ (2).

FIND OUT	Skill			
	Reading (1a)	Speaking (1b)	Writing (1c)	max. relative frequency (2)
	B1	B1	B1	B1+

Table 8.20 - CEFR levels for FIND OUT per skill and max. relative frequency across all skills, Exploring History 2

COMPARE was more demanding in *Exploring History 2* than with *Exploring History 1*. The first task asked to compare two national flags, the Western and Eastern German ones, hence, reading was not required (cf. Kröger 2009: 91, 2). Even though the first out of four tasks did not require reading, *reading for information and argument* was at level B2+ on average (see Table 8.21, 1a). Outlining the findings was at level A2 for comparing the flags as simple terminology in a sequence of sentences sufficed in this case. The demand for oral production was significantly higher for the other three tasks, leading to an average of B1+ (1b) for all four tasks. The same assessment was made for the written part of the tasks. While the *written report and essays* for the task concerning the two former German flags was rather simple at A2, the other three were rated to be at level B2, leading to an average CEFR level of B1+ for written production (1c).

The maximum relative frequency across all skills for the discourse function arrived at level B2 (2).

COMPARE	Skill			
	Reading (1a)	Speaking (1b)	Writing (1c)	max. relative frequency (2)
	B2+	B1+	B1+	B2

Table 8.21 - CEFR levels for COMPARE per skill and max. relative frequency across all skills, Exploring History 2

The majority of tasks located with AFB 2 featured ANALYZE as their discourse function. In total, 13 tasks were concerned with the activity of "study[ing] or examin[ing] something in detail to understand or explain it" (Kröger & Lohmann 2007: half-title). Per definition, the depth with which texts have to be understood is more profound than with the discourse functions discussed before. Out of the 13 tasks, five did not feature a reading part and were concerned with analyzing photographs, maps, or caricatures. The remaining eight tasks averaged at level B2 for *reading for information and argument* (see Table 8.22, 1a). The tasks featuring ANALYZE addressed all three subskills of spoken production. *Sustained monologue: giving information*, employed with five tasks, arrived at level B1; *sustained monologue: describing experience*, e.g., when photographs were to be analyzed, arrived at B2, while the average of the four tasks that also asked for *sustained monologue: putting a case arrived* at B1. The average CEFR level for spoken production was B1+ (1b). Putting the findings and possible explanations in writing required the average level of B1+ with *written reports and essays* (1c). Moreover, the maximum relative frequency of all assigned CEFR levels for ANALYZE was also B1+ (2).

ANALYZE	Skill			
	Reading (1a)	Speaking (1b)	Writing (1c)	max. relative frequency (2)
	B2	B1+	B1+	B1+

Table 8.22 - CEFR levels for ANALYZE per skill and max. relative frequency across all skills, Exploring History 2

Seven tasks throughout the two chosen chapters asked students to ASSESS. This discourse function can be challenging when judging situations, people, or problems since each judgment needs a certain amount of justification. Hence, spoken production could be more demanding than reading or writing. With *Exploring History 2*, the average demand for *reading for information and argumen*t (see Table 8.23, 1a) arrived at an average of B2 with only one task not relying on text but a caricature (cf. Kröger 2009: 85, 7), while two tasks required such a depth of understanding that it was evaluated at level C1. The average level to put the findings obtained through reading into their own words for *sustained monologue: giving information, sustained monologue: describing experience* and *sustained monologue: putting a case* were assessed to average at B2 (1b). Likewise, the average CEFR level to record the findings and their assessment was B2 (1c). The maximum relative frequency for the discourse function ASSESS arrived at level B2 (2) with 57% of all assessments made for this discourse function being at that level.

ASSESS	Skill			
	Reading (1a)	Speaking (1b)	Writing (1c)	max. relative frequency (2)
	B2	B2	B2	B2

Table 8.23 - CEFR levels for ASSESS per skill and max. relative frequency across all skills, Exploring History 2

The first discourse function of AFB 3 analyzed was EXPLAIN, which, per definition, requires a deeper understanding and the ability to put those aspects into students' own words. Out of all discourse functions of this AFB, EXPLAIN was used most often. Reading proved to be quite demanding for the nine tasks concerned with EXPLAIN. Explaining requires a deep understanding of detail, often beyond the main points. Moreover, most tasks were based on source texts, e.g., Eisenhower's *Domino Theory* in the context of the Vietnam War as published in the *Pentagon Papers*, 1953 (cf. Kröger 2009: 138, M5). Text of that quality is quite demanding per se. Consequently, the average level for *reading for information and argument* arrived at a rather high CEFR level of B2+ (see Table 8.24, 1a). While transferring the texts and their details into students' words is a cognitive process, the spoken production for *sustained monologue: giving information* arrived at B1+ for cases in which students had to put the aspects of the tasks in their own words. For justification or reasoning of those assertions, *sustained monologue: putting a case* was at level B2. Although own words had to be used, the level of required detail, reliability, and accuracy called for a higher level in the assessment. On average, the level of spoken production was B1+ (1b). *Written reports and argument* called for level B2 again (1c) taking into account the necessity of increased detail, the necessity to synthesize information from various sources, and the systematicity needed to explain aspects. The maximum relative frequency for all assigned CEFR levels for EXPLAIN in total arrived at B2 (2).

EXPLAIN	Skill			
	Reading (1a)	Speaking (1b)	Writing (1c)	max. relative frequency (2)
	B2+	B1+	B2	B2

Table 8.24 - CEFR levels for EXPLAIN per skill and max. relative frequency across all skills, Exploring History 2

Since the remaining three discourse functions in AFB 3, DISCUSS, ARGUE, and GIVE AN OPINION were used only once throughout the selected tasks, the results will be discussed in one paragraph. Starting with DISCUSS, the level required to read the relevant text was B2 (see Figure 8.6 - Average CEFR levels per discourse function and skill & per discourse function across skills, Exploring History 2, 1a), while the actual discussion was at level B1 (1b). The task concerned with DISCUSS did not require a written part (1c). The maximum relative frequency across the two required skills for this discourse function was B1+ (2). The task employing ARGUE did not require dedicated reading but was based on previous information. Hence, only *sustained monologue: putting a case* was rated along with *written report or essay*. The spoken part was evaluated to put the demand at CEFR level B2 (1b), and the same applied to the written part (1c). The maximum relative frequency across those two skills for the discourse function was accordingly also B2 (2). Lastly, GIVE AN OPINION addressed all three subskills. The reading part was assessed to ask for reading skills at level B2 (1a) while giving their own opinion in this context was already possible with CEFR level B1 (1b), which was also sufficient with writing (1c). GIVE AN OPINION mostly asked for level B1 across all three skills. Other than for the discourse functions before, there will be no overview of the assigned CEFR levels per discourse functions as the grid featuring all above-described findings follows (see Figure 8.6). It displays the three

skills and their CEFR levels, along with the red zone of targeted CEFR levels set by the curricular standards to compare the assessments with the targeted level of skill.
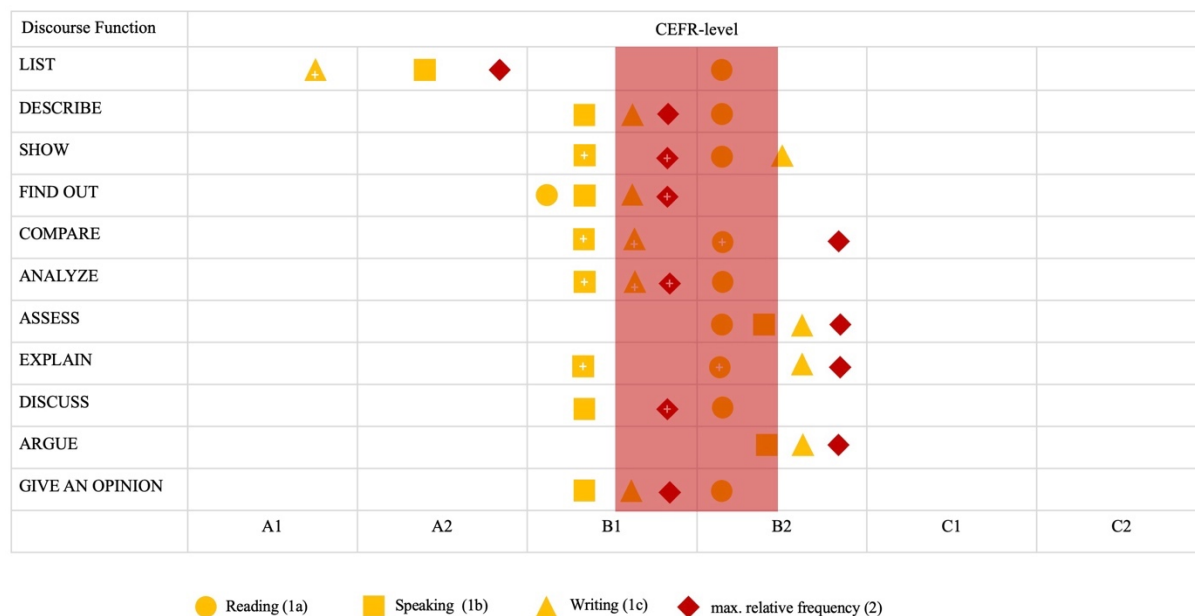
Figure 8.6 - Average CEFR levels per discourse function and skill & per discourse function across skills, Exploring History 2

The average value across all discourse functions for the respective skill, reading, arrived at B2 with all 11 discourse functions (see chapter 8.2.3 and Figure 8.3, four steps), while speaking was slightly less demanding at level B1. Lastly, writing was writing at B1+. The maximum relative frequency for all assessments across all discourse functions and skills (4) was also B1+.
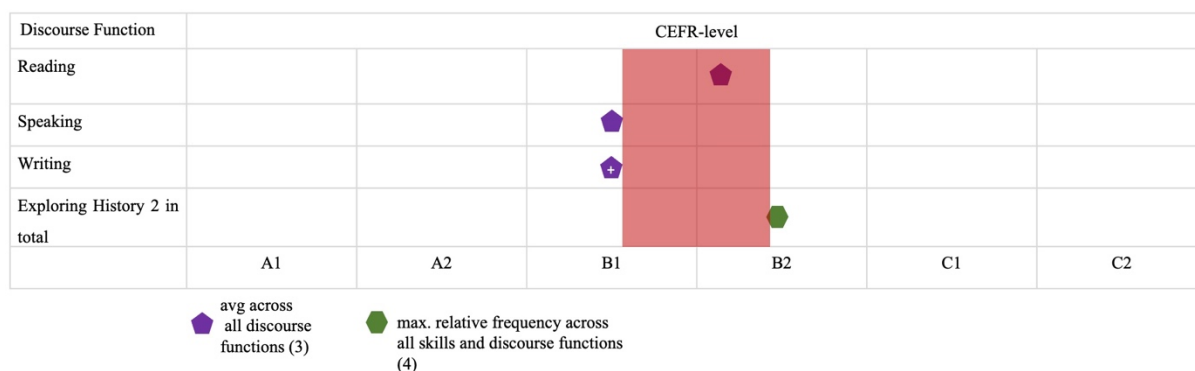
Figure 8.7 - Average CEFR levels per skill across discourse function and max. relative frequency of Exploring History 2 in total

8.3.2   Geography

For Geography, a list of discourse functions for bilingual Geography classes published by the Ministry for Schools and Education of North Rhine-Westphalia served as a reference. Rhineland-Palatinate (RLP) uses only a shortened version for this context (cf. Geisler 2012), which covers only six functions, three each for the lower two AFBs and thus, required substitution. A closer look at this list for bilingual classes revealed that the definitions for those functions to be used in bilingual classes were attenuated, e.g., while BESCHREIBEN requires students to apply appropriate use of subject-related terminology in a connected and sequential description, the bilingual pendant DESCRIBE required only coherence but excludes the explicit mentioning of terminology (ibid.). Furthermore, operating functions for the most challenging AFB 3 are not defined for bilingual classes for RLP. Since the selected textbook, *Diercke Geography* for bilingual classes uses discourse functions addressing AFB 3 in the target language, the RLP list of six operating functions for bilingual classes was complemented with entries from North Rhine-Westphalia (Ministerium für Schule und Bildung des Landes Nordrhein-Westfalen 2015a). Compared to History, Geography employs the same amount of distinct discourse functions. SHOW is not featured in Geography; however, AFB 1 features a discourse function not used in History: COLLECT. The definition of COLLECT is similar to the one given with SHOW in History. Likewise, GIVE AN OPINION is replaced by COMMENT ON in Geography with an identical definition (see Table 8.25). Furthermore, PRESENT and LOCATE enhance the list of discourse functions in AFB 1, ILLUSTRATE is added to AFB 2. For *Diercke Geography Volume 1* targeted for grades 7 and 8, 110 applications of discourse functions were analyzed. The majority was placed with AFB 1, in total 71, while 33 discourse functions were taken from AFB 2 and just six were placed with AFB 3. Other than with History, EXPLAIN for Geography is already classified as AFB 2.

Discourse Function	Definition (cf. Geisler 2021; Ministerium für Schule und Bildung des Landes Nordrhein-Westfalen 2015a)	AFB	Frequency
LIST	Write down information in a given order without commenting on it	1	14
DESCRIBE	Give details about what something or someone is like	1	32
COLLECT	Gather and write down information without commenting on it	1	5
NAME	mention information and issues without commenting on them	1	10
PRESENT	Illustrate information and (factual) connections in an ordered manner, using either diagrams/pictures or verbal elaborations	1	4
LOCATE	Spatialize case studies/ regional or zonal examples in known topographic patterns	1	6

COMPARE	Put facts next to each other and decide what they have in common or what differentiates them – which factors are the same, similar, different or opposite	2	5
EXPLAIN	context of justification, outline prerequisites and consequences of distinct structures and processes	2	25
ILLUSTRATE	Clarify correlations with the help of additional subject-related information (graphical/verbal), technical terms	2	3
COMMENT ON	State a reasoned opinion about a fact/ a thesis	3	3
DICSUSS	Talk about something in detail, coming to a reasoned opinion	3	3

Table 8.25 - List of discourse functions, definitions used in Diercke Geography Volume 1

Similar to History, the initial step to this part of the analysis was to map the discourse functions listed in Table 8.25 above to all possible levels of the CEFR and their respective skills (see chapter 12.6.2). Starting with LIST and the first skill *reading for orientation*, the possible levels required varied between A2 and B1 since students would have to scan texts of varying lengths for information that is mainly given directly. The differentiating factor would be the way in which information or pieces thereof would have to be understood, spanning from understanding specific information to understanding factual, relevant points. The scale for *reading for orientation* features a descriptor or can-do-statement fitting the required level of skill "Can scan longer texts in order to locate desired information, and gather information from different parts of a text, or from different texts in order to fulfill a specific task." (Council of Europe 2018: 62) which describes level B1. Per definition, LIST does not feature a spoken part; thus, the next skill to be assessed preliminarily would be writing. For *written reports and essays,* the maximum level of skill required would be A2 in cases that ask for sentences with simple connectors; mostly level A1 would suffice if single terms or sentences have to be written down. The upfront assessment of LIST proved to be valid throughout the analysis of the 14 tasks featuring this discourse function in *Diercke Geography Volume 1*. Out of 14 applications of the discourse function, only two required reading skills to be at level B1 since students had to "[…] scan longer texts in order to locate desired information, and gather information from different parts of a text, or from different texts in order to fulfill a specific task." (ibid.) with an increasingly complex text that did not offer the information necessary in one paragraph but spread all over the text (e.g., cf. Löbmann 2014: 37, 1). The average level for *reading for orientation* (see Table 8.26, 1a) arrived at level A2, the demand with *written reports and essays* (1c) averaged at A1+. The relative frequency was at a maximum for all CEFR levels was at A1/A1+ (2) with 73% of all assigned CEFR levels.

	Skill			
LIST	Reading (1a)	Speaking (1b)	Writing (1c)	max. relative frequency (2)
	A2	-	A1+	A1+

Table 8.26 - CEFR levels for LIST per skill and max. relative frequency across all skills, Diercke Geography 1

Most tasks in *Diercke Geography Volume 1* were concerned with the discourse function DESCRIBE*,* which is defined in the same way as in History. In general, the necessary reading skills for

reading for information and argument could be between A2 and B1, depending on the length and complexity of the text and the level of detail and understanding necessary. Moreover, in cases concerned with the description of, i.a., schematic illustrations, photographs, graphs, level A1 may also be sufficient. Speaking skills for *sustained monologue: giving information* could range from A1, "Give a simple description of an object or picture while showing it to others using basic words[…]" (Council of Europe 2018: 71) to B1, requiring students to "[…] report straightforward factual information on a familiar topic[…]" (ibid.). For the level between the two aforementioned (A1 and B1), level A2, the can-do-statement provided with *overall spoken production* matches the activity needed for DESCRIBE better than the one given with *sustained monologue: giving information* and was thus chosen for the intermediate level as students "can give a simple description or presentation […] as a short series of simple phrases and sentences linked into a list" (ibid.: 69). The linguistic requirements tied to the written part of DESCRIBE follow suit to the spoken part and thus, could also range between A2 for writing "[…] simple texts on familiar subjects of interest, linking sentences with connectors […]" (ibid.: 77) and level B1, which asks for increased precision when describing aspects as well as an increased level of detail in a linear sequence of points. The average CEFR levels for all three skills and their subskills were again calculated. Accordingly, *reading for information and argument* (see Table 8.27, 1a) arrived at level A1+ which may seem surprising in the context of the results obtained with the text analysis (see chapter 7.2). However challenging those texts were evaluated to be, the information required to work the tasks given was mostly also presented alongside graphics or in single paragraphs in the texts. For the task "DESCRIBE the seasons in the Northern and Southern hemisphere." (Löbmann 2014: 33, 1a & b), e.g., the necessary information could be obtained through the graphics only and was supplemented and enhanced with details through the text rated B1+ (see Table 7.5). The average CEFR level across all 32 cases in which DESCRIBE was applied for *overall spoken production* was at level A2, the same for *sustained monologue: giving information* (1b). The third step in each task, recording their descriptions, required students to display an average CEFR level of A2 for *written reports and essays* (1c). The overall requirement for DESCRIBE across all three types of skills and their subskills was A2 as well (2).

DESCRIBE	Skill			
	Reading (1a)	Speaking (1b)	Writing (1c)	max. relative frequency (2)
	A1+	A2	A2	A2

Table 8.27 - CEFR levels for DESCRIBE per skill and max. relative frequency across all skills, Diercke Geography 1

The next discourse function, COLLECT, requires students to "Gather and write down information without commenting on it" (Ministerium für Schule und Bildung des Landes Nordrhein-Westfalen 2015a) and differs only slightly from LIST used in History. While with LIST information has to be reproduced following a specified order, COLLECT passes up on such a preset order and simply asks for an accumulation of information. As both discourse functions are quite similar, the possible levels of skills required were as well (see chapter 8.3.1). Therefore, the analysis continues with the average values

calculated across the five cases that applied COLLECT. For *reading for orientation* (see Table 8.28 , 1a)*,* the average CEFR level was B1. The written part was only partly evaluated for COLLECT since three tasks featured a second discourse function with which the two subskills were assessed. For those tasks of COLLECT that actually had a spoken and written part, the average CEFR level for *overall written production* (1c) was A1, too. The maximum relative frequency across all skills for COLLECT was located at B1 (2).

	Skill			
COLLECT	Reading (1a)	Speaking (1b)	Writing (1c)	max. relative frequency (2)
	B1	A1	A1	B1

Table 8.28 - CEFR levels for COLLECT per skill and max. relative frequency across all skills, Diercke Geography 1

NAME is similar to LIST and COLLECT*,* focusing on the spoken part of the tasks concerned with said discourse function*.* Consequently, the preliminary considerations on possible CEFR levels needed for LIST also apply to NAME. Therefore, the discussion continues with the average values calculated for the ten tasks working with NAME*.* The average value for *reading for orientation* (see Table 8.29, 1a) was A2 as the information required was either concentrated in few distinct lines of the text or had to be retrieved from the graphics provided with each text. For the spoken part, the average for both *overall spoken production* and *sustained monologue: giving information* arrived at A1 (1b). The majority of all assigned CEFR levels with this discourse function was A1 (2) with 68% of a total of 28 assessments being placed at that level.

	Skill			
NAME	Reading (1a)	Speaking (1b)	Writing (1c)	max. relative frequency (2)
	A2	A1	-	A1

Table 8.29 - CEFR levels for NAME per skill and max. relative frequency across all skills, Diercke Geography 1

Other than History, AFB 1 in Geography features six discourse functions, one of which is PRESENT. Here, the focus shifts to a subskill in the spoken skills that was not yet discussed: *addressing audiences*. It is concerned with giving presentations or a speech in class, including the questions and answers that usually follow a presentation. Such presentations can range "from short, rehearsed statements […] to a well-structured presentation on a complex subject […]" (Council of Europe 2018: 74). Still, reading comes first. The CEFR levels possibly needed for *reading for information and argument* could span from A1, especially with rich visual support, to B1 in cases that include longer, complex texts which have to be understood on a deeper level and grade of detail to compile a presentation. The dedicated subskill for oral production, *addressing audiences,* might be needed between levels A2 for "[…] giv[ing] a short, rehearsed basic presentation on a familiar subject […] briefly giv[ing] reasons and explanations for opinions, plans and actions [while] cop[ing] with a limited number of questions." (ibid.) to level B2, at which the presentation has to be systematically developed "[…] with highlighting of significant points, and relevant supporting detail. [Being] in support of or against a particular point of view." (ibid.). Simultaneously, students "can take a series of follow up

questions with a degree of fluency [posing] no strain for him/herself or the audience." (ibid.). The written part potentially is a little less demanding, from simple, isolated words and phrases at level A1 to level B1, asking students to "[…] present a topic in a short report or poster, using photographs and short blocks of text." (ibid.: 77). In *Diercke Geography Volume 1*, multiple tasks used more than a single discourse function. This was also the case with in three out of four tasks asking to PRESENT. Hence, the required reading skills were evaluated for the first discourse function within the task and not for PRESENT as such. Only one task actually asked for *reading for information and argument* in direct connection to the discourse function. This task was assessed to demand level B1 from students (see Table 8.30, 1a). The spoken part, *addressing audiences*, was evaluated for all four tasks and arrived at an average CEFR level of A2+ (1b). Lastly, the written part was calculated to average at level A2+ (1c). All three skills and their individual assessments for all tasks arrived at level B1 as the level of linguistic requirements for the discourse function PRESENT in the context of *Diercke Geography Volume 1*.

	Skill			
PRESENT	Reading (1a)	Speaking (1b)	Writing (1c)	max. relative frequency (2)
	B1	A2+	A2+	B1

Table 8.30 - CEFR levels for PRESENT per skill and max. relative frequency across all skills, Diercke Geography 1

The last discourse function of AFB 1 in Geography is LOCATE and is, in its meaning, very specific for the subject as such. It is concerned with finding and relating aspects particular for certain regions or geographic phenomena. This usually involves using an atlas or maps of any kind, e.g., with task 1 on page 41, "Locate the Amazonas Basin. Use your atlas." (Löbmann 2014: 41, 1). Consequently, the required skill for *reading for orientation* spans from level A1 when only single terms are required to A2 when being asked to "locate specific information […] and isolate information required" (Council of Europe 2018: 62) with reading as such being very selective. When describing the location connected to a task, *sustained monologue: giving information* could be at level A1 in cases with rich visual support, but also at A2 in cases for which "[…] simple directions […] using basic expressions […]" (ibid.: 71) are essential. The same level could also be required for *written reports and arguments*, for which single, isolated phrases could be satisfying as well as a sequence of connected sentences in simple language. For the six tasks employing LOCATE in *Diercke Geography Volume 1*, the average value for *reading for orientation* (see Table 8.31, 1a) arrived at level A1+. *Giving information* for the spoken part averaged out at level A2 (1b) and the written component in *written reports and essays* arrived at level A2 as well (1c). Across all skills and their subskills, tasks to LOCATE required level A2 (2) in 53% of all assessed uses, however, level B1 was also demanded with one task and its reading.

	Skill			
LOCATE	Reading (1a)	Speaking (1b)	Writing (1c)	max. relative frequency (2)
	A1+	A2	A2	A2

Table 8.31 - CEFR levels for LOCATE per skill and max. relative frequency across all skills, Diercke Geography 1

Remarkably, out of 110 tasks, 71 were allocated with the lowest AFB, focusing on repetition and reproduction of information in a scope limited to thematic contexts and acquired skills (cf. Geisler 2012). At the same time, the highest assigned CEFR levels were at level B1. Possibly, the demand could increase with the three discourse functions in AFB 2 and the 33 tasks working with those.

The first of those discourse functions is COMPARE, which shares the definition with History as the activity to "put facts next to each other and decide what they have in common or what differentiates them – which factors are the same, similar, different or opposite" (Ministerium für Schule und Bildung des Landes Nordrhein-Westfalen 2015a). *Reading for information and argument* may ask for skills ranging from level A2, "[…] pick[…] out main information […]" (Council of Europe 2018: 63) to level B2, when "[…] highly specialized sources […]" (ibid.) are involved. When putting findings into words, the demand could be between level B1 when "[…] explain[ing] the main points in an idea or problem with reasonable precision." (ibid.: 63) would be sufficient but could also go up to level B2 if more details and a refined level of understanding are required. In cases that require written production, the CEFR level demanded could be between level A2 and level B1, depending on the needed sophistication of the composition. In *Diercke Geography Volume 1,* however, the tasks featuring COMPARE did not require writing as a second discourse function, usually LIST, was used in those tasks. The demand for those was already described above. Regarding reading, the average level for *reading for information and argument* (see Table 8.32, 1a) arrived at level B1, as did the average level for *sustained monologue: giving information* (1b). Consequently, the overall demand across the two skills for COMPARE was also found at level B1 (2) assigned with 60% of all assessments.

	Skill			
COMPARE	Reading (1a)	Speaking (1b)	Writing (1c)	max. relative frequency (2)
	B1	B1	-	B1

Table 8.32 - CEFR levels for COMPARE per skill and max. relative frequency across all skills, Diercke Geography 1

Interestingly, the next discourse function, EXPLAIN, is defined differently than it is for History and marks an activity in the "context of justification, outline prerequisites and consequences of distinct structures and processes" (Ministerium für Schule und Bildung des Landes Nordrhein-Westfalen 2015a) which also differs from the German definition "Informationen und Sachverhalte so darstellen, dass Bedingunen, Ursachen, Folgen und Gesetzmäßigkeiten verständlich werden" (Geisler 2012). The latter appears to be closer to the definition used in History, "Tell something in a way to help others understand it better" (Kröger & Lohmann 2007: half title). Still, the underlying target is the *reorganization and transfer* of knowledge in Geography (cf. Ministerium für Bildung, Wissenschaft, Weiterbildung und Kultur 2021: 27). Regardless of the definition, the required level of understanding to be able to EXPLAIN could resolve in CEFR levels between A2, if specific information in simple language has to be understood, and up to level C1, depending on the complexity, length, and required level of understanding when *reading for information and argument.* The same range of levels could be required

when putting the information into words for a *sustained monologue: giving information.* Lastly, *written reports and essays* could range between A2 and C1 when putting the explanations in writing. The textbook analyzed featured a total of 25 tasks employing the discourse function. For *reading for information and argument*, the required level of skill (see Table 8.33, 1a) was at level A2 on average. Putting the information into words to help understanding the aspect asked for an average level B1 for s*ustained monologue: giving information* (1b)*.* Lastly, *written reports and essays (1c)* arrived at level A2+ on average. Across all three skills and their 70 individual assessments, level A2 was assigned most often with 47 cases corresponding to 67%, and thus, being also the CEFR level for the maximum relative frequency (2).

	Skill			
EXPLAIN	Reading (1a)	Speaking (1b)	Writing (1c)	max. relative frequency (2)
	A2	B1	A2+	A2+

Table 8.33 - CEFR levels for EXPLAIN per skill and max. relative frequency across all skills, Diercke Geography 1

The third and last discourse function in AFB 2 is ILLUSTRATE, a discourse function not used in History, that tasks students with finding additional material or information to support the understanding of thematic aspects and contexts. *Reading for information and argument* in this regard could demand levels A2 to B2 in cases using highly specialized sources to fulfill this task. The same span could apply to *sustained monologue: giving information* or *overall spoken production* (see chapter 12.6.2). When putting the supporting information in writing, the possible levels of skill for *written reports and essays* could range from A2 to B1. In *Diercke Geography Volume 1*, only three tasks featured ILLUSTRATE*.* One task asked for creating a poster to ILLUSTRATE the different vegetation zones in Africa and the other for creating a mind map on weather, vegetation, and animals in the tropical rain forest (cf. Löbmann 2014: 35, 1; 39, 1). The third requested the drawing of a flow chart to support the description of how the system of shifting cultivation in rainforests works (ibid.: 37, 1). For the former tasks, the main text serves as a source of information, and the latter task names two additional photographs as sources of information. On average, the required skill for *reading for information and argument* (see Table 8.34, 1a) arrived at level B1 since several sources contained the information, and understanding factual, relevant points was necessary. The average CEFR level to describe the illustrations along with the information obtained through reading arrived at A2+ for *sustained monologue: giving information* (1b). Finally, *written reports and essays* were limited to single isolated phrases or simple sentences; thus, the average level required was A1+ (1c). The level most often assigned in relation to all assessments was A2 (2).

	Skill			
ILLUSTRATE	Reading (1a)	Speaking (1b)	Writing (1c)	max. relative frequency (2)
	B1	A2+	A1+	A2

Table 8.34 - CEFR levels for ILLUSTRATE per skill and max. relative frequency across all skills, Diercke Geography 1

In Geography, AFB 3 comprises only two discourse functions, DISCCUSS and COMMENT ON. Both were applied three times each in the textbook. While DISCUSS is used in a similar sense as in

History, COMMENT ON is a new discourse function that must be described further. It is defined as the activity that asks students to "state a reasoned opinion about a fact/ a thesis" (Ministerium für Schule und Bildung des Landes Nordrhein-Westfalen 2015a). The definition implies already that students would need to have a thorough understanding of facts and their details, regardless of the demand imposed by the text regarding length, complexity, or manner of information delivery. Accordingly, the level minimally necessary for *reading for information and argument* would be B1, ensuring that students "can understand straightforward, factual texts on subjects relating to their interests or studies." (Council of Europe 2018: 63). This could go up to level C1 when texts are increasingly long and complex providing not only information but also the writer's attitudes and opinions (ibid.). When commenting on something, the corresponding oral skill is *sustained monologue*: *putting a case*, which requires students to display competences between levels B1 and C1, depending on the specificity of arguments and the required level of detail. The *written reports and essays* or *creative writing* ask at least for level B1 since writing would have to be a "[…] detailed, straightforward description on a range of familiar subjects within their field of interest." (ibid.: 76). The actual tasks averaged at level B1 for reading (see Table 8.35, 1a); for spoken production, however, only two were assessed as the third asked students to answer the task in German. Said two tasks averaged at level A2+ for spoken production (1b) and level B1 for *creative writing* because the tasks ask for either a letter to a fictitious friend (cf. Löbmann 2014: 61, 3) or a diary entry (ibid.: 37,3). The maximum relative frequency was at level A2 across all skills and assessments for COMMENT ON (2).

	Skill			
COMMENT ON	Reading (1a)	Speaking (1b)	Writing (1c)	max. relative frequency (2)
	B1	A2+	B1	B1

Table 8.35 - CEFR levels for COMMENT ON per skill and max. relative frequency across all skills, Diercke Geography 1

The last discourse function in AFB 3 is DISCUSS, which is defined similarly to the definition used in History. Additionally, the possible CEFR levels for the three skills are similar as well. Within the preliminary considerations on these CEFR levels for this discourse function, reading was rated between B1 and C1, assuming that to discuss the topic in detail, as per the definition, a detailed understanding of increasingly complex text could be required. Likewise, the considerations on speaking skills resulted in high levels starting at B2 for *sustained monologue: putting a case*. Since discussions entail argumentation, the ability to communicate effectively and precisely, marking level C1, could also be needed. This assumes that discussions require speaking skills with increased accuracy, fluency, and a high level of detail. Although discussions in classrooms rarely lead to written products, the possibility of *written reports and essays* was assessed to be likely quite demanding, rated at level C1. The overall impression of this discourse function was that of a linguistically demanding activity. After the analysis of *Diercke Geography Volume 1* and the three tasks featuring the discourse function DISCUSS, it turned out that the required reading skills were at level B1 (see Table 8.36, 1a). Although at least two tasks involved dealing with information provided by given texts or materials beforehand, a discussion should

be based on the complete picture rather than the specific details addressed in the other tasks. This calls for a thorough re-reading of the text. Hence, reading on level B1 was assigned for the discourse function. For the productive part, i.e., *sustained monologue: putting a case*, level B1 (1b) was selected, deviating from the preliminary evaluation. The intense work with the text beforehand would provide students with adequately detailed knowledge to lead a sufficiently precise discussion. Furthermore, the arguments required for the discussions were basically given already in the text, which imparted a somewhat repetitive character to the three tasks. Lastly, the required skill level to put the discussion into writing was assessed to be at level B1 (1c). Again, a deviation from the preliminary evaluation, which the repetitive character of the task can explain as such. With all three tasks featuring DISCUSS assessed at level B1, the maximum relative frequency for this discourse function also arrived at level B1 (2).

	Skill			
DISCUSS	Reading (1a)	Speaking (1b)	Writing (1c)	max. relative frequency (2)
	B1	B1	B1	B1

Table 8.36 - CEFR levels for DISCUSS per skill and max. relative frequency across all skills, Diercke Geography 1

All above-described findings are placed on the grid of skills and CEFR levels again, along with the red-colored zone of targeted CEFR levels set by the curricular standards.
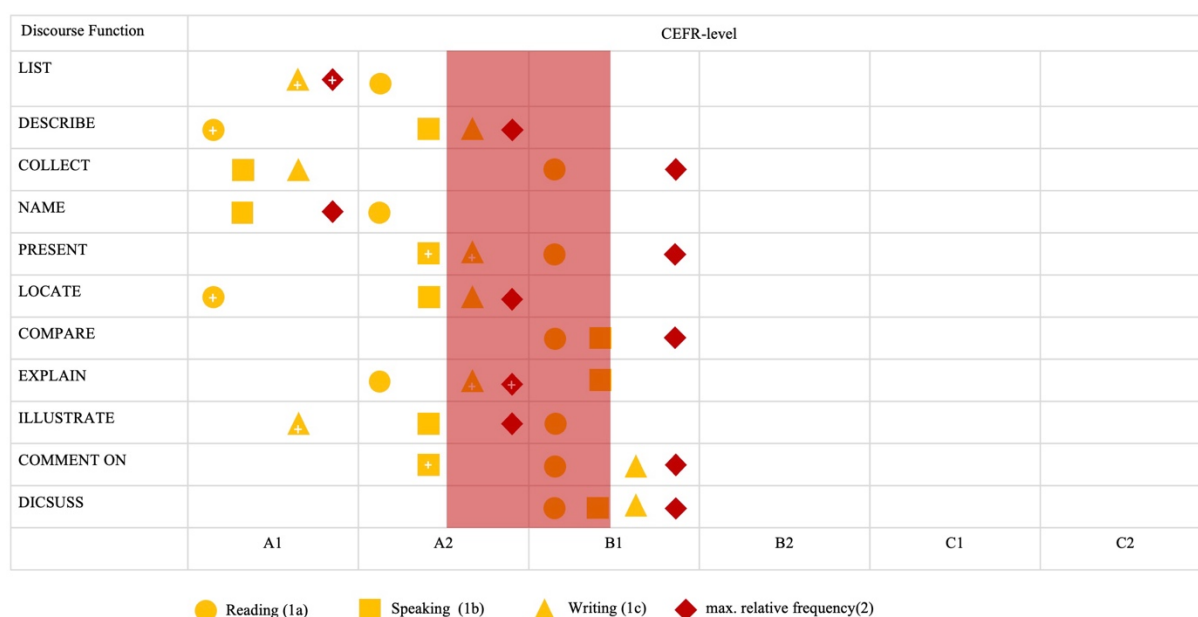
Figure 8.8 - Average CEFR levels per discourse function and skill & max. relative frequency per discourse function across skills, Diercke Geography Volume 1

Finally, the analysis calculated the average CEFR level per skill and detected the overall maximum frequency across all discourse functions and skills. The average for the required reading skills, including *reading for information and argument* and *reading for orientation*, was calculated to be at level A2+. The same level was calculated for the average of all four subskills of spoken production, while written production demanded an average skill across all discourse functions of A2 (3). The maximum relative frequency of all 351 assessments made for discourse functions and skills resulted in

CEFR level A2 (4) as the required skill level to carry out the tasks in *Diercke Geography Volume 1*. This comparably low requirement level starkly contrasts the high complexity attested to the texts analyzed (see 7.5.2). Possible reasons and justification for this contrast are discussed in chapter 8.4. The following grid gives an overview on the distribution of CEFR levels for the averages discussed above (3) and (4).
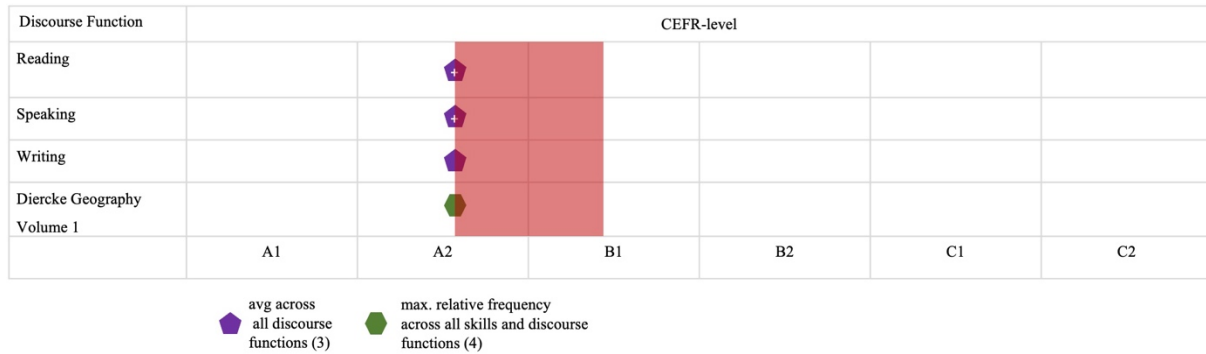
Figure 8.9 - Average CEFR levels per skill across discourse function and maximum relative frequency average of Diercke Geography Volume 1

Diercke Geography, Volume 2, grades 9 and 10

Diercke Geography also features a second volume targeted at and designed for the second age group defined for this study. It is mainly used in grades 9 and 10 and thus supposedly reflects upon the increased skills students have acquired during the years of instruction. In addition to the discourse functions listed and defined for volume 1, this volume adds four new discourse functions: EVALUATE, CLASSIFY, and ANALYZE for AFB 2 as well as JUSTIFY for AFB 3 while AFB 1 no longer features COLLECT and AFB 2 no longer entails ILLUSTRATE. The overview of definitions and the frequencies with which the distinct discourse functions were used accordingly changes into:

Discourse Function	Definition (cf. Geisler 2012; Ministerium für Schule und Bildung des Landes Nordrhein-Westfalen 2015a)	AFB	Frequency
LIST	write down information in a given order without commenting on it	1	8
DESCRIBE	give details about what something or someone is like	1	14
NAME	mention information and issues without commenting on them	1	1
PRESENT	illustrate information and (factual) connections in an ordered manner, using either diagrams/pictures or verbal elaborations	1	2
LOCATE	spatialize case studies/ regional or zonal examples in known topographic patterns	1	2

COMPARE	put facts next to each other and decide what they have in common or what differentiates them – which factors are the same, similar, different or opposite	2	6
EXPLAIN	context of justification, outline prerequisites and consequences of distinct structures and processes	2	12
CLASSIFY	assign spatial and/or factual aspects to a distinct position in an order/pattern based on determined characteristics	2	1
ANALYZE	study or examine something in detail in order to understand and explain it	2	4
EVALUATE	develop a qualified opinion/ a well-founded, science-based point of view based on information and knowledge on the subject matter and disclose/ reflect upon criteria applied	3	1
DICSUSS	talk about something in detail, coming to a reasoned opinion	3	6
COMMENT ON	state a reasoned opinion about a fact/ a thesis	3	4
JUSTIFY	develop a reasoned argument and highlight thematic correlations/connections on complex/basic considerations or theses	3	1

Table 8.37 - List of discourse functions, definitions used in Diercke Geography Volume 2

Two discourse functions were added, and the total number of tasks analyzed and the distribution of tasks per AFB has changed. In total, 62 use cases of discourse functions were analyzed, of which 27 were allocated with AFB 1, 23 with AFB 2, and 12 with AFB 3. The distribution of tasks shifted significantly from AFB 1 towards AFB 2 and 3, despite the total number being considerably lower than in *Diercke Geography Volume 1*. In the first volume, 65% of all tasks were allocated to AFB 1, whereas in the second volume, this figure decreased to 44% or 27 tasks. The number of tasks in AFB 3 increased from 5% or six tasks in volume 1 to 18% and 12 tasks in volume 2 instead. The percentage increase with AFB 2 tasks is less significant, from 33 tasks and 30% in volume 1 to 23 tasks making 38% in volume 2. Still, tasks with discourse functions allocated with AFB 1 were the majority in volume 2 but to a lesser extent than in volume1.

Similar to the first volume, the analysis started with looking at the first discourse function listed with AFB 1, LIST. Throughout the eight uses in tasks, the assigned CEFR level varied between A2 and B1 for the subskills *reading for orientation* and *reading for information and argument.* For oral production, only three out of the eight tasks had a spoken part, which varied between CEFR level A1, assumed answers consisting of a single word or sentence, and level A2. As with oral production, the

analysis only partially assessed written production because some tasks featured more than the discourse function LIST. Accordingly, other discourse functions covered the missing steps for speaking and writing. For those cases which employed a written part, the CEFR levels were between A1 and A2. Consequently, the average level required for both subskills of reading arrived at B1 (see Table 8.38, 1a). For those tasks covering the spoken part, the average CEFR level was at A1+ (1b); for those tasks containing written production, the average level was also A1+ (1c). The required level assigned most often in relation to all assessments within LIST was not clearly to make out as A1, A2, and B1 were assigned five times each. Hence, it is not possible to find a maximum relative frequency (2). Regardless, the maximum demand at level B1 should be kept in mind.

	Skill			
LIST	Reading (1a)	Speaking (1b)	Writing (1c)	max. relative frequency (2)
	B1	A1+	A1+	n.a.

Table 8.38 - CEFR levels for LIST per skill and max. relative frequency across all skills, Diercke Geography 2

The following discourse function was the one of AFB 1 applied most often. DESCRIBE was used in 14 tasks for which the required reading skill varied from A1 to B1. Level A1, e.g., was assigned in cases asking students to describe a picture or schematic drawing annotated with single expressions or simple sentences. The average CEFR level for *reading for orientation* and *reading for information and argument* was calculated to be A2+ (see Table 8.39, 1a). This level translates into reading skills ensuring the understanding of short texts and simple phrases containing specific information in simple language (cf. Council of Europe 2018: 62f.). When describing, i.e., speaking about the observations, the speaking skills required would have to be between A1 and A2. The levels assigned to four tasks that entailed a spoken part arrived at a calculated average of level A1+ (1b). The tasks requiring a written product averaged at level B1 (1c). Other than LIST, DESCRIBE had a distinct CEFR level assigned most often. In relation to a total of 41 assessments for DESCRIBE, level B1 was assigned most often, namely 20 times. This was equivalent to 49% of all assessments. Hence, the maximum relative frequency was to be found at CEFR level B1 (2).

	Skill			
DESCRIBE	Reading (1a)	Speaking (1b)	Writing (1c)	max. relative frequency (2)
	A2+	A1+	B1	B1

Table 8.39 - CEFR levels for DESCRIBE per skill and max. relative frequency across all skills, Diercke Geography 2

NAME, LOCATE, and PRESENT were only rarely used. Precisely one task used NAME and two employed LOCATE. While *reading for orientation* for NAME was assessed at level A2 (see Table 8.40, 1a), both productive skills were rated as level A1 (1b and 1c). With one single task using NAME, the maximum relative frequency is the CEFR level used most often. In this case, level A1 was assigned with two out of three assessments (2). For PRESENT, the two use cases averaged at level A2 for *reading for information and argument (1a)*. The spoken subskill specific for this discourse function, *addressing audiences*, arrived at an average level of A2+ (1b). As was the case with PRESENT and its written part

for volume 1, the required level for writing was A2 (1c). Out of four assessments, three were located with level A2 making this the most frequent CEFR level for this discourse function in relation to all assessments (2). The last discourse function, one of the rarely used of AFB 1, was LOCATE. The two tasks concerned with locating specific places or landmarks averaged at A2 for *reading for orientation* (1a). The spoken part (1b) and the written part (1c) required the same level. Consequently, level A2 was also the maximum value for relative frequency in this case (2).

	Skill			
	Reading (1a)	Speaking (1b)	Writing (1c)	max. relative frequency (2)
NAME	A2	A1	A1	A1
PRESENT	A2	A2+	A2	A2
LOCATE	A2	A2	A2	A2

Table 8.40 - CEFR levels for NAME, PRESENT, and LOCATE per skill and max. relative frequency across all skills, Diercke Geography 2

AFB 2 was comprised of five discourse functions, of which three the first volume did not feature. COMPARE, however, was also used in volume 1. Six tasks applied this function in total in volume 2. Among those six tasks, the level required for reading skills varied between A2 and B1, whereas one task did not feature reading text from the textbook but students' written production. Hence, the analysis exempted this task for reading as said skill was not addressed. For the remaining five tasks, the average level of skill for reading was A2+ (see Table 8.41, 1a). Other than with reading, all six tasks required spoken production. Accordingly, the assigned CEFR levels arrived at an average CEFR level of B1 (1b), calling for oral production with increased fluency, level of detail, and precision. Like reading, written production was featured in five out of six tasks. Here, all five tasks required writing at level B1 resulting in an average level of B1 (1c). With those three figures for reading, speaking, and writing as well as 11 out of 16 assessments assigning level B1, the maximum relative frequency was likewise at that level (2).

	Skill			
COMPARE	Reading (1a)	Speaking (1b)	Writing (1c)	max. relative frequency (2)
	A2+	B1	B1	B1

Table 8.41 - CEFR levels for COMPARE per skill and max. relative frequency across all skills, Diercke Geography 2

With a total of 12 applications, EXPLAIN was the discourse function used most often for tasks of AFB 2. All 12 tasks required *reading for information and argument,* and the assigned CEFR levels were between A2 and B2. The task requiring the latter CEFR level asked students to "EXPLAIN why more and more people become dissatisfied with the EU." (Ahrend 2013: 121, 2). To accomplish this, students would need to retrieve both direct and indirect information from the text and understand most aspects of personal opinions given, which they then would have to categorize and abstract to comply with the task's mandate satisfactorily. Hence, level B2 was assigned. Nevertheless, the average CEFR level for reading was calculated to be B1 (see Table 8.42, 1a). For *sustained monologue: giving information,* CEFR levels B1 and B2 were assigned among the 12 tasks. The average of those was level

B1+ (1b), which can be explained by the required increase in fluency, level of detail, and precision when explaining increasingly complex aspects. Out of the 12 tasks, ten had a written part to them with CEFR levels between A2 and B1. With the most of those tasks requiring B1, the average level for *written reports and essays* arrived at level B1 as well (1c). Across all three skills, the CEFR level B1 had the maximum relative frequency for EXPLAIN (2) with 73% assessments being at that level.

	Skill			
EXPLAIN	Reading (1a)	Speaking (1b)	Writing (1c)	max. relative frequency (2)
	B1	B1+	B1	B1

Table 8.42 - CEFR levels for EXPLAIN per skill and max. relative frequency across all skills, Diercke Geography 2

Moving on to the newly introduced discourse functions allocated to AFB 2, CLASSIFY and ANALYZE, an initial description of the actual underlying activity is called for to understand the purpose and the preliminary evaluations of the discourse functions. As per definition, the underlying intention of CLASSIFY is that students know characteristic aspects of distinct spatial or factual categories and recognize those in descriptive information about the shape of a landscape, e.g., enabling them to classify said landscape with one of the categories correctly (see Table 8.37). Depending on the material available, reading skills would have to be at least at level A2, especially when dealing with brief descriptive notes provided with visual aids. The requirements could be as high as level B2 for reading skills if a complex, lengthy text provides the necessary information. Likewise, speaking skills would have to be at least at level A2 when giving information, with students being able to produce a sequence of sentences with connections in slow and simple language (cf. Council of Europe 2018: 71f.). Again, speaking skills for *giving information* or *putting a case* could increase in demand up to level B2 to sufficiently carry out the spoken part of the task with increased precision, level of detail, and accuracy. The written part could require skills to be between A2, in cases series of linked simple sentences suffice, to level B1 asking for more sophisticated writing summarizing aspects in a linear sequence with an increased level of detail (cf. Council of Europe 2018: 77). The material taken from *Diercke Geography Volume 2* employed CLASSIFY only with one task in the thematic context of the European Union. This task asked students to CLASSIFY statements on people's takes on life in the EU and come up with suitable categories to sort the statements (cf. Ahrend 2013: 107). Since the material provided the statements as an edited transcript of actual utterances, the required reading skill for *reading for information and argument* was level A2 with the requirement of understanding the main points of short texts along with the ability to pick out the main information (cf. Council of Europe 2018: 63) (see Table 8.43, 1a). When elaborating on the categories decided for, the required level for *sustained monologue: giving information* to initially describe the categories followed by *sustained monologue: putting a case* to explain and justify their decisions, students would have to display speaking skills at level B1 (1b). They would have to "[…] explain the main points with reasonable precision" and "[…] briefly give reasons and explanations for opinions […]" (Council of Europe 2018: 71f.). A written part was ruled out for this task since it would

have been limited to maximally jotting down the headings. Consequently, the maximum relative frequency for CLASSIFY resulted in level B1 across the two actually assessed sets of skills (2).

	Skill			
CLASSIFY	Reading (1a)	Speaking (1b)	Writing (1c)	max. relative frequency (2)
	A2	B1	-	B1

Table 8.43 - CEFR levels for CLASSIFY per skill and max. relative frequency across all skills, Diercke Geography 2

ANALYZE requires students to "study or examine something in detail in order to understand and explain it" (Ministerium für Schule und Bildung des Landes Nordrhein-Westfalen 2015a), which already conveys the necessity of understanding texts and information on a deeper level. Consequently, the skills needed when *reading for information and argument* when working in the context of the discourse function ANALYZE start at level A2 as a minimum and could go up to level C1, depending on the length and complexity, the familiarity of students with the topic, and the manner in which information is given, that is either direct or indirect (see also chapter 12.6.2). After the cognitive processing of the information obtained through reading, students explain what they understood through their analysis, i.e., they have to transform their cognitive processes into speech. This could require speaking skills starting from level B1, at which increased fluency along with an increased level of detail and increased precision can be assumed when talking (cf. Council of Europe 2018: 69ff.). The requirements could even go up to level C1 when *giving information* with tasks that call for a "[…] clearly detailed distinction between ideas, concepts and things that closely resemble one other" (Council of Europe 2018: 71). When *putting a case*, argumentative spoken skills could go as high as level B2. Those cases ask students to "develop a clear argument, expanding and supporting their points of view at some length with subsidiary points and relevant examples" (ibid.: 71) to justify their explanations following the analysis. The possible requirements for *written reports and essays* follow suit and thus range from level B1 for summaries containing accumulated factual information (cf. Council of Europe 2018: 77) to level C1 in cases that ask for a "[…] clear, well-structured exposition of complex subjects, underlining the relevant salient issues." (ibid.). The four tasks featured in the material from *Diercke Geography Volume 2* did not require such high levels of skills. For two out of four tasks, the analysis concerned a cartoon, one of which was even in German. The required reading skills ranged from A1 for the English cartoon to B2 for a collection of extracts taken from official documents of and on the European Union (cf. Ahrend 2013: 108 f.). This wide range resulted in an average demand imposed by *reading for information and argument* for *analyze* at CEFR level B1 (see Table 8.44, 1a). To sufficiently fulfill the requirements of *sustained monologue: giving information,* the level of those four tasks was at B1. Consequently, the average required skill was at level B1 (1b). Furthermore, *written reports and essays* required skills between level B1 and B2, leading to an average level of B1+ across all four tasks for written production (1c).

With all three skills averaging at level B1 and B1+, respectively, the maximum relative frequency across all three skills was also with CEFR B1 (2).

ANALYZE	Skill			
	Reading (1a)	Speaking (1b)	Writing (1c)	max. relative frequency (2)
	B1	B1	B1+	B1

Table 8.44 - CEFR levels for ANALYZE per skill and max. relative frequency across all skills, Diercke Geography 2

The addition to AFB 3, EVALUATE, has to be broken down first into the possibly required levels of skill before discussing the actually required levels as evaluated based on the tasks (see chapter 12.6.2). To EVALUATE on aspects related to the subject, students would have to come to an understanding that allows for a qualified opinion that is supported by "factual information and knowledge on the subject matter" as well as being able to "disclose or reflect upon criteria applied" (Ministerium für Schule und Bildung des Landes Nordrhein-Westfalen 2015a). As per this definition, the level of understanding texts when *reading for information and argument* would have to be at least at B1 for "straightforward, factual texts" (Council of Europe 2018: 63) and could increase to level C1 with texts more complex requiring students to "identif[y] finer points of detail including attitudes as well as stated opinions" (ibid.). When students are asked to put their evaluation into words, the skill level required for *sustained monologue: putting a case* could range from B1 with students "giv[ing] simple reasons to justify a viewpoint" (Council of Europe 2018: 72) to level C1, allowing students to "[…] argue a case on a complex issue, formulating points precisely and employing emphasis effectively" (ibid.). Composing *written reports and essays* may require skills between levels B1 and C1. The single task employing EVALUATE required reading skills at level B1 (see Table 8.45, 1a), as was the case with speaking (1b). Only the written part was more challenging being assessed to be at level B2 (1c). The CEFR level assigned most often for the discourse function across all three skills was B1 (2).

EVALUATE	Skill			
	Reading (1a)	Speaking (1b)	Writing (1c)	max. relative frequency (2)
	B1	B1	B2	B1

Table 8.45 - CEFR levels for EVALUATE per skill and max. relative frequency across all skills, Diercke Geography 2

The next discourse function was the one used most in AFB 3 for *Diercke Geography Volume 2*. A total of six tasks asked students to DISCUSS a given topic, aspect, or question. Since DISCUSS was already used in the first volume, further elaboration on possible requirements outside of the actual task is not necessary, and the requirements imposed by said tasks can be described directly. As for *reading for information and argument*, the tasks required reading skills at level B1 in unison. Accordingly, the texts were of increasing length and complexity but provided information in a straightforward manner allowing students to understand factual, relevant points without further challenges (cf. Council of Europe 2018: 63). Accordingly, the average calculated CEFR level also arrived at level B1 (see Table 8.46, 1a). Moreover, this discourse function required speaking skills at level B1 for *sustained monologue: giving information*. When *putting a case*, students would have to display also skills at level

B2 for two out of six tasks to show increased reliability and level of detail along with systematicity when laying out their arguments (ibid.: 72). As a result, the average skill for speaking for DISCUSS arrived at level B1+ (1b). Discussions are basically regarded as oral discourse only. Hence the written part of this discourse function was left out of the evaluation. Consequently, out of all assessments per subskill for the tasks featuring DISCUSS, level B1 was assigned most often leading to that level to also be the one at the maximum relative frequency (2).

	Skill			
DISCUSS	Reading (1a)	Speaking (1b)	Writing (1c)	max. relative frequency (2)
	B1	B1+	-	B1

Table 8.46 - CEFR levels for DISCUSS per skill and max. relative frequency across all skills, Diercke Geography 2

However similar the next discourse function may seem to DISCUSS, COMMENT ON still entails a different activity. DISCUSS is about the process of coming to a reasoned opinion through "talking about something in detail […]" (Ministerium für Schule und Bildung des Landes Nordrhein-Westfalen 2015a) while COMMENT ON bases on the assumption that this reasoned opinion is already found and students are to express thoughts and ideas on that (ibid.). Having clarified that, the reading skill for the four tasks was between levels A2 and B1. Tasks asking students to comment on brief statements on aspects appearing as transcripts of interviews or, e.g., a bullet point list of facts on the port of Shanghai (cf. Ahrend 2013: 35, 39). The length, directness, and simplicity in which information was provided did not allow for a higher level of required reading skills. The average for reading resulted in level A2+, accordingly (see Table 8.47, 1a). The spoken part was more demanding. COMMENT ON mainly required *sustained monologue: putting a case* as opinions should be – per definition – reasoned, and thus, the required skill here was at level B1 for all four tasks. Students would have to display abilities to voice their opinions in a linear sequence of points with an increased level of detail and fluency and be able to elaborate on simple reasoning (cf. Council of Europe 2018: 72). The average for this skill also arrived at level B1 (1b). Other than DISUCSS, the discourse function COMMENT ON allowed for a written part. To comment on the given aspects in writing, students would have to be able to write on level B1 to sufficiently and satisfyingly carry out the task. This entails straightforward connected text in a linear sequence, including an increased level of detail when summarizing while still using a rather simple language (ibid.: 75). As all four tasks were at level B1, the average for this skill also arrived at level B1 (1c). Across all three skills, level B1 was assigned most often in relation to the total of 12 assessments (2).

	Skill			
COMMENT ON	Reading (1a)	Speaking (1b)	Writing (1c)	max. relative frequency (2)
	A2+	B1	B1	B1

Table 8.47 - CEFR levels for COMMENT ON per skill and max. relative frequency across all skills, Diercke Geography 2

Lastly, AFB 3 entails another, newly added discourse function, JUSTIFY. This discourse function is defined to "develop a reasoned argument and highlight thematic correlations/connections on

complex/basic considerations or theses" (Ministerium für Schule und Bildung des Landes Nordrhein-Westfalen 2015a), seemingly overlapping with abovementioned discourse functions DISCUSS and COMMENT ON. However, the difference is that JUSTIFY asks students to come up with their own reasons and arguments without discussing the aspect with their peers. This is evident in the task found in Diercke Geography Volume 2, which utilizes the discourse function of "[JUSTIFY] your arrangement" (Ahrend 2013: 107). This task reflects on a previous activity where students were required to order a collection of statements provided by EU citizens, using a categorization derived from these statements. Accordingly, students would have to explain and through that support their order. Generally, JUSTIFY could require reading skills starting at level B1 for straightforward, factual texts that give the required information directly, to level C1 in cases that require reading a long and complex text including the writer's own opinion or attitude (cf. Council of Europe 2018: 63). In the specific case of the task given above, the required reading skill was at level A2 (see Table 8.48, 1a), since this task was a follow-up task based on the work completed so far. Additionally, the statements given were brief and in comparably simple language. The spoken part of JUSTIFY, *sustained monologue: putting a case*, was generally analyzed to possibly require speaking levels between B1 and C1. Level B1 was as much an option for simple reasoning in students' own words as was level C1 for complex issues which require "formulating points precisely and employing emphasis effectively" (Council of Europe 2018: 72). In the specific given task here, the required level of skill for *sustained monologue: putting a case* to justify the order and categorization of statements was B1 (1b) as students had to give reasons for their choice in their own words (Council of Europe 2018: 72). Although this specific task did not require anything in writing (1c), written skills, in general, could range from B2 with *written reports and essays* "which develop an argument, giving reasons in support of or against a particular point of view […]" (ibid.: 77) to C1 requiring the writing of "clear, well-structured expositions of complex issues […] and expand and support points of view at some length with […] relevant examples." (ibid.). With only two assessments in total, a maximum relative frequency could not be calculated, or much rather was inconclusive since A2 and B1 were assigned once. However, the demand for reading at B1 indicates the demand this task using JUSTIFY.

	Skill			
JUSTIFY	Reading (1a)	Speaking (1b)	Writing (1c)	max. relative frequency (2)
	A2	B1	-	n.a.

Table 8.48 - CEFR levels for JUSTIFY per skill and max. relative frequency across all skills, Diercke Geography 2

The following figure (see Figure 8.10) depicts all results for the discourse functions used in *Diercke Geography Volume 2* and their allocated CEFR levels per skill as an average. Likewise, the maximum relative frequencies across the three skills for each discourse function are displayed.

Again, the level of skill students should have achieved as per the curricular standards (cf. KMK 2003) in the respective grades is represented by the red-colored zone, whereas the three different skills are allocated with the CEFL levels indicated per column.
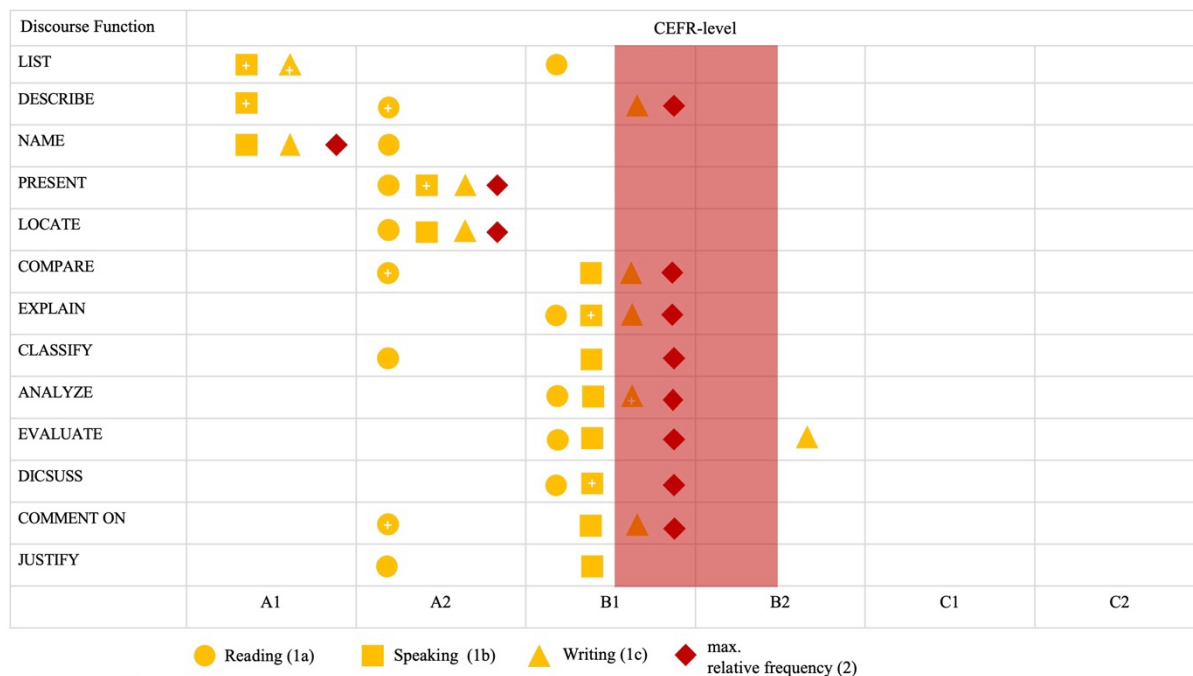
Figure 8.10 - Average CEFR levels per discourse function and skill & max. relative frequency per discourse function across, Diercke Geography Volume 2

Similarly, the final step of the analysis calculated the average required level per skill across all tasks. Apart from writing, the average levels increased for reading and speaking from A2+ with volume 1 to B1 with volume 2. This increase reflects students' assumed development from grades 7 and 8 to grades 9 and 10. The overall maximum relative frequency of CEFR levels assigned for all assessments of tasks taken from *Diercke Geography Volume 2* arrived at level B1, which appears completely in line with the targeted level as defined by the administrative regulations (cf. KMK 2003). The discussion in chapter 8.4 comments on the implications of those results.
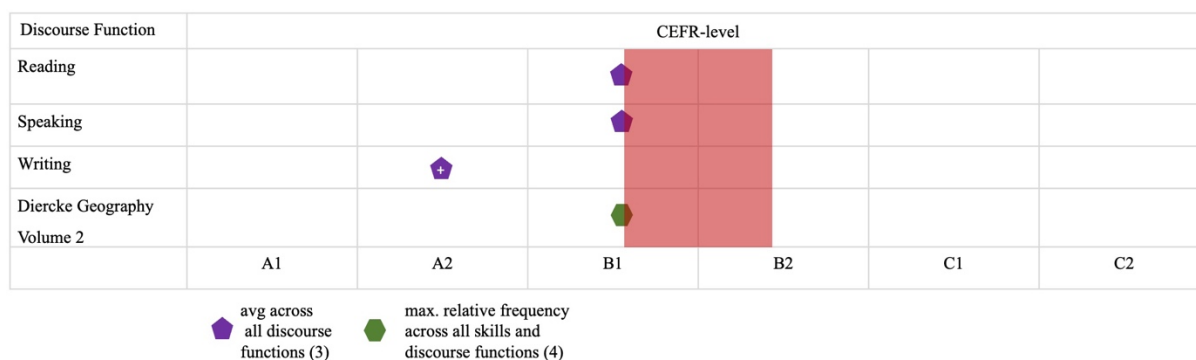
Figure 8.11 - Average CEFR levels per skill across discourse function and max. relative frequency of Diercke Geography Volume 2

8.3.3 Biology

The material selected from *Discover Biology Volume 1* featured 14 different discourse functions. The KMK has compiled a list of discourse functions frequently used in science classes (Physics, Biology, Chemistry) for German schools in foreign countries (cf. KMK 2013b). This general list is specialized for "Gymnasium" and provides all discourse functions and elaborates on the activity connected to or the purpose of the discourse function. It is more extensive than the list explicitly dedicated to bilingual teaching in Biology (cf. Ministerium für Schule und Bildung des Landes Nordrhein-Westfalen 2015). Hence, a comprehensive list was compiled combining the available lists mapping German discourse functions to their English equivalents. This is a preliminary step toward mapping those discourse functions to CEFR descriptors. The next step was, just like with History and Geography, to map the discourse functions onto all potential levels within the CEFR for each skill (see chapter 12.6.3). Based on these preliminary steps, the actual analysis was possible.

Before turning to the results of the analysis of the material selected to represent potential linguistic requirements at grade 7, the discourse functions and their definitions as used in the material are listed (see Table 8.49). The analysis looked at 70 tasks for grade 7. DESCRIBE and EXPLAIN were the discourse functions used most often, both allocated in AFB 2. Accordingly, this frequency represents this level of cognitive demand the most, while discourse functions of AFB 3 followed in second place. Interestingly, tasks and discourse functions allocated with AFB 1 were used least often.

Discourse Function	Definition (cf. KMK 2013b; Ministerium für Schule und Bildung des Landes Nordrhein-Westfalen 2015)	AFB	Frequency
DRAW	create a picture, diagram, or model of observable or given structures using an appropriate level of precision	1	1
NAME	attribute terms and actual situations to a given structure	1	4
PRESENT	to give an account for/ to illustrate actual situations correlations, methods, events, etc. in a structured manner	1	2
APPLY	relate obtained (biological) knowledge to a new problem/question at hand	2	4
COMPARE	Put actual situations, objects, living things and processes next to each other and decide what they have in common or what differentiates them – which factors are the same, similar, different or opposite	2	5
DESCRIBE	give details about what structures, actual situations or correlations are like using own words and appropriate subject-specific terminology	2	15
EXPLAIN	give information about an actual situation, contextualize regarding rules, principles and causes to help to understand it better	2	27

ILLUSTRATE	put essential aspects on an actual situation/object/process across using examples and additional information	2	1
DISCUSS	talk about statements or theses in detail, contrast and come to a reasoned conclusion	3	2
GIVE AN OPINION	express a reasoned opinion based on the application of subject-related knowledge and methodology	3	5
HYPOTHESIZE	word a founded assumption concerning a biological phenomenon	3	2
JUSTIFY	attribute actual situations to rules, principles, or casual connections	3	2

Table 8.49 - List of discourse functions, definitions used in Discover Biology Volume 1, grades 7 and 8 material

However, following the order from AFB 1 to AFB 3, the first discourse function analyzed was DRAW. Students would have to read and understand the task to a certain depth of detail and be able to elaborate briefly on their proceedings. Accordingly, reading skills could be required to be between level A2 allowing for students to "[…] identify specific information in simpler written material […]" (Council of Europe 2018: 63) and level B2 in cases that involve complex texts from "highly specialized sources" (ibid.). With drawings or images, the demand for students speaking skills could range from level A1 for single isolated words or phrases (cf. Council of Europe 2018: 71) to level B2 for "[…] clear, detailed descriptions of how to carry out a procedure" (ibid.). Lastly, written production and its potential necessary CEFR level could be at level A1 since with drawings, only single terms are needed, if at all (see chapter 12.6.3). In the case of the single task of the selected material, the required level for *reading for information and argument* was B1 as it asked students to DRAW cause-and-effect diagrams for which a long, increasingly complex text providing straightforward information (cf. Mathews & Olmesdahl 2010: 35, task 1) had to be read (see Table 8.50, 1a). The spoken part, i.e., the elaboration on the drawing, was not evaluated (1b) because the task featured a specific discourse function for that part, DESCRIBE. As was anticipated in the initial assessment, the written part only involved single phrases and terms and was thus assessed to be at level A1 (1c). With only two assessments in total, the maximum relative frequency would not provide useful information to determine the demand of this discourse function as such (2).

	Skill			
DRAW	Reading (1a)	Speaking (1b)	Writing (1c)	max. relative frequency (2)
	B1	-	A1	n.a.

Table 8.50 - CEFR levels for DRAW per skill and max. relative frequency across all skills, Discover Biology 1, grades 7 and 8

The potentially required levels of skill for NAME are comparable to those allocated with DRAW. When reading a task to identify the information to be named, students would at least have to display reading skills at level A2 to "[…] find specific information in practical, concrete, predictable texts […]" (Council of Europe 2018: 62). This demand could increase to level B1, especially when the text's length increases. Depending on the quality and quantity of information that has to be given when fulfilling the

spoken part of the discourse function, the required CEFR levels could vary between A1 for single terms and B1 to "[…] report on straightforward factual information on a familiar topic, e.g., to indicate the nature of a problem[…]" (Council of Europe 2018: 71). The anticipated maximum level of skill for writing was at level A2 at which the demand extends to a "[…] series of simple sentences" (ibid.: 77). The actual demand for reading for the four tasks featuring NAME was between levels A2 and B1 with an average required level at A2+ (see Table 8.51 1a). For the spoken production, including subskills *sustained monologue: giving information, describing experience,* and *overall spoken production,* levels varied between A1 and A2 but averaged out at A2 (1b). As anticipated, the written part of those four tasks demanded writing skills between levels A1 and A2. The average here was calculated to be A1+ again (1c). The maximum relative frequency of all assigned CEFR levels across all three skills and their subskills arrived at level A2 (2).

NAME	Skill			max. relative frequency (2)
	Reading (1a)	Speaking (1b)	Writing (1c)	
	A2+	A2	A1+	A2

Table 8.51 - CEFR levels for NAME per skill and max. relative frequency across all skills, Discover Biology 1, grades 7 and 8

The following discourse function, PRESENT, ties in with the skill *addressing audiences.* By definition, students have to "[…] give an account of […] actual situations, correlations, methods, events in a structured manner" (KMK 2013b). To prepare for such a presentation, students would have to read for information and argument. They would have to display the ability to at least " […] identify specific information in simpler written material […]" (Council of Europe 2018: 63), which places the potential minimum of required reading skills at level A2. This demand could increase to level B1 in cases that either need a deeper, more detailed understanding or which entail longer complex texts (ibid.). The necessary level of skill to *address[ing] audiences* could range from level A2, at which students "can give a short, rehearsed basic presentation [and] can cope with a limited number of questions" (Council of Europe 2018: 74). This could increase to level B2 at which both the level of sophistication of the presentation as such but also of the number and complexity of questions to handle increases significantly (ibid.). Most presentations limit the required level for written products to A2 with simple phrases and sentences. In cases requiring connected text, the demand could increase to level B1 as well. With the given tasks, the level of written products was at B1 (see Table 8.52, 1c). The tasks asked students to prepare a complete text to present their findings, providing them with a list of subject-related terms to be used. Accordingly, a *written report and essays* composed of "straightforward connected text […] within their field of interest" (Council of Europe 2018: 77) was required. Reading skills were at level B1 (1a) as was the actual *addressing* of an *audience* (1b). With the two tasks employing PRESENT, level B1 was assigned most often.

Hence, the maximum relative frequency arrived at this level (2).

	Skill			
PRESENT	Reading (1a)	Speaking (1b)	Writing (1c)	max. relative frequency (2)
	B1	B1	B1	B1

Table 8.52 - CEFR levels for PRESENT per skill and max. relative frequency across all skills, Discover Biology 1, grades 7 and 8

Continuing with APPLY, this discourse function and its definition entail that students either base their work on prior knowledge or knowledge they have recently gathered, e.g., through reading. They would have to identify the underlying concept, transfer it to a specific context, and elaborate on their proceedings. Therefore, students would have to be able to at least "[…] understand the main points of short texts […] pick out the main information […]" (Council of Europe 2018: 62f.). Accordingly, the minimally required reading skills would have to be at level A2 and could increase to level B1 in cases that need a deeper level of understanding than just select pieces of information. Students would have to display speaking skills starting from level A2 to elaborate on the knowledge and how they applied it. Furthermore, the demand could potentially go up to level B1 when *giving information* as students would have to be able to "[…] explain the main points in an idea or problem with reasonable precision." (Council of Europe 2018: 71). The same selection of CEFR levels could be necessary when composing *written reports and essays.* "Short series of linked sentences […]" (Council of Europe 2018: 77) could suffice but might as well have to be extended to "[…] brief reports to a standard conventionalized format, which pass on routine factual information […]" (ibid.). The four tasks analyzed revealed that the CEFR levels *reading for information and argument* varied between levels A2 and B1. The calculated average for reading skills arrived at level A2+ (see Table 8.53, 1a). The same spread applied to the spoken production part of this discourse function. As the assessment for all tasks was equally distributed onto levels A2 and B1, the average was also A2+ (1b). Lastly, the CEFR level identified as necessary for written production was, on average, A2 (1c). Of 12 assigned CEFR levels, six were level A2, marking a 50% share and thus, the maximum relative frequency was at this level (2).

	Skill			
APPLY	Reading (1a)	Speaking (1b)	Writing (1c)	max. relative frequency (2)
	A2+	A2+	A2	A2

Table 8.53 - CEFR levels for APPLY per skill and max. relative frequency across all skills, Discover Biology 1, grades 7 and 8

To determine similarities and differences between two or more given aspects, students have to carry out the discourse function COMPARE and its related activities. As a first step, students have to *read for information and argument* displaying at least skills at level A2, allowing them to "[…] understand the main points […], pick out the main information" (Council of Europe 2018: 63) of texts that come with visual support. In cases dealing with increasingly specialized information, the demanded level of skill could go up to level B2. Following the gathering and evaluating of information, students would have to verbalize their findings. For this, the level of speaking skills should minimally be at A2

to "[…] give a simple description or presentation as a short series of simple phrases […]" (ibid.: 69). This skill or much rather the quality demanded could increase to level B2, depending on the textual basis, at which "[…] systematically developed descriptions and presentations, with appropriate highlighting of significant points […]" (ibid.) would be expected to be displayed by students. A *written report and essay* that "[…] summarize[s], report[s], and give[s] an opinion on accumulated factual information […]" (ibid.: 77) would require writing skills at level B1. The five comparisons featured in the selected material required reading skills at level A2, with four out of five tasks asking for this level of reading skill. The fifth task asked students to compare two schematic drawings for which reading was not necessary (cf. Mathews & Olmesdahl 2010: 48) and for which further reading was not necessary. Based on the four assigned CEFR levels, the average level for *reading for information and argument* was at level A2 (see Table 8.54, 1a). When talking about their observations and findings connected to the five tasks, students in all cases were required to display skills of *sustained monologue: describing experience* at level A2, allowing them to "[…] use simple descriptive language to […] compare objects […]" (Council of Europe 2018: 70). With such homogeneous individual assessments, the average level of skill arrived at level A2 (1b) as well. For the written part, the individually assigned CEFR levels varied between A2 and B1, causing the average level for writing to arrive at level A2+ (1c). All the determined CEFR levels per skill combined placed the maximum relative frequency for COMPARE at level A2 (2).

COMPARE	Skill			
	Reading (1a)	Speaking (1b)	Writing (1c)	max. relative frequency (2)
	A2	A2	A2+	A2

Table 8.54 - CEFR levels for COMPARE per skill and max. relative frequency across all skills, Discover Biology 1, grades 7 and 8

The second most frequently used discourse function asks students to DESCRIBE the given matter. The analysis counted 15 uses. To be able to DESCRIBE something with a satisfying level of detail, students would have to initially show reading skills starting at level A1 in cases that provide visual support. The required level of understanding, complexity of the text and detail could increase to the extent of reading skills resembling the descriptions provided with level B1. Here students would be able to understand "[…] straightforward factual texts […] and the main points in descriptive notes […]" (Council of Europe 2018: 63). To carry out the discourse function, speaking skills at level A1 could already suffice when "[…] giving a short description of an object or picture while showing it to others using basic words, phrases, and formulaic expressions […]" (ibid.: 71) with sufficient preparation time. The required level of speaking skills could increase to level B1 in dependence on the level of detail and precision and the level of complexity needed in the actual description (ibid.). *Written reports and essays* in connection with the discourse function would minimally require writing skills at level A2 for "[…] simple texts on familiar topics […]" (Council of Europe 2018: 77). However, they could also demand level B1, mainly when being tasked to adhere to a specific "[…] standard conventionalized format" (ibid.) which could apply to the written records of observations of experiments in Biology. The 15 tasks

within *Discover Biology Volume 1* asked for reading skills placed at either level A2 or level B1, while most tasks required the former level. The average level across all tasks for *reading for information and argument* arrived at level A2+ (see Table 8.55, 1a). The same average level was also calculated (1b) for *sustained monologue: giving information* or *describing experience,* amongst which both levels A2 and B1 were evenly distributed during the individual evaluation of the required level of skill per task. While both reading and speaking required level B1 in single tasks, *written reports and essays* in all cases demanded level A2 causing the overall average level for that skill to also be at level A2 (1c). With a majority of all assessments at level A2, namely 27 out of 35, the maximum relative frequency was also at level A2 with 77% (2).

	Skill			
DESCRIBE	Reading (1a)	Speaking (1b)	Writing (1c)	max. relative frequency (2)
	A2+	A2+	A2	A2

Table 8.55 - CEFR levels for DESCRIBE per skill and max. relative frequency across all skills, Discover Biology 1, grades 7 and 8

EXPLAIN was the discourse function employed most often among the selected tasks. In total, 27 tasks asked students to "give information about an actual situation, to contextualize regarding rules, principles and causes to help to understand[…]" (KMK 2013b). Again, this definition was different than the ones given for EXPLAIN in History (see Table 8.5) and Geography (see Table 8.25). The considerations (see chapter 8.2.1) preceding the assessment of the actual tasks deemed reading skills between levels A2 and C1 as potentially suitable, with the latter applying to texts that can be found within an academic background providing an extended level of detail (cf. Council of Europe 2018: 63). The actual explaining, i.e., verbalizing their thoughts and ideas on the matter, could require students to display skills of *sustained monologue: giving information* at level A2. An increase to level C1 is possible in cases that require the communication of "[…] clearly detailed distinctions between ideas or concepts […]" (ibid.: 71). For written recordings of the explanations, skills starting at level A2 for *written reports and essays* could suffice, whereas an increase to level C1 would also be possible especially in cases that require "[…] clear, well-structured expositions of complex subjects underlining the relevant, salient issues." (ibid.: 77). For the material selected and targeted at grade 7, however, the higher levels were not expected to apply when considering possibly required levels of skill. The actual tasks and the findings of the analysis supported that expectation. Between the 27 tasks, the required level of reading skills varied between levels A2 and B1, which asked students to gather "[…] factual, relevant points" from "[…] straightforward longer texts" (ibid.: 63). With a majority of tasks at level B1, the average level of skill calculated for reading for information and argument arrived at level B1 (see Table 8.56, 1a). Verbalizing their findings required students mostly to display speaking skills at level B1. Only one task was evaluated to demand level B2 as it asked for complex elaborations on the exchange process of oxygenation of blood in higher altitudes (cf. Mathews & Olmesdahl 2010: 2). Single tasks required speaking skills at level A2 to be carried out sufficiently. Nevertheless, the average level of speaking

needed in the context of the selected tasks arrived at level B1 (1b). To record the explanations, the levels of written skills varied between A2 and B1 for all but one task, which was again the above-described tasks requiring level B2 for writing skills. On average, written reports and essays were placed at level A2+ to sufficiently carry out this part of the discourse function (1c). A total of 43 of 71 assigned CEFR levels was at B1 and hence, the maximum relative frequency arrived at that level (2).

	Skill			
EXPLAIN	Reading (1a)	Speaking (1b)	Writing (1c)	max. relative frequency (2)
	B1	B1	A2+	B1

Table 8.56 - CEFR levels for EXPLAIN per skill and max. relative frequency across all skills, Discover Biology 1, grades 7 and 8

The final out of five discourse functions allocated in AFB 2 was ILLUSTRATE. Although the term seems to allude to a graphical representation, ILLUSTRATE is an activity close to EXPLAIN as it entails conveying information on "essential aspects on an actual situation, object, or process" (KMK 2013b). Level A2 would suffice to manage the required reading for the discourse function. At this level, students "can understand the main points of a text [and] can pick out the main information […]" (Council of Europe 2018: 63). But with more complex aspects to be illustrated or explained, the demanded level of reading skill could increase to level B2, especially when information is obtained from highly specialized sources (ibid.). To verbally ILLUSTRATE any given aspect, students would at least have to be able to "[…] give a simple description or short presentation […] as a short series of simple phrases and sentences linked into a list." (ibid.: 69), implying CEFR level A2. More likely, students would have to employ skills resembling the description of level B1, which requires them to be able to "[…] explain the main points in an idea or problem with reasonable precision." (ibid.: 71). Very complex cases could also require speaking skills at level B2. Putting illustrations or elaborations on a given aspect with the mandate to clarify and exemplify into writing could require writing skills for *written reports and essays* between levels A2 and B1. The latter should be sufficient as a maximum required level as it targets the writing of "[…] very brief reports to a standard conventionalized format […]" (ibid.: 77), which fitted the context of the subject and grade analyzed. The single task featuring ILLUSTRATE was on par with the preceding general assessments. Asking students to "ILLUSTRATE the importance of plasma to blood." (Mathews & Olmesdahl 2010: 47, 3) required *reading for information and argument* of a rather lengthy and complex text and understanding factual, relevant points (cf. Council of Europe 2018: 63). This description corresponded to CEFR level B1 (1a). To verbalize the aspects being asked, the level of skill required for *sustained monologue: giving information* was placed at level B1 (1b). Writing skills were assessed to be necessary at the same level (1c) and hence, overall assessment of ILLUSTRATE in that single case was also B1. However, a single task and its assessed CEFR level did not suffice to compute a relative frequency (2).

The remaining four discourse functions, DISCUSS, GIVE AN OPINION, HYPOTHESIZE, and JUSTIFY are all placed with AFB 3. Students engage in detailed discourse and compare statements during discussions to develop a reasoned opinion. These statements and the essential information needed

to enter a discussion could be obtained through reading, which would at least have to be at level B1. At this level, students would be able to "[…] understand straightforward, factual texts of subjects relating to their study […]" (Council of Europe 2018: 63). This demanded level could increase to level C1 in cases concerned with complex, lengthy texts requiring an understanding of finer details (ibid.). Likewise, speaking skills for both *sustained monologue: giving information* and *putting case* could range between B1 and C1. Although only rarely put in writing, if students were to compose a *written report and essay* on the discussion, writing skills would need to start at least at level A2 and could increase to level C1.

Of the selected material, two tasks featured DISCUSS. Both required reading skills to be at level B1 leading to an average level of that skill to be the same (see Table 8.57, 1a). The same level was asked for the subskills of spoken production (1b), i.e., talking about the task, one concerned with lung capacity (cf. Mathews & Olmesdahl 2010: 41, 3), and one with the body's reaction to a respiratory tract infection (ibid.: 56). One of the two tasks did allow for a written part based the context it was embedded in. The necessary level to compose a *written report and essay* was placed with CEFR level A2 (1c). As the majority of assigned CEFR levels across the three skills was at level B1, the maximum relative frequency was also at this level with 67% of all ratings (2).

DISCUSS	Skill			max. relative frequency (2)
	Reading (1a)	Speaking (1b)	Writing (1c)	
	B1	B1	A2	B1

Table 8.57 - CEFR levels for DISCUSS per skill and max. relative frequency across all skills, Discover Biology 1, grades 7 and 8

While DISCUSS strives to come to a reasoned opinion through discourse, GIVE AN OPINION and connected activities take this a step further as students "express a reasoned opinion based on the application of subject-related knowledge and methodology" (KMK 2013b). Independent of the actual tasks, reading skills were generally assessed to be at least at level B1 to gather sufficient information from a text to carry out the discourse function properly. Cases involving increasingly complex and lengthy texts could also increase the demanded level of reading skills to B2 and C1. The same gradation could apply to skills related to *sustained monologue: putting a case.* The minimum level students would have to display would be B1; however, an increase to level C1 is possible based on the complexity and level of detail needed. The minimal level required to compose *written reports and essays* on their own opinion would be level B1. The companion's statement on competences connected to level B2, though, appears more fitting as students are able to "[…] write a report or essay which develops an argument, giving reasons in support of or against particular points of view […]" (Council of Europe 2018: 77). Although an increase to even level C1 is generally possible in connection with GIVE AN OPINION, this level seemed unlikely in the context of the selected material. In total, five tasks mandated students to come to an opinion on select biological questions. The necessary levels of reading skills varied between A2 and B1, with four out of five tasks were placed with the higher level. Accordingly, the average level of reading required with the selected tasks arrived at level B1 (see Table 8.58 1a). To

elaborately GIVE AN OPINION on the five tasks and their contexts, the necessary level of skill for *sustained monologue: putting a case* was at an average CEFR level B1 (1b). Contrary to the ex-ante assessment of written skills, *written reports and essays* for two out of five tasks sufficed at level A2, while the one task asked for skills at level B1. The remaining two tasks did not ask for a written part. Hence, the average level for writing skills was calculated to be at A2+ (1c). In 62% of all assessments level B1 was assigned, leading to this level to also be the maximum relative frequency (2).

GIVE AN OPINION	Skill			
	Reading (1a)	Speaking (1b)	Writing (1c)	max. relative frequency (2)
	B1	B1	A2+	B1

Table 8.58 - CEFR levels for GIVE AN OPINION per skill and max. relative frequency across all skills, Discover Biology 1, grades 7 and 8

Subjects that entail experiments also ask students to hypothesize or speculate on possible results or reactions before conducting the experiment. *Discover Biology Volume 1* follows suit featuring two tasks out of the selection, asking students to make assumptions based on their prior and subject-related knowledge for a given question or situation. To be able to obtain the information to base assumptions on, HYPOTHESIZE would generally require reading skills starting at level A2, at which the readers are able to gather essential basic information on the matter at hand (cf. Council of Europe 2018: 63). Depending on both the complexity of the experiment or question and the text providing the necessary information, this demand could increase to level C1, at which the ability to understand finer points in texts with a potentially unfamiliar context (ibid.). Speaking skills, either for *overall spoken production* or *sustained monologue: putting a case* could be required at the competences connected to CEFR level A2 to verbalize ideas or assumptions on possible outcomes. The more details and substantiation through examples are required, the higher is the linguistic demand. It could potentially increase to level B2, at which speakers "can develop an argument systemically with appropriate highlighting of significant points, and relevant supporting detail" (ibid.: 72). In cases that ask for a written record of a hypothesis, writing skills should at least be at level A2 to sufficiently carry out the task. This requires putting "[…] impressions and opinions in writing on topics of personal interest […] using basic everyday vocabulary and expressions" (ibid.: 77). While level B2 could be possible, level B1 fitted the discourse function's potential demand to "[…] summarize…about accumulated factual information…with some confidence […] using simple language […] give and justify their opinion" (ibid.) best. The analysis of the two tasks showed that one task on the question of the muscular movement of breathing during sleep (Mathews & Olmesdahl 2010: 31, 2) required reading skills at level A2, while the task on the increase in heart rates of trained athletes (cf. Mathews & Olmesdahl 2010: 57, C3) needed skills at level B1. Hence, the average reading skill connected to the material and the discourse function was at level A2+ (see Table 8.59, 1a). Students giving "briefly [...] reasons and explanations for opinions, plans and actions" (Council of Europe 2018: 72) while using simple language and being given time to prepare (ibid.) sufficed for spoken production. Since this level applied to both tasks, the average for this skill also arrived at level

B1 (1b). Lastly, both tasks entailed a written part, one at level A2 and the other at level B1. Accordingly, the average required level of skill for writing was calculated to be at level A2+ (1c). Out of seven assessments, four were at level B1, hence this level was also the maximum relative frequency (2).

HYPOTHESIZE	Skill			
	Reading (1a)	Speaking (1b)	Writing (1c)	max. relative frequency (2)
	A2+	B1	A2+	B1

Table 8.59 - CEFR levels for HYPOTHESIZE per skill and max. relative frequency across all skills, Discover Biology 1, grades 7 and 8

Although already partly included with the above-described discourse functions, JUSTIFY was also used on its own. JUSTIFY aims to transfer prior knowledge on principles and rules onto unknown situations, supported by the ability to prescind said situations onto a meta-level. This rather challenging cognitive process calls for reading skills that allow for the understanding of most and especially the most relevant aspects of a straightforward text (cf. Council of Europe 2018: 63), starting at level B1 and potentially increasing to level C1 depending on the length, complexity and necessity to "[…] identify[…] finer points of detail […]" (ibid.). Possible elaborations and explanations on the given aspect may require speaking skills starting at level B1, at which speakers "can give simple reasons to justify a viewpoint on a familiar topic" (ibid.: 72). This skill may increase to level C1 as well especially in cases which call for "argu[ing] a case on a complex issue, formulating points precisely and employing emphasis effectively" (ibid.). *Written reports and essays* need at least writing skills comparable to those described with level B1. At this level, the writer can "[…] summarize […] accumulated factual information […] with some confidence […] using simple language to give and justify their opinion" (ibid.: 77). For both tasks of the material featuring JUSTIFY, the demand for reading was placed with level B1 (see Table 8.60, 1a) as was the case with speaking (1b) and writing (1c). Accordingly, the maximum relative frequency with the assigned levels of skill required to carry out the discourse function JUSTIFY was B1 (2).

JUSTIFY	Skill			
	Reading (1a)	Speaking (1b)	Writing (1c)	max. relative frequency (2)
	B1	B1	B1	B1

Table 8.60 - CEFR levels for JUSTIFY per skill and max. relative frequency across all skills, Discover Biology 1, grades 7 and 8

All analyzed discourse functions and skills were between levels A2 and B1. Although one single task required level B2 for speaking and writing, the majority of tasks or the necessary level of skill to carry out the activities connected to each discourse function remained at the lower end of the potential spectrum set up before the analysis of the actual tasks. To further clarify the obtained results, all levels of skills and the average level per discourse function are placed on a grid.

The red-colored zone marks the targeted level of language skills as set with the educational standards by the KMK (cf. KMK 2003).
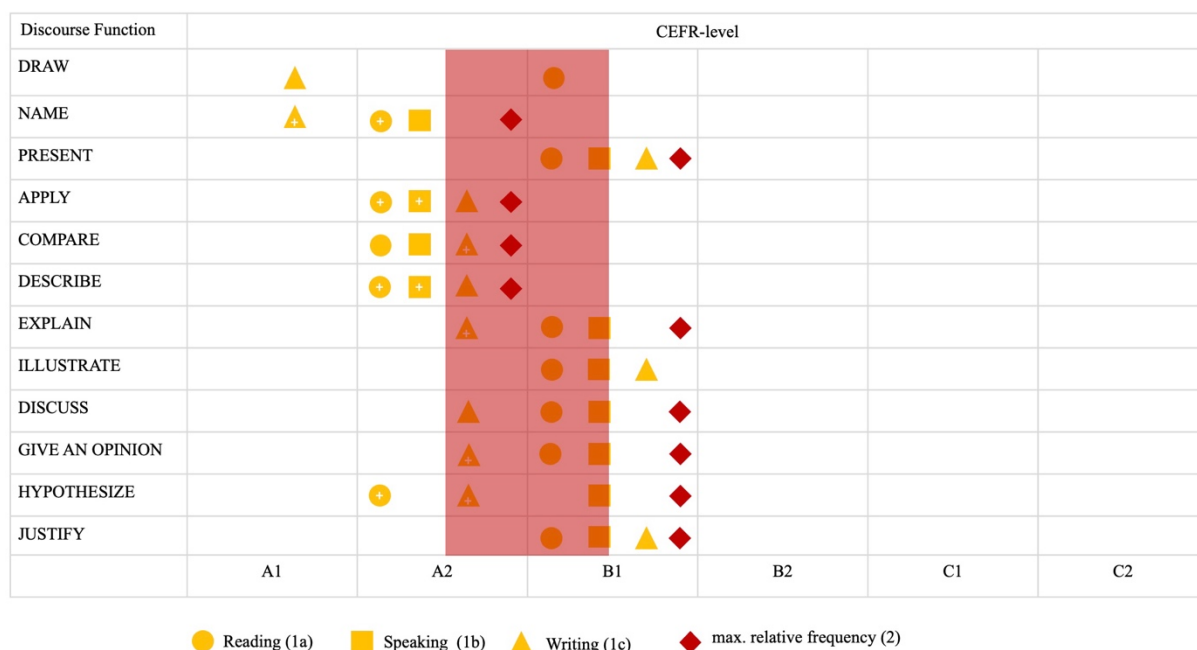
Figure 8.12 - Average CEFR levels per discourse function and skill & max. relative frequency per discourse function across skills, Discover Biology Volume 1, grades 7 & 8

Lastly, the averages for the skills across all discourse functions give an impression of the demand imposed on students being taught Biology bilingually. For all three skills, speaking, reading, and writing, the average level arrived at A2+ (see Figure 8.13, 3), which consequently led to an overall maximum relative frequency, i.e., including all 210 assessments made during the analysis, of B1 (4).
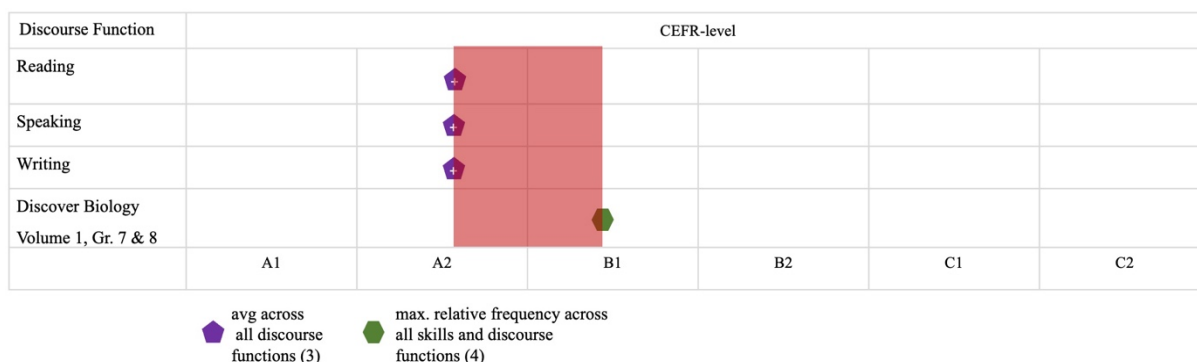
Figure 8.13 - Average CEFR levels per skill across discourse function and overall maximum relative frequency across discourse function, Discover Biology Volume 1, grades 7 and & 8

Discover Biology Volume 1, grades 9 and 10

The 57 tasks selected for bilingual Biology classes in grades 9 and 10 featured three additional discourse functions: LIST, RECORD, and DERIVE. Both LIST and RECORD are allocated with AFB 1 and aim at written production. The overview of discourse functions used in the material changes to this:

Discourse Function	Definition (cf. KMK 2013b; Ministerium für Schule und Bildung des Landes Nordrhein-Westfalen 2015)	AFB	Frequency
LIST	write down/ give information in a given order without commenting on it	1	6
RECORD	take precise notes on the test procedure, observations, results and possibly on analyses	1	2
DRAW	create a picture, diagram, or model of observable or given structures using an appropriate level of precision	1	3
PRESENT	to give an account for/ to illustrate actual situations correlations, methods, events, etc. in a structured manner	1	2
APPLY	relate obtained (biological) knowledge to a new problem/question at hand	2	3
COMPARE	put actual situations, objects, living things and processes next to each other and decide what they have in common or what differentiates them – which factors are the same, similar, different or opposite	2	3
DESCRIBE	give details about what structures, actual situations or correlations are like using own words and appropriate subject-specific terminology	2	9
EXPLAIN	give information about an actual situation, contextualize regarding rules, principles and causes to help to understand it better	2	17
DERIVE	come to an appropriate conclusion based on knowledge and findings	2	1
DISCUSS	talk about statements or theses in detail, contrast and come to a reasoned conclusion	3	5
HYPOTHESIZE	word a founded assumption concerning a biological phenomenon	3	1
JUSTIFY	attribute actual situations to rules, principles, or casual connections	3	2

Table 8.61 - List of discourse functions, definitions used in Discover Biology Volume 1, grades 9 and 10 material

Six tasks asked to LIST information as per the definition. In general, to read to be able to list information needs at least reading skills at level A2 as "specific information has to be found in […] texts provided they are written in simple language." (Council of Europe 2018: 62). The demand could increase to level B1 in cases concerning longer texts in which the targeted information has to be located in different parts of the text and relevant information has to be identified first (ibid.). When verbalizing the information identified to be suitable for the required list, speaking skills at level A1 could suffice if only the single terms have to be mentioned (ibid.: 71). If, however, the selection of specific information over

other information has to be explained, the necessary level of skill for *sustained monologue: giving information* could increase to level B1. At this level the "nature of a problem" (ibid.) can be indicated along with detailed elaborations on how the selection was carried out. As LIST as such indicates, the potentially needed level of writing skill is most likely level A1, allowing for the writing of "simple isolated phrases and sentences" (ibid.: 77). It is however possible that lists consist of single sentences which would mean a required level of writing skills at A2 (ibid.). For the six tasks given, the majority required reading skills were placed with level A2, leading to the calculated average for this skill at the same level (see Table 8.62, 1a). True to the initial assumption on all possibly needed levels of speaking skills, the tasks' assigned CEFR levels varied between A1 and A2, whereas the latter was used more often than the former. Accordingly, the average level arrived at level A2 for spoken production (1b). The same applied to written production. The two presumed CEFR levels A1 and A2 were the only two used to evaluate the demand for *written reports and essays.* With the majority of assessments again at level A2 the average level across all six tasks was also A2 (1c). Consequently, with level A2 being assigned most often, the maximum relative frequency was at that level as well (2).

LIST	Skill			
	Reading (1a)	Speaking (1b)	Writing (1c)	max. relative frequency (2)
	A2	A2	A2	A2

Table 8.62 - CEFR levels for LIST per skill and max. relative frequency across all skills, Discover Biology 1, grades 9 and 10

The second discourse function concerned with mainly written production, RECORD*,* required reading skills starting from level A2 as well. The level could increase to level B2 in cases involving complex descriptions of test procedures qualifying as "highly specialized sources" (Council of Europe 2018: 63). Elaborations on the notes taken following the instructions as per the above-given definition necessitate speaking skills starting at level A2 for "simple descriptions […] using basic expressions along with sequential connectors […]" (ibid.: 71), or the description of "[…] habits and routines, past activities […]" (ibid.: 70). Cases involving a more complex procedure to be described, demand on speaking skills could go up to level B2 expecting speakers to be able to "[…] give a clear, detailed description of how to carry out a procedure" (ibid.: 71). As implied by the can-do-statements given in the *CEFR Companion Volume with New Descriptors,* the minimal level of writing skills to be displayed would be B1 at which the composition of "[…] very brief reports to a standard conventionalized format […]" is mentioned as a competence. However, the given context may also allow for *written reports and arguments* at level A2 with a "series of linked sentences" (ibid.: 77). The analyses of the two tasks featuring RECORD were on par with the upfront assessment above. The necessary reading to obtain the information from the given texts for one task was placed with level B1 skills (see Table 8.63, 1a). The other task had RECORD as a third discourse function and hence, the required level of reading was assessed for the first discourse function, DESCRIBE (cf. Mathews & Olmesdahl 2010: 67, task 1). Next, the oral description of the recording was not required at all as both tasks entailed another discourse function to cover the spoken part (1b). For the written part, the actual mandate of the discourse function,

skills at level A2 were necessary leading to a calculated average of A2 (1c). An assessment of the discourse function across all three skills arrived at level A2 (2).

	Skill			
RECORD	Reading (1a)	Speaking (1b)	Writing (1c)	max. relative frequency (2)
	B1	-	A2	A2

Table 8.63 - CEFR levels for RECORD per skill and max. relative frequency across all skills, Discover Biology 1, grades 9 and 10

Two tasks in the selection for grade 9 also asked to PRESENT posters on either healthy food or dental hygiene (cf. Mathews & Olmesdahl 2010: 63, 4; 77, 5). Both featured tasks also employed another discourse function. Accordingly, PRESENT was to elaborate on the previous discourse functions' results. The reading part connected to each task was assessed for the other discourse functions and not necessary in the context of PRESENT. The actual presentation demanded levels A2 and B1 with addressing audiences, primarily because in both cases, posters served as visual support that would also help with follow-up questions (cf. Council of Europe 2018: 74). Consequently, the average level necessary here was A2+ (see Table 8.64, 1b). The composition of a poster featuring not only text but also visual aids was assessed to require level A2 writing skills in both cases (1c). Accordingly, the CEFR level assigned most often both absolute and in relation to the total number of assessments was A2 (2).

	Skill			
PRESENT	Reading (1a)	Speaking (1b)	Writing (1c)	max. relative frequency (2)
	-	A2+	A2	A2

Table 8.64 - CEFR levels for PRESENT per skill and max. relative frequency across all skills, Discover Biology 1, grades 9 and 10

Remaining with the composition of visuals, the selected material featured three tasks asking students to DRAW. This activity asks students to "create a picture, diagram, or model of observable or given structures using an appropriate level of precision" (KMK 2013). The demanded level of reading for information and argument was at B1 in all three instances (see Table 8.65, 1a). For speaking however, the level for *sustained monologue: giving information* was A2 (1b) since a short description of the product along with the necessary steps sufficed (cf. Mathews & Olmesdahl 2010: 69, 1a; 77, 1; 78, 1). With the three tasks asking for either a bar chart, a schematic drawing, or a flow chart, very little writing was necessary. Additional elaborations and ascriptions would have to be in isolated terms. Those were assessed to be on par with the competences described at level A1 with *overall written production* (1c). DRAW, incorporated in the specific use cases, required CEFR level of A2 most often in relation to the total number of assessments (2).

	Skill			
DRAW	Reading (1a)	Speaking (1b)	Writing (1c)	max. relative frequency (2)
	B1	A2	A1	A2

Table 8.65 - CEFR levels for DRAW per skill and max. relative frequency across all skills, Discover Biology 1, grades 9 and 10

All discourse functions placed with AFB 2 are to some extent interrelated as they seem to build on one another cognitively. The discourse function DESCRIBE, used nine times among the selected tasks, aims at simply giving details on the matter. The demanded level of reading skills varied between A1 and B1 throughout the tasks arriving at an average level of A2+ (see Table 8.66, 1a). The essential part of the discourse function, the actual verbal description, required speaking skills for *overall spoken production*, *sustained monologue: giving information,* or *describing experience* between levels A2 and B1. With the distribution of all assessments, the average level for this category was A2+ (1b). Putting the description in writing asked for a "series of simple sentences" (Council of Europe 2018: 77) in all cases and thus, the average level was A2 for this category (1c). With the obtained individual results per skill, the majority of assessments, namely 61%, were at level A2 leading to this level being the maximum relative frequency (2).

DESCRIBE	Skill			
	Reading (1a)	Speaking (1b)	Writing (1c)	max. relative frequency (2)
	A2+	A2+	A2	A2

Table 8.66 - CEFR levels for DESCRIBE per skill and max. relative frequency across all skills, Discover Biology 1, grades 9 and 10

Taking the cognitive aspect beyond a mere description of things leads to APPLY, the activity "relat[ing] obtained (biological) knowledge to a new problem/question at hand" (KMK 2013b). The three tasks working with APPLY required reading skills between levels A2 and B1. The competences for the calculated average CEFR level corresponded with the description of A2+ (see Table 8.67, 1a), i.e., the ability to identify and extract the main information in shorter texts of a field of interest (cf. Council of Europe 2018: 63). The same range of CEFR levels applied for both speaking and writing as well. The frequency and distribution of the distinct levels for spoken production led to an average level of A2+ (1b). The same result was calculated for written production (1c). Consequently, with all three skills being evaluated as demanding skills at level A2+, the majority of levels assigned in relation to all assessments for APPLY was A2 (2).

APPLY	Skill			
	Reading (1a)	Speaking (1b)	Writing (1c)	max. relative frequency (2)
	A2+	A2+	A2+	A2

Table 8.67 - CEFR levels for APPLY per skill and max. relative frequency across all skills, Discover Biology 1, grades 9 and 10

Having deconstructed a subject matter to be able to APPLY prior knowledge also fosters the activity of comparison. To COMPARE as a discourse function asks for the determination of similarities between two or more aspects. Three tasks featured this discourse function. The necessary reading demanded reading skills between levels A2 and B1, which resulted in an average level for this skill across the three tasks at A2+ (see Table 8.68, 1a). The same distribution was determined for the spoken part of those tasks with the same average CEFR level for this skill (1b). Likewise, the recording of the results of each comparison in writing required *written reports and essays* to resemble the characteristics

allocated with levels A2 and B1 with an average demand of A2+ (1c). Finally, considering all three skills, the CEFR level for the maximum relative frequency was A2 (2).

	Skill			
COMPARE	Reading (1a)	Speaking (1b)	Writing (1c)	max. relative frequency (2)
	A2+	A2+	A2+	A2

Table 8.68 - CEFR levels for COMPARE per skill and max. relative frequency across all skills, Discover Biology 1, grades 9 and 10

Another discourse function that benefits from the initial two cognitively is EXPLAIN. Here, students need to "give information about an actual situation, contextualize regarding rules, principles and causes to help to understand it better" (KMK 2013b). EXPLAIN was again the discourse function most frequently used. The 17 tasks featuring EXPLAIN varied between levels A2 and B1 regarding the necessary level of reading skills. With twice as many assessments with level B1, the average required level for *reading for information and argument* arrived at level B1 (see Table 8.69, 1a). Four of five spoken production subskills were needed to assess the required level for the selected tasks. Regardless of the subskill, all but one tasks were at level B1, resulting in the corresponding average level for this skill at B1 (1b). The written part then mainly required skills corresponding to the description of level A2 competences for written reports and essays. Only two tasks were assigned level B1; hence, the average for written skills arrived at A2+ (1c). EXPLAIN as a discourse function had the majority of all required language skills placed at level B1, considering all three categories of skills (2).

	Skill			
EXPLAIN	Reading (1a)	Speaking (1b)	Writing (1c)	max. relative frequency (2)
	B1	B1	A2+	B1

Table 8.69 - CEFR levels for EXPLAIN per skill and max. relative frequency across all skills, Discover Biology 1, grades 9 and 10

The last of the five discourse functions of AFB 2 is an additional one that was not previously analyzed. DERIVE asks students to "come to an appropriate conclusion based on knowledge and findings" (KMK 2013b). While cognitively, this discourse function builds on the activities carried out with the previous discourse functions, linguistically, however, each discourse function is independent. Generally, DERIVE could require reading skills starting from level A2, at which main information can be picked out, especially in cases in which visual aids are provided (cf. Council of Europe 2018: 63). With increasing specialization of the sources from which the information has to be obtained, the required level of reading skill could increase to level B2. The same range of skills, from A2 to B2, could be necessary for spoken production. With A2 being the minimal requirement level, B2 is likewise possible with complex and specialized cases that call for an increased level of detail (ibid.: 69f.). *Written reports and essays* would have to start at a level A2, at which students "can give their impressions and opinions in writing about topics of personal interest […] using basic everyday vocabulary and expressions" (ibid.: 77). Although rather unlikely in the given context of the higher classes, the demand on written products could increase to level C1, which is connected to writing about complex issues (ibid.). The selected

material contained a single task employing DERIVE, asking students to derive "the role each of these [foods, discussed at length throughout the chapter] should play in a balanced diet" (Mathews & Olmesdahl 2010: 69, 2b). To gather the information necessary to carry out the task, students had to read several sections of the chapter on foods and nutrition, corresponding to reading skills for information and argument at level B1 (1a). To elaborate on their thoughts and apply prior knowledge about nutrition, *sustained monologue: putting a case* was evaluated to require level B1 as well (1b). Putting the derived combination of foods into writing demanded writing skills resembling the qualities described at level A2 with *written reports and essays*, i.e., a sequence of sentences with connectors sufficed (cf. Council of Europe 2018: 77). With a single task employing the discourse function, a maximum relative frequency was not computed (2). The remaining three discourse functions, DISCUSS, JUSTIFY, and HYPOTHESIZE are allocated with AFB 3. Seemingly reflecting on the AFB, linguistic demand was at level B1 for all three categories of skills with each discourse function, except with JUSTIFY, which required level A2 writing skills (1c). The maximum relative frequency with each of the three discourse functions, however, was still with level B1 (2).

The distribution of CEFR levels per skill and discourse functions is illustrated in Figure 8.14 , including the red-colored zone marking the level of language as per the curricular standards (cf. KMK 2003).
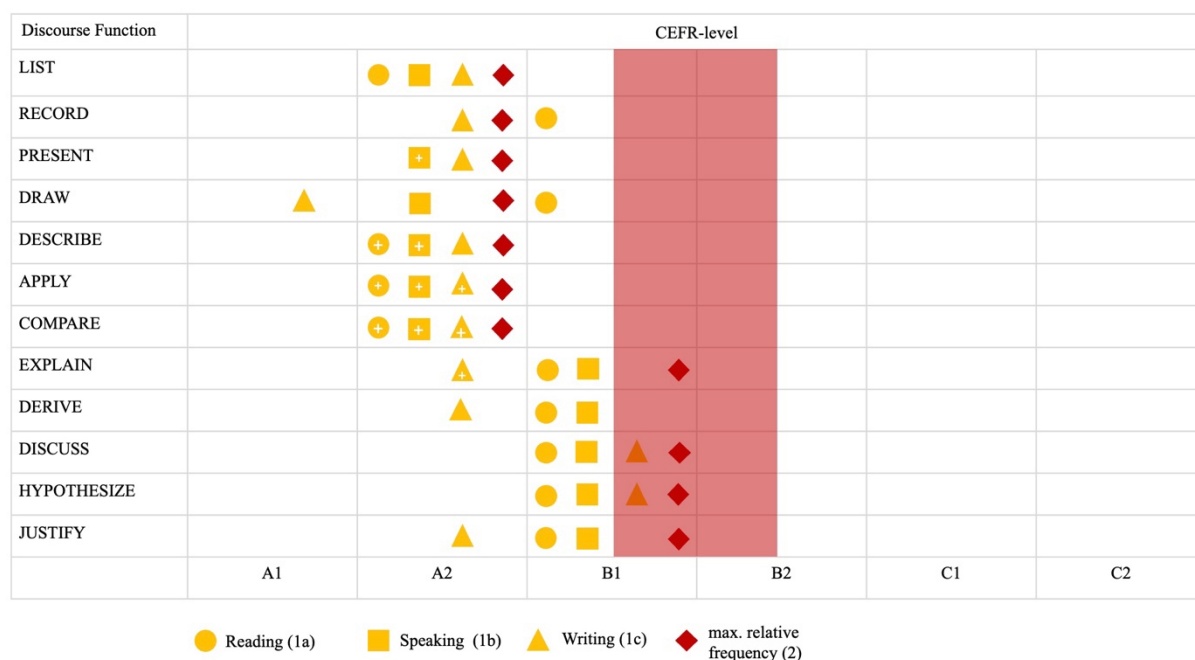
Figure 8.14 - Average CEFR levels per discourse function and skill & max. relative frequency per discourse function across skills, Discover Biology Volume 1, grades 9 & 10

The individual skill categories averaged out at level B1 for reading, A2+ for speaking and writing (see Figure 8.15, 3) for all three skills across the 57 tasks and use cases. The maximum relative frequency was at level B1, with 51% over 46% at level A2 (4).

Discourse Function	CEFR-level					
Reading						
Speaking						
Writing						
Discover Biology Volume 1, Gr. 9 & 10						
	A1	A2	B1	B2	C1	C2

avg across all discourse functions (3)

max. relative frequency across all skills and discourse functions (4)
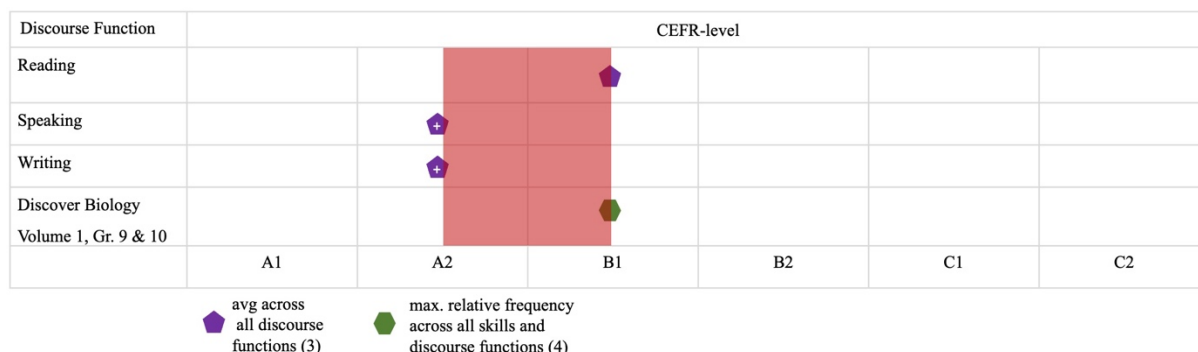
Figure 8.15 - Average CEFR levels per skill across discourse function and overall maximum relative frequency, Discover Biology Volume 1, grades 9& 10

8.3.4 Mathematics

As mentioned in chapter 7, Mathematics material has a different character; hence, tasks and in-class discourse are different as well. Following the structure applied in chapter 7.4, the analysis of Mathematics tasks and exercises also differentiated between grades 7 and 9 and the two topics of Algebra and Geometry. Before defining the various discourse functions and elaborating on expected required CEFR levels, a brief note on reading in the context of Mathematics material, precisely the material chosen for this study, is necessary. Other than the three subjects of History, Geography, and Biology, Mathematics does not base the tasks on explicatory or source texts. The tasks as such are the text to be read with, at times, additional information provided by accompanying visual support. Consequently, this affected the assessment of reading.

The KMK has compiled a list of mathematical discourse functions frequently used in Mathematics classes for German schools in foreign countries (cf. KMK 2013). This general list is specialized for *Gymnasium* and provides all discourse functions and elaborates on the activity connected to or the purpose of the discourse function. It is more extensive as the list provided by the *Institut für Qualitätsentwicklung für das Bildungswesen*, which published a base list for German standards in the A-level exams (*Abitur*) (cf. Institut zur Qualitätsentwicklung im Bildungswesen 2019). Rhineland-Palatinate's curricular standards refer to the first list, but the second seems more appropriate for this study as it is more comprehensive and reflects on all stages of Mathematics teaching and not only the *Abitur*. As a preliminary step towards mapping those discourse functions to CEFR descriptors, those German terms were mapped to English equivalents. Since a comprehensive and equivalent list could not be retrieved, neither from a German source nor from an English-speaking country, a list based on different sources serves as a base for analyzing the linguistic requirements for mathematical operating functions (see chapter 12.6.3). Neither list provided a subdivision of discourse functions into the three

AFBs (see chapter 4.2), as was the case for the other three subjects. The definitions of the discourse function and their actual frequency of use are given in Table 8.70 below:

Discourse Function	Definition (for sources please refer to gloss at the end of this table)	Frequency		
		Algebra	Geometry	Total
ANALYZE	analyze requires students to examine carefully, take apart mathematically, and break down into components or essential characteristics to identify causes, key factors, and possible results [2]	25	5	30
CALULATE	recall knowledge produced from memory and apply to determine solution [4]	32	11	43
APPLY	apply requires a student to use mathematical knowledge in a variety of situations [2]	14	4	18
VERIFY	- provide evidence while reasoning through solution. See relationships and connections. Identify conjectures and tests.[4] - [Students] apply their mathematical knowledge […] in systematically exploring and solving problems […] and give valid reason when […] selecting from possible solutions[3] - requires students to demonstrate that a mathematical concept is true or accurate[2]	1	-	1
PROVE	- students explain and verify mathematical relationships […] they use mathematical language and notation […][3] - prove requires students to demonstrate that an argument is universally true where each step and conclusion must be supported by evidence and/or reasoning. This can be shown through a variety of strategies[2]	3	2	5
EXPLAIN	- students use mathematical terminology […] to communicate mathematical ideas, and link concepts and processes within and between mathematical contexts[3] - explain requires a student to provide verbal evidence to support a conclusion or solution[2]	3	4	7
NAME	to identify, specify or mention by name[1]	-	16	16
DRAW	[…] requires a student to create a picture, diagram or model to illustrate a mathematical concept using an appropriate level of precision[2]	-	5	5
SKETCH	- requires students to communicate a mathematical problem concept through pictures, diagrams, models, symbols or algebraic expressions[2] - [by using] appropriate mathematical language […] in graphical form[3]	-	2	2
DESCRIBE	requires that students illustrate their thinking or justifications through verbal (oral or written) statements that refer to a drawing/diagram/model[2]	-	1	1

DECIDE	[students] use mathematical terminology […] and give valid reasons when comparing and selecting from possible solutions, making connections with existing knowledge and understanding[3]	-	1	1
JUSTIFY	provide evidence while reasoning through solution. See relationships and identify connections. Identify conjectures and tests[4]	-	2	2
DEMONSTRATE	- students explain and verify mathematical relationships […], and link mathematical ideas to existing knowledge […]. […] They demonstrate fluency in the solution of familiar and unfamiliar problems[3] - - students demonstrate understanding and application of the content through […] mathematically appropriate method[s] that clearly communicate[s] the steps leading to the solution or conclusion needed[2]	-	1	1

[1]Kilgo 2021
[2]New York State Education Department 2017
[3]NSW Standards Authority 2019
[4] Sonoma County Office of Education 2021

Table 8.70 - List of discourse functions, definitions used in with Mathematics corpus compiled

Tasks which did not feature one of the discourse functions listed were adjusted, i.e., a discourse function was chosen from the list resembling the tasks imperative as close as possible. Likewise, a single imperative in a task often led to a transfer of said imperative verb into several discourse functions listed, e.g., with "Construct an equilateral triangle of side 6cm." (Capewell 2002: 237, 3a), an unlisted imperative as a discourse function was given. Hence, the imperative word CONSTRUCT was first split into discourse functions APPLY, to ensure the necessary evaluation of the process of "[…] us[ing] mathematical knowledge in a variety of situations" (New York State Education Department 2017), and DRAW, since the task asked for realization of the mathematical concept of an equilateral triangle. Both discourse functions were then analyzed for their linguistic requirements. The material chosen featured a total of 13 different discourse functions (see Table 8.70) of which seven were used in the thematic context of Algebra. Out of those seven, two functions were being used exclusively for Algebra.

<u>Mathematics, Basic Algebra and Geometry, grade 7</u>

For grade 7 material, 123 discourse functions were employed, with three exclusively for Geometry. The number of tasks analyzed amounted to 132, with 68 tasks dealing with Algebra and 54 focused on Geometry. Especially two discourse functions stood out as being most frequently used in both thematic fields, ANALYZE and CALCULATE, while the remaining 11 were used less frequently. In Geometry, NAME was used comparably often as the aforementioned two.

In the preliminary considerations on possibly required CEFR levels for each skill to satisfyingly and sufficiently carry out the task featuring the first discourse function listed, ANALYZE, the requirements for reading, either *for orientation* or *information and argument*, ranged from level A2 to level B2. Within those two, students can either "find specific information […] in simpler written material" (Council of Europe 2018: 62) or can "obtain information, ideas and opinions from highly specialized sources" (ibid.: 63). When speaking about their findings or results, students' skills could be required to master abilities between level A2, at which connected sentences in sequence and simple language (cf. Council of Europe 2018: 71) sufficed, to again level B2. This level would be required when "systematically developed descriptions […] with appropriate highlighting of significant points and relevant supporting detail" (Council of Europe 2018: 69) were needed. For *overall written production*, i.e., to record their findings, students could be required to "write a series of simple phrases and sentences linked with connectors" (ibid.: 75). The demand for writing could increase to level B2 for *written reports and essays* in cases a "clear and detailed description of a complex process" (ibid.: 77) would be needed.

With those preliminary settings of the level of expected demand, the actual application of ANALYZE with the tasks chosen from the corpus for Mathematics placed the demand on the lower end of the above-described spectrum. The following task taken from *Framework Maths 7,* "[ANALYZE and] write an equation to describe each of Marcus' magic tricks*"*(Capewell 2002: 219, 2), was exception as this task refers to one of the rare explanatory texts given beforehand. Nevertheless, the chosen

material did not ask for extensive reading. Hence, based on the corpus of Mathematics material for grade 7, the CEFR level for *reading for orientation* and *reading for information and argument* varied from A1 to A2, never going beyond reading more than two sentences for the Algebra part. With the tasks concerned with Geometry, tasks at times were comprised of more sentences but never demanded a higher CEFR level than A2. As for the spoken part, i.e., *sustained monologue: giving information* presenting the findings of the analysis required speaking skills between levels A2 and B1 for both Algebra and Geometry. The remaining 13 tasks all featured a second discourse function, e.g., EXPLAIN. Speaking skills were assessed in the context of the additional discourse function. The same applied to the written part. Only two tasks featuring ANALYZE had a written part. For those two, the required CEFR level was limited to level A1 because it was either a single number or a single function. Looking at the average CEFR levels per skill, the average for both Algebra and Geometry was calculated. For reading (see Table 8.71, 1a), the average arrived at CEFR level A2. Those tasks that actually featured a spoken part arrived at an average required level of A2+ (1b). Lastly, the average level for writing was at level A1 although this average figure only included two tasks (1c). The maximum relative frequency across all three skills for both Algebra and Geometry combined arrived at CEFR level A2 (2).

	Skill			
ANALYZE	Reading (1a)	Speaking (1b)	Writing (1c)	max. relative frequency (2)
	A2	A2+	A1	A2

Table 8.71 - CEFR levels for ANALYZE per skill and max. relative frequency across all skills, Algebra and Geometry, grade 7

Not surprisingly, the tasks featured CALCULATE most frequently as this discourse function is the subject's core and essence. With a total of 40 occurrences, of which 32 were in the context of Algebra and 8 used with Geometry, preliminary considerations for this discourse function placed possibly required reading skills between CEFR levels A1 for "simple, important information" (Council of Europe 2018: 62). This could apply with tasks simply stating, e.g., an equation to solve. However, with tasks that probably entail a longer text to read asking students to retrieve information for a calculation, reading skills could increase to level B1. In such cases, *reading for orientation* could require students to "scan longer texts in order to locate desired information, and gather information from different parts of a text […] to fulfill a specific task" (ibid.). Furthermore, speaking skills between levels A1 and B2 could be needed to talk about results and the procedural steps of the calculation. While level A1 would apply to single sentences simply giving the solution as a number or figure, level B2 could apply for more complex calculations with several steps which students would have to elaborate on "giv[ing] a clear, detailed description of how to carry out a procedure" (ibid.: 71). The preliminary analysis expected written parts in connection with CALCULATE to be generally less demanding. Often, noting down equations or solutions was deemed to be sufficient. Hence, the expected skill level was limited to a range between A1 for single phrases or digits and A2 for short sentences or a short sequence of sentences, respectively. The selected tasks featuring CALCULATE supported the prior assessment for reading skills which alternated between level A1 and A2 arriving at an average CEFR level of A1+ for both Algebra and

Geometry (see Table 8.72, 1a). Determining the average demand per topic separately, the average level for Algebra arrived at level A2 while Geometry arrived at level A1 and vice versa for spoken production. For Geometry, the required skill level averaged out at level A2, with the individual level at exclusively A2. The levels allocated to Algebra varied between A1 and A2 again and averaged at level A1+. The average across both topics for spoken production arrived at level A2 (1b). Despite the preliminary assumption that the written part of calculate could be quite demanding, the selected tasks required level A1 only for both Algebra and Geometry. Consequently, the average CEFR level for this skill also arrived at level A1 (1c). With required CEFR levels for all three skills in the context of calculate raging between level A1 and A2, the maximum relative frequency among the CEFR levels for this discourse function arrived at level A1 (2).

	Skill			
CALCULATE	Reading (1a)	Speaking (1b)	Writing (1c)	max. relative frequency (2)
	A1+	A2	A1	A1

Table 8.72 - CEFR levels for CALCULATE per skill and max. relative frequency across all skills, Algebra and Geometry, grade 7

The next discourse function, APPLY, was often used in conjunction with ANALYZE. In 18 cases, tasks required to APPLY, i.e., "to use mathematical knowledge in a variety of situations" (New York State Education Department 2017) following the analysis of a given task. After taking it apart mathematically, students had to choose and APPLY the appropriate mathematical approach. APPLY as such could require levels A1 to B1 when *reading for orientation*, depending on the quantity and complexity of the text. The tasks selected from the various textbooks, as outlined in chapter 6.2.4, required reading skills between A1 and A2 only. Here, reading was limited to either "find[ing] and understand[ing] simple, important information […]" or "locat[ing] specific, predictable information in simple […] material" (Council of Europe 2018: 62). While the tasks concerned with Algebra required level A1 exclusively, most tasks on Geometry were at level A2. Accordingly, the average level of skill required for reading the selected tasks featuring apply arrived at A1+ (see Table 8.73, 1a). When applying mathematical knowledge, spoken elaborations in class appear to be inevitable. Hence, in an initial assessment independent from actual tasks, the required levels for *sustained monologue: giving information* ranged between A2 and possibly level C1. The former level of skill would be limited to "giv[ing] simple directions […] using basic expressions […] along with sequential connectors" (Council of Europe 2018: 71) when describing the proceedings of the application. However, the latter level of skill could be required in cases students have to "give instructions on carrying out a series of complex professional or academic procedures" (ibid.). The level of skill required for the selected tasks remained at the lower end of this spectrum, assessed exclusively at level A2 (1b). The written parts required level A1 in all cases (1c), although the required level could potentially go up to level B2 in cases students were asked to "write a detailed description of a complex process" (ibid.: 77). With all assigned CEFR levels combined for all three skills for APPLY, the maximum relative frequency arrived at A1 (2).

	Skill			
APPLY	Reading (1a)	Speaking (1b)	Writing (1c)	max. relative frequency (2)
	A1+	A2	A1	A1

Table 8.73 - CEFR levels for APPLY per skill and max. relative frequency across all skills, Algebra and Geometry, grade 7

The discourse function DECIDE seems to be closely connected to APPLY as it asks students to "[…] use mathematical terminology […] and give valid reasons when comparing and selecting from possible solutions, making connections with existing knowledge and understanding." (NSW Standards Authority 2019). Accordingly, DECIDE could follow any activity that involves activities tied to APPLY. To decide on possible solutions, students would have to display reading skills between levels A2 and B2. However, the high level of its specificity potentially calling for B2 appeared unlikely with grade 7 material. The elaborations on a possible decision could require *overall spoken production* as well as *sustained monologue: putting a case* at least level A2, ensuring students "can give a simple description of […] daily routines […] as a short series of simple sentences […]" (ibid.: 69), or "can present opinion in simple terms […]" (ibid.: 72). Those requirements could increase to level C1 with complex issues asking for precision and a high level of detail. Putting decisions into writing would need a minimum level of skill at A2 with a "[…] series of simple phrases and sentences linked with simple connectors" (ibid.: 75). The level of skill could, however, go up to C1 with complex subjects that call for "[…] clear, well-structured expositions […]" (ibid.: 77). Among the selected tasks, only one task concerned with a geometric problem featured DECIDE. It asked students to "DECIDE whether the angles shown on the diagram are alternate, corresponding or allied."(CGP 2014: 159, 2a). With no more text than solely the task as such, the required level of reading skill was A1 (1a). Naming and elaborating on the decision concerning the angles illustrated, sustained monologue: putting a case sufficed at level A2 by simply giving the kind of angle and giving short reasons for the decision (1b). The given task required only short written answers asking students to simply note the quality of each angle with its technical term. Hence, the required level of skill for overall written production arrived at A1 (1c). The maximum relative frequency was not detected.

The list of discourse functions for Mathematics features four seemingly similar functions. All four, JUSTIFY, VERIFY, PROVE, and DEMONSTRATE are concerned with applying mathematical knowledge to dissect a problem and giving reasons for the chosen approach or as evidence for or against a mathematical rule or universal truth. Each discourse function in the order given here adds to the next, i.e., JUSTIFY is mainly concerned with giving reasons as evidence (cf. Sonoma County Office of Education 2021). VERIFY builds on the skill of giving reasons to further the activity to informedly select an approach to a specific problem. The next step is the activity entailed in PROVE, which does not exclusively focus on verifying distinct approaches but transfers that knowledge to the universal level. Lastly, DEMONSTRATE uses this competence to transfer said knowledge to unfamiliar problems and aspects. This scaling, however, describes the cognitive development of those four discourse functions. The following examines whether the linguistic demand reflects this.

JUSTIFY is defined as an activity that aims at "provid[ing] evidence while reasoning through solution" (ibid.). In general, reading skills should at least be at level A2 to ensure that the relevant information can be retrieved from text (cf. Council of Europe 2018: 63). With increasing complexity and quality of the source, e.g., from school textbooks to academic texts, the required reading skill could increase to level B2. The level of speaking should be at A2 as a minimum and could likewise increase to level B1 with increased precision and level of detail required in elaborations (ibid.: 71f.). Lastly, *written reports and essays* would require at least level B1 to "[…] summarize, report and give an opinion about accumulated factual information." (ibid.: 77), that is, if more than an equation or term is needed. Again, with increasing complexity for written elaborations, the level could increase to B2. Two tasks in the corpus featured the discourse function of JUSTIFY focusing on Geometry. Other than initially assessed, the actual tasks both required level A1 for *reading for information and argument* (see Table 8.74, 1a) while spoken production *sustained monologue: putting a case* was in line with the initial considerations at level B1 for both tasks (1b). Both tasks also asked for more in writing than an equation. One task would be sufficiently fulfilled with *written reports and essays* at level A2. The second task, however, required level B1. Accordingly, the average level for writing was A2+ (1c). The maximum relative frequency was at level B1 with three out of six assessments being at that level (2).

	Skill			
JUSTIFY	Reading (1a)	Speaking (1b)	Writing (1c)	max. relative frequency (2)
	A1	B1	A2+	B1

Table 8.74 - CEFR levels for JUSTIFY per skill and max. relative frequency across all skills, Algebra and Geometry, grade 7

As per its definition, VERIFY appears more demanding both cognitively as well as linguistically as it requires students to "apply their mathematical knowledge […] in systematically exploring and solving problems […] and give valid reason when […] selecting from possible solutions" (NSW Standards Authority 2019). Since the linguistic aspect is in focus, the CEFR levels deemed probable for *reading for information and argument* ranged from A2 to C1. With the low grade chosen for this analysis, the high level of demand C1 entails is very unlikely as students would be asked to "[…] understand in detail […] complex text likely to be encountered in […] academic life, identifying finer points of detail […]" (Council of Europe 2018: 63). The likelihood of level A2 being demanded requiring students to "understand short factual descriptions […]" (ibid.) was much greater with material for a grade 7. Students verifying and giving reasons for their judgment in their own words came with an increased demand for the spoken part of the discourse function. Accordingly, CEFR levels started at B1 with *sustained monologue: putting a case* asking students to "[…] briefly give reasons for opinions, plans, and actions" (ibid.: 72). Level C1 and its skill to "[…] develop and argument systematically […] highlighting significant points with supporting examples […]" (ibid.) in more complex calculations or mathematical processes was also considered possible. Lastly, the CEFR levels potentially needed with VERIFY and a written product could range from A1, or A2 respectively, to C1. The lower levels most likely apply to tasks sufficiently answered by an equation or mathematical term with a short

accompanying sentence. In comparison, the higher levels are most likely reserved for complex academic mathematical aspects and, thus, reasonably unlikely in the context of grade 7 Mathematics. Accordingly, the single task concerned with the discourse function VERIFY, "Solve each of the equations you made in question 2 and VERIFY your answers to make sure they are correct." (Capewell 2002: 219, 6), required reading skills at level A2 for both reading the actual task and the text provided beforehand (1a). Elaborating on findings and justifying choices and actions was a little more challenging for both the oral production and vocabulary needed and thus, required level B1 for the spoken part *sustained monologue: putting a case* (1b). Finally, this task was sufficiently accomplished with writing down an equation and its solution for a variable leading to the *overall written production* being assessed at level A1 (1c). With a single task and three assessments, the maximum relative frequency does not provide sufficiently relevant results and hence, was not determined. Likewise, the results are not individually presented in a table at this point but are visualized in Figure 8.16.

Having analyzed the task that featured VERIFY, the discourse function PROVE appears to be a synonym judging from the vocabulary and its general meaning alone. Essentially, VERIFY and its activities are part of the function PROVE. While VERIFY can be about a specific aspect or problem, PROVE is instead concerned with "demonstrat[ing] that an argument is universally true where each step and conclusion must be supported by evidence and/or reasoning." (New York State Education Department 2017). Rating potentially needed skills preliminarily, PROVE could require reading skills between level B1 for "straightforward, factual texts[…]" (Council of Europe 2018: 63) and level B2 in cases the text given qualifies as a "highly specialized source within the[…] field" (ibid.). Furthermore, speaking skills may have to be at level B1 to satisfyingly follow the discourse function's mandate and elaborate on the universal truth and "[…] briefly give reasons for opinions […]" (ibid.: 71). With more complex mathematical aspects or required elaborations, the demand on speaking skills could increase to level C1 in cases students would have to "[…] argue a case on a complex issue […]" (ibid.). With such potentially high demand for speaking skills, the possibly required writing skills are comparably low. Accordingly, single equations or mathematical terms could suffice for *written reports and essays* or *overall written production* corresponding to level A1. This could increase up to level C1 in cases requiring "clear, well-structured expositions of complex subjects [...]" (ibid.: 77). Again, for all three skills, the higher end of the described spectrum is rather not to be expected with grade 7 material. This assumption was confirmed by the assessment of the five tasks concerned with PROVE. The three tasks for Algebra required reading skills at level A1 with the provided short text and simple phrases. With Geometry, the needed reading skills increased slightly to level A2. For reading in total, an average skill of A1+ was calculated (see Table 8.75, 1a). With the requirement to explain and justify the findings, the demand connected with *overall spoken production* and *sustained monologue: putting a case* increased to level B1 for both Algebra and Geometry. The average was accordingly also calculated to be B1 (1b). Lastly, the written parts of those five tasks ranged from A1 to A2 and thus, confirmed the initial assessments made beforehand. On average, the required skill to fulfill the specific tasks was A1+ again

(1c). Looking at the linguistic demand imposed by PROVE across all three skills, the CEFR level used most often in the assessment and hence, arriving at the maximum relative frequency, was at level B1 (2).

	Skill			
PROVE	Reading (1a)	Speaking (1b)	Writing (1c)	max. relative frequency (2)
	A1+	B1	A1+	B1

Table 8.75 - CEFR levels for PROVE per skill and max. relative frequency across all skills, Algebra and Geometry, grade 7

A third discourse function requiring similar processes like PROVE and VERIFY is DEMONSTRATE. Again, the word as such and its general meaning does not provide a distinction between the three. The differences become more apparent when looking at the definition provided above (see Table 8.70). While VERIFY was concerned with distinct tasks and PROVE with the mandate to trace universal truths, DEMONSTRATE incorporates the activity of verification to explain mathematical relationships and tie those to prior knowledge, i.e., the aspect of transferring existing knowledge and applying it to the given problem plays a more prominent role with DEMONSTRATE. The task of explaining something calls for a deeper understanding of details when reading text. Accordingly, the skill possibly required when reading in the context of this discourse function would have to be at least level B1, allowing for "[…] understand[ing] most factual information [in] straightforward factual texts […]" (Council of Europe 2018: 63). This demand could increase to level C1 the more challenging and complex the texts get. Being tasked to explain and apply mathematical (prior-)knowledge to "unfamiliar problems" (NSW Standards Authority 2019) demands an increased skill when it comes to spoken production. Students would have to be able to "[…] explain the main points in an idea or problem with reasonable precision, […] to indicate the nature of a problem, [and] to describe how to do something" (Council of Europe 2018: 71) with an increased level of detail. These competences would be given with level B1. If confronted with increasingly complex procedures, the required level of skill could increase to level C1 (ibid.). In cases that require a written answer composed of more than equations or terms, the required level of skill again would minimally have to be at level B1 with a possible increase to level C1. The selected tasks featured only a single application for DEMONSTRATE. Despite the initial assessment of potentially needed levels of skill, *reading for information and argument* for the specific task was at level A1 only (1a). True to the upfront considerations on possible CEFR levels, the spoken part of the task required level B1 to ensure a sufficient level of detail and precision (1b). The same applied to the written part (1c). The fact that only a single task asked students to fulfill the activities connected to demonstrate, the maximum relative frequency (2) was not computed. The results for categories 1a to 1c are summarized in Figure 8.16.

Although the discourse function EXPLAIN plays also a part in other discourse functions such as PROVE, students can also be "require[d] […] to provide verbal evidence to support a conclusion or solution" (New York State Education Department 2017) "[…] us[ing] mathematical terminology […] to communicate mathematical ideas, and link concepts and processes within and between mathematical

contexts" (NSW Standards Authority 2019) independent of further aspects. EXPLAIN entails the description of possible mathematical approaches and iterations using technical terms and subject-related language. Accordingly, skills for *reading for information and argument* would at least have to be at level A2, potentially increasing up to level B2 in cases that involve "highly specialized sources" (Council of Europe 2018: 63). As the can-do-statements of the CEFR Companion Volume mention the verb "explain" for the first time with *sustained monologue: giving information* at level B1, this would be the lower end of the spectrum of skills potentially needed for this discourse function when it comes to oral production. This level could go up to C1 in cases asking students to "give instructions on carrying out a series of complex professional or academic procedures" (ibid.: 71). The potential demand for written production then decreases to level A2 as the lowest level. Since EXPLAIN may require a little more than just an equation or mathematical term, level A1 was not deemed sufficient. Again, in cases featuring complex procedures, writing skills demand could increase up to level C1. The application of EXPLAIN in the selected material concerned seven tasks, three dealing with algebraic and four with geometric problems. *Reading for information and argument* was required at level A2 for both topics. All tasks were comparable to task 6 of *Mathematics for Key Stage Three*: "Investigate regular polygons with 3 to 10 sides. Which ones tessellate? Use what you know about angles to EXPLAIN why." (CGP 2014: 163, 6). Accordingly, the average level of skill for both areas arrived at level A2 (see Table 8.76, 1a). As per the given can-do-statements, the required level of skill for all uses of EXPLAIN for *sustained monologue: giving information* or *putting a case* was level B1. Correspondingly, the average was at level B1 (1b). Although stated above that EXPLAIN potentially needed a little more than an equation or mathematical term with written production, the selected tasks showed that this could indeed be sufficient. For five out of seven tasks, an equation was sufficient to explain the mathematical problem or aspect in writing. For the remaining two tasks, except for one, all assessments were at A1, leading to an average demand of level of skill to arrive at A1+ (1c). Out of all assessments, 46% were at level B1which then also qualified as the maximum relative frequency (2).

	Skill			
EXPLAIN	Reading (1a)	Speaking (1b)	Writing (1c)	max. relative frequency (2)
	A2	B1	A1+	B1

Table 8.76 - CEFR levels for EXPLAIN per skill and max. relative frequency across all skills, Algebra and Geometry, grade 7

The discourse function NAME was used exclusively with tasks for Geometry. In total, 16 tasks were concerned with "identify[ing], specify[ing] or mention[ing] by name" (Kilgo 2021). The definition already implies that the linguistic requirements connected to NAME are comparably low. Hence, reading skills to "[…] find specific information" or "[…] find and understand relevant information" (Council of Europe 2018: 62) could potentially range from level A2 to B1. Simply naming single words or aspects could already be covered with a single "phrase or formulaic expression" (ibid.: 71) making level A1 sufficient. However, required skills for *sustained monologue: giving information* could increase to level B1 in cases requiring report on "straightforward factual information on a familiar topic"

(ibid.). The written proportion connected to NAME and the potentially required level of skill could start again at level A1, in cases single terms or expressions are sufficient, and may increase to level A2. Looking at the 16 tasks featuring NAME, all tasks were comparable to "Measure and NAME the size of each angle." (Baker 2008: 155, 3) and thus required level A1 for all three skills. Accordingly, the average values for reading (see Table 8.77, 1a), speaking (1b), writing (1c). Accordingly, the maximum relative frequency across all three skills (2) arrived at level A1.

NAME	Skill			
	Reading (1a)	Speaking (1b)	Writing (1c)	max. relative frequency (2)
	A1	A1	A1	A1

Table 8.77 - CEFR levels for NAME per skill and max. relative frequency across all skills, Algebra and Geometry, grade 7

The same levels applied for another discourse function exclusively used with Geometry (see Table 8.78). The five applications of DRAW, which "requires a student to create a picture, diagram, or model to illustrate a mathematical concept." (New York State Education Department 2017), all remained in the lowest level of potential requirements (1a-c, 2). The initial assessment independent of actual application, however, expected reading skills required to potentially start at level A2 meaning students would be able to "[…] identify specific information in simpler written material […]" (Council of Europe 2018: 63), potentially increasing up to level B2 in cases that involve "highly specialized sources" (ibid.). The preliminary consideration of potentially needed skill levels for oral production corresponded to the above-described findings. The up-front assessment started at level A1 for "simple descriptions" (ibid.: 71) with the help of visual support and went up to level B2 to cover cases that could ask students to describe in detail how to draw a complex model (ibid.). The overall written production was also estimated to be sufficient at level A1. The actual findings were as follows:

DRAW	Skill			
	Reading (1a)	Speaking (1b)	Writing (1c)	max. relative frequency (2)
	A1+	A1	A1	A1

Table 8.78 - CEFR levels for DRAW per skill and max. relative frequency across all skills, Algebra and Geometry, grade 7

Comparable to VERIFY and PROVE, the two discourse functions, DRAW and SKETCH, seem synonymous. SKETCH, however, is not only concerned with the creation of a picture but "requires students to communicate a mathematical problem [or] concept through pictures, diagrams, models, symbols or algebraic expressions" (New York State Education Department 2017) by using "appropriate mathematical language […] in graphical form" (NSW Standards Authority 2019). While DRAW entails the illustration of a given, prescribed aspect, SKETCH is rather used to help develop a solution for a mathematical problem using visualization, i.e., the image is not based on specifications given but has to be thought of by students themselves. Students would have to show increased reading skills beyond level A1 to do so. Accordingly, the preliminary assessment of potential levels needed started at level A2 for *reading for information and argument* as students would have to "identify specific information in simpler written material" (Council of Europe 2018: 62) to come up with a visualization of the described

problem. The required reading skills could increase up to level B2 in cases that incorporate "highly specialized sources" (ibid.). To explain and talk about the choices and graphical realization, a minimum of level A2 for *sustained monologue: giving information* or *putting a case* would be required. As with reading, speaking requirements could increase to level B2 if "clear, detailed descriptions on how to carry out a procedure" (ibid.: 71) were required. As complex as reading and speaking potentially could be in connection to SKETCH, writing was estimated to remain at level A1 only. The two tasks featuring SKETCH deviated from the initial assessment, with reading parts for both requiring level A1 (1a). Speaking was assessed at level A2 for both tasks (1b), while the written part was as expected at level A1 (1c). The maximum relative frequency for the two applications of SKETCH was determined at level A1(2).

	Skill			
SKETCH	Reading (1a)	Speaking (1b)	Writing (1c)	max. relative frequency (2)
	A1+	A1+	A1	A1

Table 8.79 - CEFR levels for SKETCH per skill and max. relative frequency across all skills, Algebra and Geometry, grade 7

The following discourse function is closely connected to the previous two. DESCRIBE is the discourse function covering elaborations, explanations, or, in general, oral production based on illustrations and the process of arriving at a description. Students are required to "illustrate their thinking or justifications through verbal (oral or written) statements that refer to a drawing/diagram/model." (New York State Education Department 2017). With this definition in mind, the possibly required reading skill level could start at level A1, especially when an illustration is the main source of information. The needed skill could, however, increase to level B2 in cases that are concerned with "highly specialized sources" (Council of Europe 2018: 63). Corresponding with simple descriptions being potentially sufficient, level A2 was assigned for *overall spoken production*. However, the skill of *sustained monologue: giving information* could impose an increased demand up to level C1 in cases requiring "giv[ing] instructions on how to carry out complex professional or academic procedures" (ibid.: 71). Writing could be covered by single phrases or equations at level A1 but could also become increasingly demanding up to level C1, at which students would have to "expand and support points of view at some length with subsidiary points […]" (ibid.: 77). The latter seems rather unlikely in the context of this part of the study. The single task featuring DESCRIBE being concerned with a geometric problem required reading skills at level A1 (1a) since it asked to describe pictured angles in diagrams using three letters (cf. Baker & Hogan 2008: 155, 3). The actual description, i.e., *sustained monologue: giving information*, also required level A1 as the description was limited to short isolated phrases including three letters for the lines enclosing the angle (1b). Accordingly, noting down said three letters as the written product resulted in level A1 (1c). The maximum relative frequency (2) was not determined with one single use case.

All discourse functions and their assessments based on the selected tasks are shown in the illustration below allowing for both an overview and a comparison.
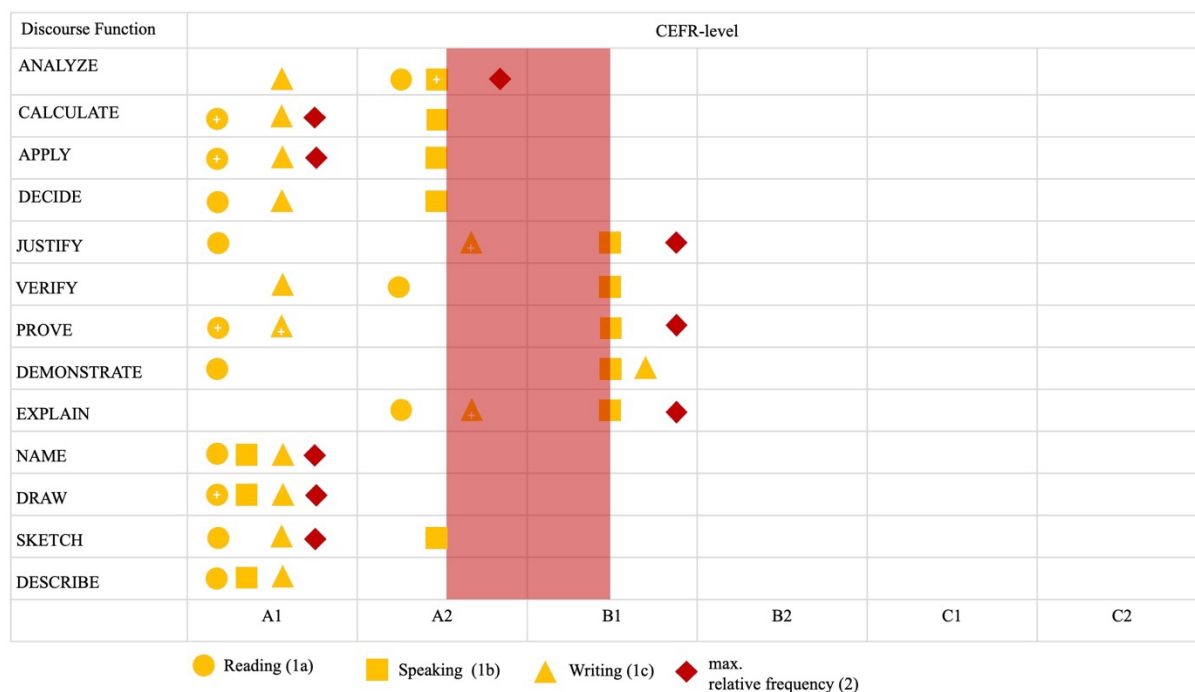
Figure 8.16 - Average CEFR levels per discourse function and skill & max. relative frequency per discourse function across skills, Mathematics corpus grade 7

Calculating the averages per skill across all discourse functions and the analyzed 132 tasks for Mathematics in grade 7, reading arrived at level A1+, speaking at level A2, and writing at A1+ (see Figure 8.17, 3). The maximum relative frequency across all discourse functions and skills with a total of 329 individual assessments arrived at level A1 (4).
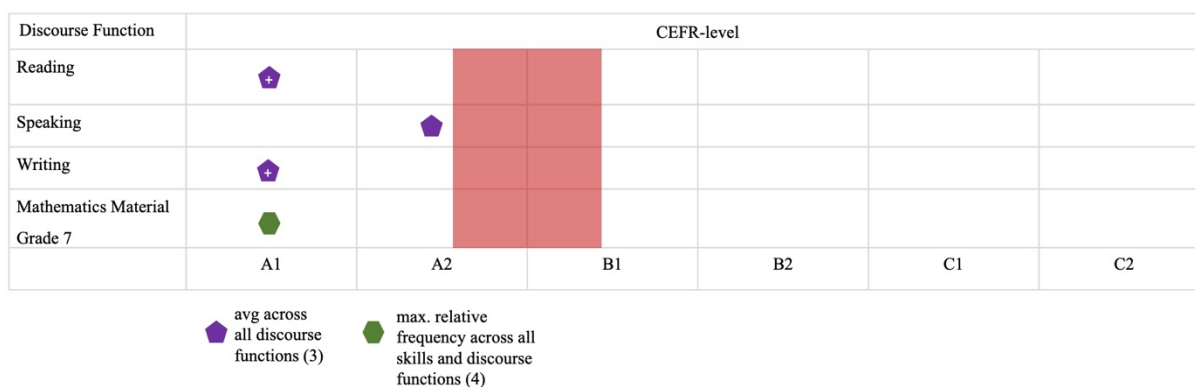
Figure 8.17 - Average CEFR levels per skill across discourse function and overall maximum relative frequency, Mathematics corpus grade 7

<u>Mathematics, Algebra 2 and Geometry, grade 9</u>

Mathematics material also included tasks targeting grade 9 according to the Rhineland-Palatinate curriculum (cf. Ministerium für Bildung, Wissenschaft, Weiterbildung und Kultur 2007), subdivided into Algebra and Geometry and featured advanced-level problems and calculations. In total, 348 discourse function uses were analyzed, of which 121 were concerned with Algebra and the remaining 227 with Geometry. Often, advanced material tasks featured more than one discourse function, while the basic material for grade 7 mostly worked with one discourse function. Another consequence of this concentration of discourse functions per task was that reading skills were not evaluated for consecutive discourse functions, especially with cases that required no increased skill in reading than those initially assessed with a preceding discourse function. For the task, e.g.,

> In ABCD, AB is parallel to DC and Db = BC, angle DBC = 70 and angle DAB = 55. CALCULATE the angle BCD and angle ABD and JUSTIFY your answer.
>
> (Parsons 2000: 42, Q15),

reading skills were assessed for CALCULATE. Since no further-developed skills were required to fulfill the JUSTIFY-part of the task, reading skills were not assessed a second time. The material for grade 9 included one new discourse function, EVALUATE, which will be defined in the course of describing the results. The other discourse functions were identical to those used in the grade 7 material. Of the 13 discourse functions, two, VERIFY and EVALUATE, were only used with tasks concerned with Geometry. Detailed results of the analysis are described in the following.

Starting again with ANALYZE, the discourse function used with 48 cases, the level of reading skills connected with *reading for orientation* or *reading for information and argument* varied between A1 and A2 for both topics. On average, reading skills (see Table 8.80, 1a) were evaluated to be at level A2 allowing for students to "[…] understand a short factual description [and] pick out the main information […]" (Council of Europe 2018: 63). Out of the 48 tasks only 20 were assessed for spoken skills. The remaining tasks had a different discourse function and reading and writing were assessed with these. The CEFR levels for the assessed tasks ranged from level A2 for "[…] simple descriptions […] as a short series of sentences […]" (ibid.: 69) to level B1 with a single task, asking for the ability to "[…] describe how to do something [or] explain the main points in an idea or problem […]" (ibid.). The average required level of skill arrived at A2 in this regard for Algebra and Geometry (1b). For written production, ten tasks were assessed for analyze. All tasks required level A1 (1c).

Accordingly, the maximum relative frequency across skills for the discourse function arrived at level A2 assigned with 81% of 78 individual assessments (2).

	Skill			
ANALYZE	Reading (1a)	Speaking (1b)	Writing (1c)	max. relative frequency (2)
	A2	A2	A1	A2

Table 8.80 - CEFR levels for ANALYZE per skill and max. relative frequency across all skills, Algebra and Geometry, grade 9

Like with grade 7 material, the most frequently used discourse function was CALCULATE. In total, 79 uses were recorded. Thirteen were in the context of an algebraic task and 66 in the context of a geometric problem. Other than with tasks using ANALYZE, most tasks using CALCULATE were assessed for every skill. Accordingly, *reading for information and argument* and *reading for orientation* ranged from levels A1 to level A2 for both Geometry and Algebra. On average, the required CEFR level arrived at level A1+ to CALCULATE with the given material (see Table 8.81, 1a). The section on oral production revealed that the level needed to fulfill the spoken part ranged from level A1 to A2 with Algebra problems. Assumingly, the mere pronunciation of an equation or term along with the solution would have been sufficient for numerous tasks. With Geometry problems, however, the levels assigned ranged from A1 to B1. Especially with cases that called for further elaborations on the applied procedure, as was the case with task 10 of *Edexcel GSCE Mathematics, Higher Course*, page 186 "Calculate the value of x. side 1 = x cm; side 2 = x + 7cm; Hypothenuse = x + 8cm." (Casson 2011: 186, 10). This specific calculation required the application of the *binomial theorem*, on which students would have to elaborate. Hence, tasks like this were assessed to be at level B1 for *sustained monologue: giving information*, according to which students can "[…] describe how to do something, giving detailed instructions" (Council of Europe 2018: 71) when discussing said task in class. The average CEFR level calculated for all subskills of speaking and both topics, Algebra and Geometry, arrived at A2 (1b). The required CEFR level for written products was evaluated to be at level A1 for all tasks and both topics (1c). The maximum relative frequency of CEFR levels assigned to all three skills and their subskills arrived at level A1 with 66% of all assessments being at that level (2).

	Skill			
CALCULATE	Reading (1a)	Speaking (1b)	Writing (1c)	max. relative frequency (2)
	A2	A2	A1	A1

Table 8.81 - CEFR levels for CALCULATE per skill and max. relative frequency across all skills, Algebra and Geometry, grade 9

Keeping the order of discourse functions set with grade 7 material, the following discourse function to be analyzed was APPLY. A total of 44 uses were recorded, of which 26 were allotted to Algebra and 18 to Geometry. When looking at required reading skills, most uses of APPLY within Algebra did not receive an assessment since another discourse function preceded APPLY, e.g., ANALYZE (cf. Baker & Hogan 2008: 107, 2). With Geometry, however, most uses received an assessment for the reading part. Here, the levels ranged from A1 to A2 based on the brief tasks given. The two tasks of Algebra, which actually were assessed, arrived at level A2. Calculating the average for

the levels assigned across both thematic fields, the required level for reading in the context of APPLY with the selected material was at A1+ (see Table 8.82, 1a). With spoken production, most tasks were between levels A2 and B1. The latter occurred exclusively with geometric problems. Here, explaining of processes or justification of decisions, without explicitly featuring the discourse function JUSTIFY, required increased speaking skills to either "[…] describe how to do something" or to "[…] briefly give reasons for and explanations for […] actions" (Council of Europe 2018: 71f.). All assigned CEFR levels in both topic areas averaged out at level A2+ (1b). Regardless the topic, written production was required at level A1 for all tasks featuring APPLY (1c). The maximum relative frequency including all three skills arrived at level A2 (2).

APPLY	Skill			
	Reading (1a)	Speaking (1b)	Writing (1c)	max. relative frequency (2)
	A1+	A2+	A1	A2

Table 8.82 - CEFR levels for APPLY per skill and max. relative frequency across all skills, Algebra and Geometry, grade 9

As was described above, the application is often accompanied with a decision on which approach to apply. Hence, the corresponding discourse function DECIDE, or moreover, the results of the analysis of tasks featuring this discourse function followed directly after APPLY. Surprisingly, though, DECIDE was featured in less tasks than APPLY. In total, 14 tasks asked students to DECIDE on something. Seven occurrences were noted for Algebra and seven for Geometry. In line with single skills not being assessed with every discourse function, six of the tasks featuring DECIDE were assessed for reading in the context of the discourse function while eight tasks had preceding discourse functions for which reading was assessed (see Table 8.83, 1a). With speaking, the subskill used most often was *sustained monologue: putting a case* with CEFR level A2 for both topic fields. Hence, the average level required for spoken production was A2 (1b). Written production was at level A1 (1c). The maximum relative frequency, putting all assessments in relation to the level assigned most often, resulted in level A2 (2).

DECIDE	Skill			
	Reading (1a)	Speaking (1b)	Writing (1c)	max. relative frequency (2)
	A2	A2	A2	A2

Table 8.83 - CEFR levels for DECIDE per skill and max. relative frequency across all skills, Algebra and Geometry, grade 9

Abovementioned and discussed similarities and differences of JUSTIFY, VERIFY, and PROVE also applied to the grade 9 material. While grade 7 material also featured a fourth discourse function resembling the former three, DEMONSTRATE was not used among the selected tasks. The first discourse function, JUSTIFY, was used in 28 cases, of which only two occurred with Algebra. Again, none of the 25 cases received an individual assessment of required reading skills as all tasks featured a prior discourse function that required students to either ANALYZE, CALCULATE, DRAW, or APPLY a theory before being asked to justify their decision (e.g., cf. Parsons 2000: 28, Q2). Accordingly, section 1a of the categories remains vacant (see Table 8.84). Giving reasons for a decision, i.e., justifying in the sense of the discourse function, requires a certain level of skill for *sustained monologue: putting a case*.

Hence, the tasks given in the selected material were assessed at level B1, the lowest possible level when asked to explain decisions. The can-do-statement within the *CEFR Companion Volume with New Descriptor*s features the discourse function "Can give simple reasons to justify […] opinions, plans and actions." (ibid.: 72). With all single tasks evaluated to be level B1, the average arrived at the same level (1b). Other than the reading part, writing skills were assessed in five cases with CEFR levels between A1 and B1. The latter applied to tasks which involved the justification of chosen theorems, such as the *Pythagoras Theorem* (cf. Casson 2001: 186, 11). On average, the required level of skill for written production across both thematic fields arrived at level A2 (1c). The average skill across two out of three skills for JUSTIFY arrived at level B1 which was mainly caused by the evaluation of spoken production (2).

	Skill			
JUSTIFY	Reading (1a)	Speaking (1b)	Writing (1c)	max. relative frequency (2)
	-	B1	A2	B1

Table 8.84 - CEFR levels for JUSTIFY per skill and max. relative frequency across all skills, Algebra and Geometry, grade 9

The next discourse function of this group, VERIFY, was used three times and only for Geometry problems. Neither reading nor writing were assessed in those three cases, only *overall spoken production* as well as *sustained monologue: putting a case* were assigned levels between A2 and B1. The maximum relative frequency was not calculated (see Table 8.85, 1b). The last discourse function, PROVE, was used with ten tasks in total. For neither the six tasks in Algebra nor the four in Geometry, reading skills were assessed. In all ten cases, *sustained monologue: putting a case* arrived at level B1 (1b). Likewise, all cases were evaluated to require writing skills at level A1 (1c). Although reading was not assessed at all, the maximum relative frequency of all 17 assessments was allocated with level B1 (2).

	Skill			
	Reading (1a)	Speaking (1b)	Writing (1c)	max. relative frequency (2)
VERIFY	-	B1	-	-
PROVE	-	B1	A1	B1

Table 8.85 - CEFR levels for VERIFY and PROVE per skill and max. relative frequency across all skills, Algebra and Geometry, grade 9

Regardless the grade, Mathematics features tasks that ask students to explain their choice of approach, a particular idea or problem, or the procedural steps leading to a solution. The selected material featured 13 uses of EXPLAIN, five concerned with Algebra and eight with a Geometry problem. Since EXPLAIN was the last discourse function in each task, reading was not assessed (see Table 8.86, 1a). For the skill of actually fulfilling the discourse function's mandate of explaining, CEFR level B1 was required for both thematic fields with *sustained monologue: giving information* and *putting a case* (1b). The assigned levels for the written part of the tasks were between A1 and A2 with both topics. With more tasks being A2, the use cases averaged out at Level A2 (1c).

With all three skills and their subskills evaluated, the maximum relative frequency was at level B1 (2).

	Skill			
EXPLAIN	Reading (1a)	Speaking (1b)	Writing (1c)	max. relative frequency (2)
	-	B1	A2	B1

Table 8.86 - CEFR levels for EXPLAIN per skill and max. relative frequency across all skills, Algebra and Geometry, grade 9

The discourse function, NAME, asks students to repeat, point to, or list information obtained from either a text or illustration. In total, 28 tasks used NAME in the selected material. The tasks were evenly distributed between Algebra and Geometry. The analysis identified levels A1 and A2 as necessary to obtain the requested information. At this level, reading skills allow for "[…] locat[ing] specific information […] and isolate the information required" (Council of Europe 2018: 62). Accordingly, most uses of the discourse function required level A2 (see Table 8.87, 1a). Although spoken production expressing single simple, isolated phrases (cf. Council of Europe 2018: 69f.) may suffice for NAME corresponding to level A1, the selected tasks also required level A2 asking for "[…] a short series of simple phrases and sentences linked into a list." (ibid.: 69). Calculating the average for both Geometry and Algebra, the CEFR level resulted in level A1+ (1b). For the written part, the requirements arrived at level A1 on average, with the necessity to note down single words and phrases, or equations for that matter (1c). All single assessments, in total 57, were analyzed for the maximum relative frequency. Level A1 (2) held the highest share for NAME for all three skills combined.

	Skill			
NAME	Reading (1a)	Speaking (1b)	Writing (1c)	max. relative frequency (2)
	A2	A1+	A1	A1

Table 8.87 - CEFR levels for NAME per skill and max. relative frequency across all skills, Algebra and Geometry, grade 9

Following the bundling of discourse functions applied above, the results for DRAW and SKETCH are discussed in one paragraph for their shared features. Four of the 30 uses of DRAW were in Algebra, and the required reading skills alternated between A1 and A2. This led to an average level of skill at A2 since a more significant proportion of tasks was assigned level A2 (see Table 8.88, 1a). As DRAW is concerned with the graphic presentation of the application of a mathematical concept, the tasks also demanded spoken production. Here, all tasks for both thematic fields needed speaking skills at level A2 since a "sequence of sentences" (Council of Europe 2018: 75) was asked for to, e.g., elaborate on the chosen approach to DRAW. Consequently, the average level of required skill also arrived at A2 (1b). Written production, on the other hand, was not required with the 30 tasks since either writing was not necessary at all or covered by another discourse function. The maximum relative frequency was at CEFR level A2 across all three skills with 77% of all assessments (2). Next, the discourse function, SKETCH, involving illustrations as a mean to develop a solution rather than depicting a pre-set matter, was used with eight tasks. Of those eight, two concerned an Algebra problem, while the remaining six were about Geometry. The required level of speaking skills was for both thematic fields at level A2 (see Table 8.88, 1a). Likewise, the level of speaking skills, i.e., *sustained monologue: giving information*, was at level

A2 for all eight cases featuring SKETCH (1b). Somewhat true to its character, this discourse function or, much rather, the eight tasks employing it, did not require a written part at all, and thus, this category remained without an assessment (1c). With the former two skills being evaluated to correspond to level A2 when looking at their requirements, the maximum relative frequency for the discourse function arrived at level A2 as well (2).

	Skill			
	Reading (1a)	Speaking (1b)	Writing (1c)	max. relative frequency (2)
DRAW	A1+	A2		A2
SKETCH	A2	A2	-	A2

Table 8.88 - CEFR levels for DRAW and SKETCH per skill and max. relative frequency across all skills, Algebra and Geometry, grade 9

Nine tasks asked students to DESCRIBE problems, issues, or matters. Neither task featured an individual reading part as tasks were composed as follows:

> Here is a sequence of shapes made using counters.
>
> A) […]
>
> B) DESCRIBE what you notice about the number of red counters in each shape […]
>
> (Baker & Hogan 2008: 111, 4).

The actual activity of describing, however, demanded speaking skills at level A2 (see Table 8.89, 1b) to be able to "[…] give a simple description […] as a short series of simple phrases and sentences linked to a list." (Council of Europe 2018: 69) . For written production, neither thematic field was evaluated for written production since, as stated above, the tasks were designed with preceding and following sub-tasks which included a written part towards the end.

	Skill			
DESCRIBE	Reading (1a)	Speaking (1b)	Writing (1c)	max. relative frequency (2)
	-	A2	-	A2

Table 8.89 - CEFR levels for DESCRIBE per skill and max. relative frequency across all skills, Algebra and Geometry, grade 9

The last discourse function was added to the collection for grade 9 tasks. EVALUATE was not used with the material selected for grade 7. To evaluate

> Students use deductive reasoning in problem-solving and in presenting arguments and formal proofs. They interpret and apply formal definitions and generalizations and apply mathematical ideas […] compare strengths and weaknesses of different strategies and solutions
>
> (NSW Standards Authority 2019).

Seemingly, this discourse function requires a deeper level of understanding when it comes to reading. Hence the initial assessment independent of specific tasks placed the level of *reading for information and argument* at B1 as a minimum, with a potential increase to B2. In the actual single task, however, the text given only demanded level A2 for both being brief and straightforward (see Table 8.90, 1a). The spoken part, i.e., *overall spoken production*, *sustained monologue: giving information* or *putting a case* would generally call for speaking skills at level B1 for straightforward descriptions, simple reasoning, and information communication with increased precision (cf. Council of Europe 2018: 69ff.). This demand could increase to levels B2 to even C1, with more complex problems of academic quality. The task of the selected material was presumed to be sufficiently fulfilled with speaking skills displaying proficiency at level B1 (1b). For written production, the initial assessment of skills levels potentially required spanned across the full spectrum of levels, from A1 to C1. Producing single words and phrases was possible, as was level C1, which could be required in highly complex written products or even text of academic quality. In the specific use case given, written production was not evaluated as a second discourse function was used: "EVALUATE*, is there a connection between the number of sides of a shape and the number of lines of symmetry it can have? JUSTIFY your answer!" (Capewell 2002: 245, 1c). Consequently, the written part was assessed for the second discourse function leaving this category empty for EVALUATE (1c). For a single task, the maximum relative frequency was not detected (2).

	Skill			
EVALUATE	Reading (1a)	Speaking (1b)	Writing (1c)	max. relative frequency (2)
	A2	B1	-	-

Table 8.90 - CEFR levels for EVALUATE per skill, Algebra and Geometry, grade 9

To sum findings up, the average values per category as outlined above are shown in a graphic overview allowing also to mark the targeted level of language skills in grade 9 using a red-colored zone.
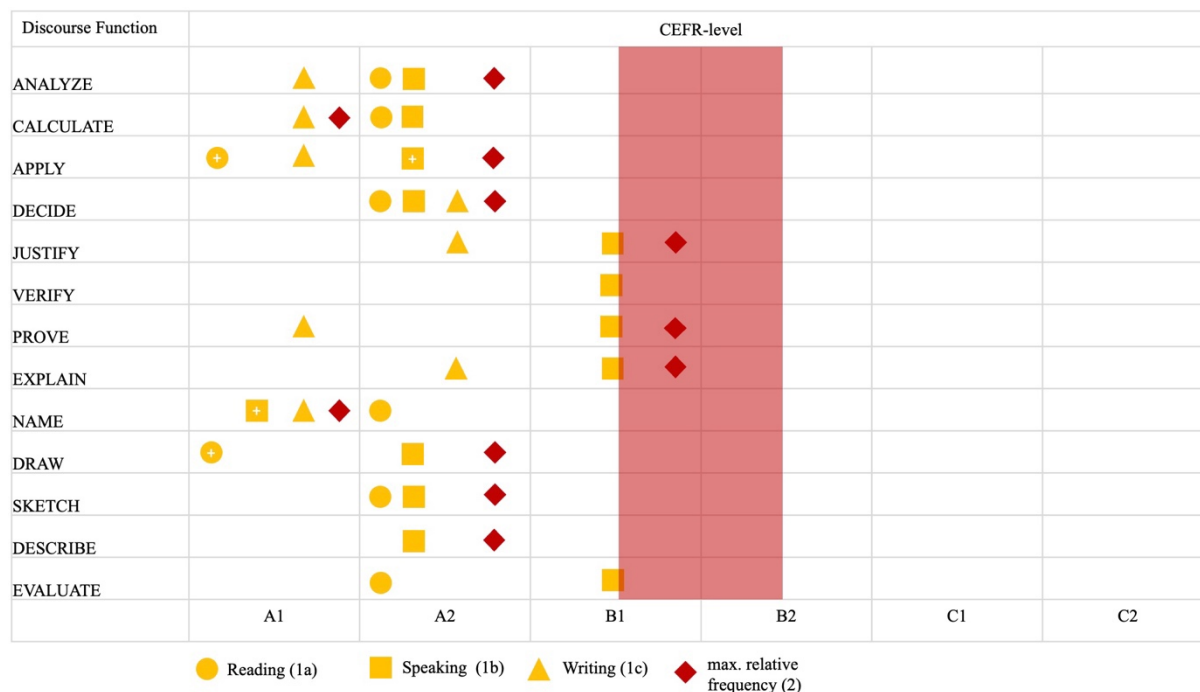
Figure 8.18 - Average CEFR levels per discourse function and skill & max. relative frequency per discourse function across skills, Mathematics corpus grade 9

Calculating the averages per skill across all discourse functions, the minimal level required for reading and speaking for both thematic fields arrived at A2 (see Figure 8.19, 3), while writing was A1+. Overall, language skills for Mathematics in grade 9 needed to be at least at level A2 (4) to be able to both sufficiently and successfully fulfill the tasks selected.
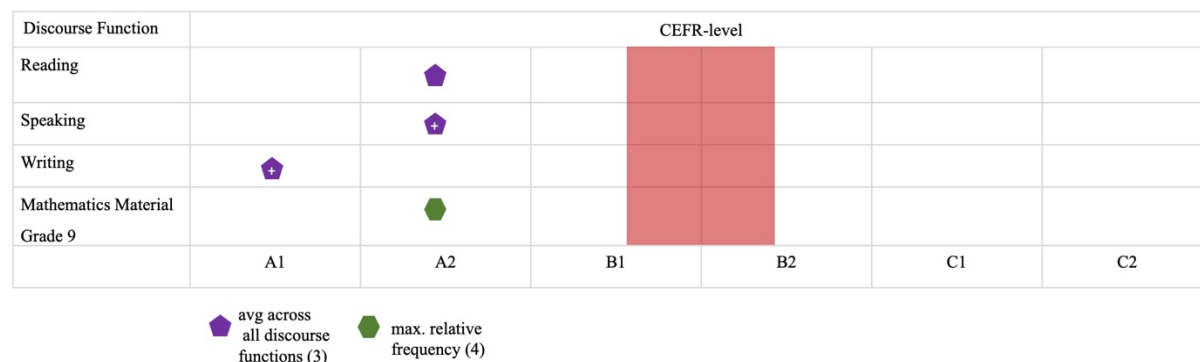
Figure 8.19 - Average CEFR levels per skill across discourse function and overall maximum relative frequency, Mathematics corpus grade 9

The implications of above findings, i.e., the adequacy of the subjects for the grades selected with regard to the requirements they impose on students, are discussed in the upcoming chapter.

8.4   Discussion of results

In coherence with the approach selected for the discussion and comparison of the results yielded from the text analysis (see chapter 7.5), the results obtained from the analysis of tasks are discussed per subject first and then in comparison to one another.

8.4.1   History

Starting with the results yielded from the material for History for the lower age group, the analysis showed that the required level of skills for most discourse functions is too demanding for students in grades 7 or 8. As per the curricular standards, the targeted CEFR level at that stage would be A2 on the verge of B1 (see Table 7.15 - Targeted CEFR levels per grade), as was indicated by the red-colored zone with the illustrations per grade and subject. The overall assessment for the tasks of *Exploring History 1*, based on the maximum relative frequency, indicated excessive demand at B1 (see Figure 8.5). Even if assuming that bilingual students display more advanced language skills than their peers without bilingual instruction, this result implies that History as a subject is too demanding at this early stage of schooling. The averages calculated per skill across all discourse functions supported this assessment. Looking at the discourse functions and their required levels of skill individually, only those allocated with AFB 1 (LIST, DESCRIBE, and SHOW) were appropriate for bilingual instruction in the lower grades (see Figure 8.4). More than half of all discourse functions were too demanding, with required levels of skill going up to level B2.

For the material selected from *Exploring History 2*, an increase in demand was identified setting on at discourse functions of AFB 1 already. Although this was expected, the extent to which the required skills increased partly went beyond the targeted skill level with that age group of 15-to-16-year-olds. Six of 11 discourse functions required skills at B2 or even B2+. This level is overtaxing, especially if the administrative directive on curricular standards places the targeted level at the threshold from B1 to B2 by the end of year 10 (cf. KMK 2003: sec. 5). Shifting the focus onto the individual skills, the calculated averages per skill would be within the targeted level (1a & b, see Figure 8.6), only with reading being too demanding at level B2. The level for which the maximum relative frequency was detected, B2, was again at the upper end of the targeted levels. Accordingly, the analyzed chapters and their tasks could again be regarded as potentially overtaxing. Additionally, the excessive demand single discourse functions imposed on students require a cautious interpretation since the calculated averages do not completely reflect those extremes. Placing those discourse functions on the CLIL Matrix (see chapter 4.3), the discourse functions deemed appropriate for the lower grades would be those of low linguistic and low cognitive demand, including FIND (OUT) and COMPARE. For the higher grades, discourse functions of AFB 2 with increasingly higher demand on cognition and language would be appropriate, while those in the upper right quadrant still imposed a too high demand.

Language proficiency requirements – tasks
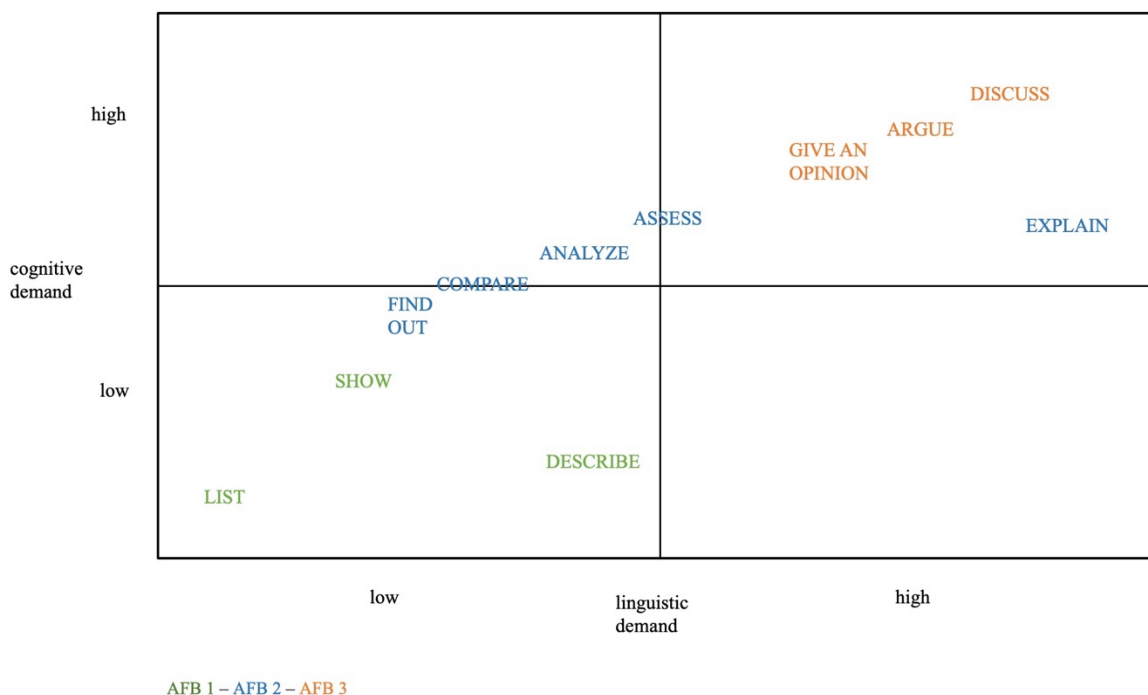
Figure 8.20 - History discourse functions on CLIL Matrix (adapted from Cummins, 1984)

The results for the analysis for History show that for the lower grades and the selected material from *Exploring History 1*, respectively, the demand is too high for students to be able to carry out the tasks. Excessive scaffolding, didactic reduction, and supporting material and information would be indispensable to teaching bilingual History at this early stage so that students have a chance to participate with success. Those supporting activities, in turn, would most likely be too demanding for teachers, especially those who did not receive bilingual or CLIL teaching training. For the higher grades, i.e., 9 and 10, the selected material still appeared too demanding but to an extent that could be managed if students were given language support. This assessment reflects the classification of language use in the History classroom as outlined in chapter 4.2. The discourse functions, especially of AFB 2 and 3, encourage free production and interaction. As described in chapter 4.1, texts in History hold a special position as the subject matter as such is only accessible through text and often employs everyday language that becomes subject-related, academic language. Accordingly, discourse in History mostly employs language of BICS-quality in an academic context (see Figure 4.1 - The four subjects in focus allocated with the types of communication, kinds of knowledge and quality of language used (BICS, CALP)). This shows in the assessment of the tasks and their discourse functions, as well. With said quality of language in History, with free production and horizontal knowledge, employing a relatively high share of BICS language, the demand is too high to even overtaxing. These qualities of language in History are very prominent with EXPLAIN of AFB 2 and with the discourse functions of AFB 3. These discourse functions are all placed with high linguistic and high cognitive demand on the CLIL Matrix, with extreme values at CEFR level C1 (see Figure 8.4, Figure 8.6 and chapter 12.8). Additionally, discourse functions in AFB 2 already had CEFR level B2 as their maximum relative frequency (see Figure 8.7), indicating that a majority of individual assessments were at that level. B2, however, is the

upper end of the targeted level for grades 9 and 10 and thus could be too demanding if not overtaxing in these grades. For students in the lower grades, this quality of language required is insurmountable.

8.4.2 Geography

Geography was the second subject for which material was analyzed. The calculated average for the material of *Diercke Geography Volume 1*, aimed at the lower age group, arrived at a level that exceeded the language skills targeted in grade 7 (see Figure 7.15). However, the majority of the averages calculated for the three skills across the discourse functions was placed within the A2-band. Likewise, the overall maximum relative frequency with all discourse functions and skills arrived at level A2 (see Figure 8.9). Seemingly, Geography is a subject appropriate for early bilingual teaching. When looking at the discourse functions individually, though, a total of five discourse functions appeared to be potentially overtaxing for students. Particular tasks or parts of tasks were exceedingly demanding at level B1. However, a significant portion of assessments were assigned to the A2-band, aligning with the skill level being targeted. Other than the material for History, the progress detected in material moving on from the first volume to the second volume resembled the assumed progress students potentially make from grade 7 to grade 9. Of the material selected for the latter age group, all discourse functions were in the targeted range of language competences (see Figure 8.8), except for the demand with writing with EVALUATE. With the targeted level of language competence by the end of grade 9 to be at CEFR level B1 beginning to slowly move onto B2, the CEFR levels for the specific skills meet the expectation set by the administrative directive for curricular standards (cf. KMK 2003: sec. 5). This was also true for the calculated averages across the discourse functions for the skills as such at the overall maximum relative frequency (see Figure 8.10). The distribution of discourse functions on the CLIL Matrix shows that those deemed overtaxing for the lower grades (DISCUSS, COMMENT ON) are placed in the quadrant for high cognitive and linguistic demand. All other discourse functions were placed with low demand or, at most, on the threshold.

For the higher grades, the discourse functions of AFB 3 with increasingly higher demands on cognition and language were evaluated to be appropriate regarding their linguistic demand, although placed in the upper right quadrant.
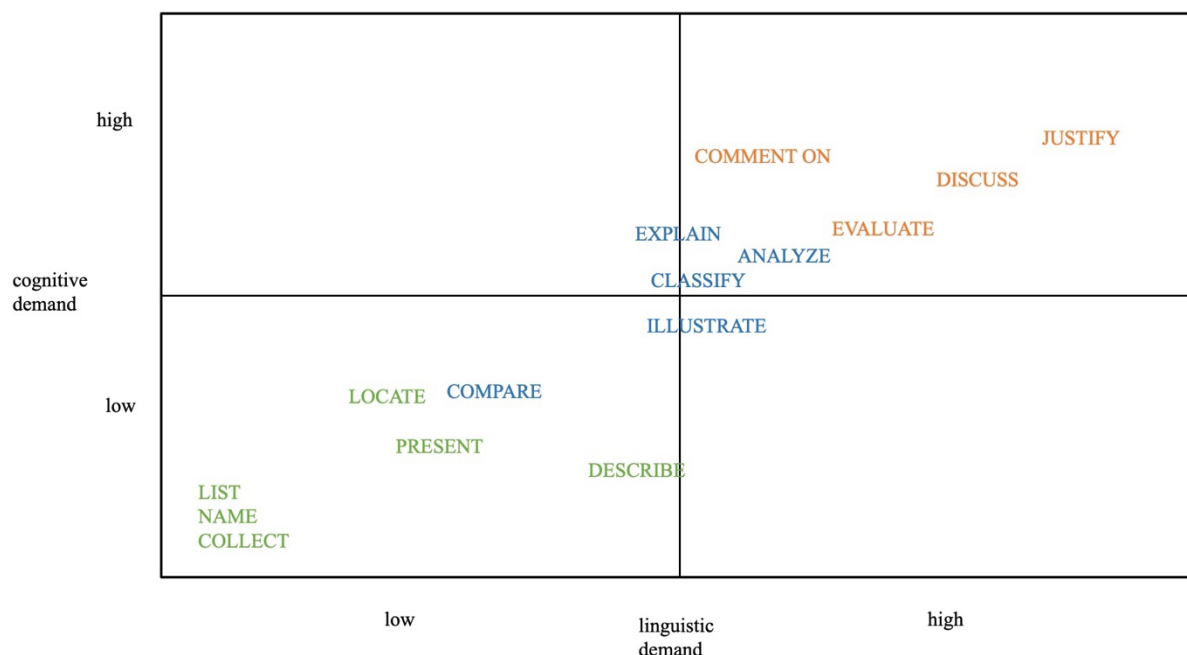
Figure 8.21 - Geography discourse functions on CLIL Matrix (adapted from Cummins, 1984)

Based on the assessment of the selected material, the tasks of Geography can be regarded as a bilingual subject appropriate to be taught at lower grades already. With respect to the three discourse functions evaluated as too demanding and single skills evaluated to be above the targeted level, a suitable starting point could be either the second half of grade 7 or grade 8. Especially with regards to Mentz's classification of Geography (see Figure 4.1 - The four subjects in focus allocated with the types of communication, kinds of knowledge and quality of language used (BICS, CALP)) as engaging both limited and controlled production as well as free production of language, a later onset of bilingual teaching could ensure that students had the opportunity to develop their language skills to an extent that allows for the demanding aspects of bilingual Geography teaching and especially the material, which was excessively demanding particularly for the lower grades.

8.4.3 Biology

With Biology classified as a subject fostering language competence within a controlled productive manner, i.e., using systematized phrasing and pre-set language chunks, and a strong focus on subject-specific language (see Figure 4.1 - The four subjects in focus allocated with the types of communication, kinds of knowledge and quality of language used (BICS, CALP)), the averages calculated for the three skills of the subject for the lower age group arrived at level A2+. The overall maximum relative frequency arrived at level B1. Both results point to Biology being entirely appropriate for the targeted

age group (see Figure 8.12 and Figure 8.13). The maximum relative frequency of level B1 already hints at the aspect that some discourse functions and their demanded level of skill could be too challenging with an early onset of CLIL Biology. A total of four discourse functions exceeded the targeted level if only by single skills but also by the average across the skills per discourse function (see Figure 8.12). At an early stage in bilingual Biology teaching, even more than just the four could be too demanding for students. Hence, starting with Biology as a bilingual subject would also be advisable later in grade 7 or in grade 8. In contrast to Geography, the character of language in Biology (see chapter 4.2), i.e., the fixed language patterns to be used, is regarded to allow for an onset of bilingual Biology teaching earlier than Geography. With the overall level of demand remaining the same for the higher age group, students would be able to catch up with the assumed progress they make from grade 7 to grade 9. Based on the evaluations of the material selected for the older age group, the average demand for bilingual Biology remained at level A2+ for speaking and writing while the demand for reading increased to level B1. The maximum relative frequency of CEFR levels across all assessments also remained at level B1 (see Figure 8.15). For all discourse functions and their individual assessments, the determined and calculated levels were within the range of the targeted level of B1 with an onset of B2 (see Figure 8.14), not even those discourse functions placed with high linguistic and high cognitive demand on the CLIL were evaluated to be too demanding.
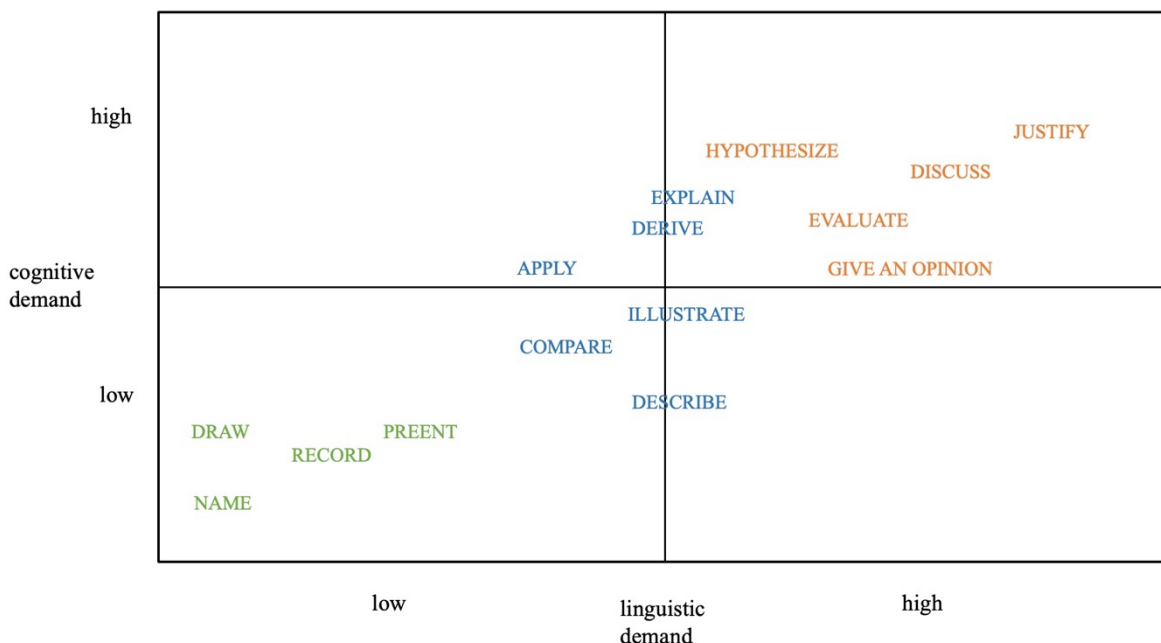
Figure 8.22 - Biology discourse functions on CLIL Matrix (adapted from Cummins, 1984)

Due to the character of language in Biology being patterned and with a tendency to apply fixed phrasing, Biology seems appropriate as a subject being taught early on in bilingual strands. The repetitive character of patterned language, the comparably higher share of CALP, and the higher share of vertical knowledge alludes to Biology being more suitable to start with than Geography. Preferably, CLIL

Biology teaching would start with the second half of year 7 because this way, it allows for students to have accrued sufficient language skills to be able to successfully participate in such classes.

8.4.4   Mathematics

The analysis of the selected tasks for Mathematics revealed that linguistic demand was very low for both age groups. The overall average across all discourse functions arrived at level A1+ for reading and writing, while speaking required A2 (see Figure 8.17). The maximum relative frequency for Mathematics tasks in grade 7 was detected at level A1. For grade 7, most discourse functions were evaluated to impose lower linguistic demand than the targeted level A2 to B1 (cf. KMK 2003: sec. 5) (see Figure 8.16). In contrast to grade 7 material, the average CEFR level for reading increased to level A2, speaking increased to A2+ while writing remained at A1+. Likewise, the maximum relative frequency remained at level A2 (see Figure 8.19). The results detected for the single discourse functions remained in the range with B1 as the maximum requirement (see Figure 8.18).

Even though Mathematics does not apply AFBs, an order regarding the cognitive demand of a discourse function was possible to determine. Placing the discourse functions on the CLIL-matrix revealed that most discourse functions come with high(er) cognitive demands but comparably low(er) linguistic demands. Even the ones with the highest placement for cognitive demand are relatively undemanding regarding language. The classification of Mathematics as a non-interactive subject with limited production, authoritative communication of vertical knowledge explains the low(er) linguistic demand. The aspect of Mathematics being a subject with high shares of CALP language may seem to contradict this assessment in total.

However, the fact that in either language of schooling, subject-related language, i.e., technical terms, would have to be defined and explained, levels out the aspect of CALP language and its perceived high demand.
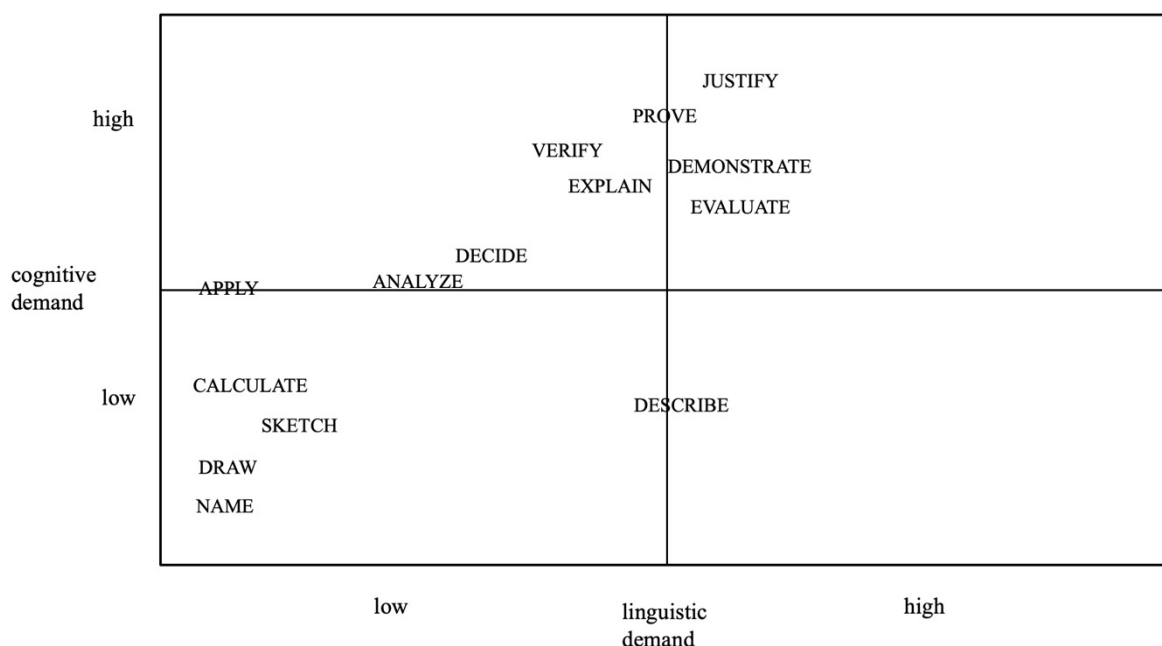
Figure 8.23 - Mathematics discourse functions on CLIL Matrix (adapted from Cummins, 1984)

This is reflected in the evaluation of the individual discourse functions, for which an increase in linguistic demand could hardly be determined from grade 7 onto grade 9. The latter aspect led to the assumption that – linguistically – the demand imposed on the older age group may be too low. Consequently, if knowledge gain bases, among other factors, on the challenge presented by the material as implied by Vygotsky's *Zone of Proximal Development* (see 4.2), bilingual Mathematics may only be suitable for lower age groups since the potential for knowledge gain in the higher age group may be minimal.

8.4.5 Comparison of selected discourse functions

Out of all discourse functions, two were used in all four subjects, DESCRIBE and EXPLAIN. This overlap serves as a base for further comparison of the distinct subjects' demands. Starting with the definitions for the discourse functions, DESCRIBE is defined similarly for History, Geography, and Biology and tasks students with giving details based on their perception of a particular aspect, visual, or other material. In Mathematics, students are to "[…] illustrate their thinking or justifications through verbal (oral or written) statements that refer to a drawing/diagram/mode" (New York State Education Department 2017). Although elaboration here is also on present material, the cognitive demand seems to be comparably higher than with the other three since a justification is required. Linguistically, the level of demand is very much alike for all four subjects, with the discourse function placed on or close to the threshold from low to high linguistic demand on the CLIL Matrix (see Figure 8.20 to Figure 8.23).

The discourse function came with the lowest levels of demand in all four subjects. In Mathematics, DESCRIBE was at level A1 (see Table 8.91) in grade 7 (green columns) and A2 in grade 9 (blue columns). In Biology, DESCRIBE demanded skills resembling competences between CEFR level A2 and A2+ for both age groups. The evaluation of the tasks in Geography for the lower age group placed DESCRIBE between levels A1+ and A2, while the level for the older age group was between A1+ and B1. Lastly, DESCRIBE was evaluated as the most challenging when used in History. Here, the material aimed at the lower age group set the demand imposed with this discourse function mostly to level B1, hinting at possible over-taxation already with an AFB 1-classified discourse function. This demand increased to even level B1 to B2 with the material selected for the higher age group. Regarding the AFBs with which DESCRIBE was allocated per subject, it has to be noted that for History and Geography, the discourse function was placed in AFB 1 while for Biology, DESCRIBE is part of AFB 2. Mathematics does not use AFBs at all; the position on the CLIL-Matrix, however, implies increased cognitive demand, which may make it comparable to the placement used in Biology.

DESCRIBE	Skill						max. relative frequency	
	Reading		Speaking		Writing			
History	A2	B2	B1	B1	B1	B1	B1	B1
Geography	A1+	A2+	A2	A1+	A2	B1	A2	B1
Biology	A2+	A2+	A2+	A2+	A2	A2	A2	A2
Mathematics	A1	-	A1	A2	A1	-	-	A2
grades	7	9	7	9	7	9	7	9

Table 8.91 - Comparison of CEFR levels per skill and max. relative frequency, per subject for DESCRIBE (grade 7 green, grade 9 blue)

This comprehensive overview shows not only that in most cases the demand increases from the lower to the higher group, but also the differences in demand across the subjects.

The second discourse function shared by all four subjects was EXPLAIN. The definitions put to the discourse function per subject differ considerably. While History gives free rein to students to "[…] tell something in a way that helps to understand it better" (Kröger & Lohmann 2007: half-title), Geography limits this freedom to put things into their own words by mandating students to "[…] outline prerequisites and consequences of distinct structures and processes" (Ministerium für Schule und Bildung des Landes Nordrhein-Westfalen 2015a). Likewise, Biology's definition further limits the extent to which students can use language freely to help to understand things better by tasking them to "[…] give information about an actual situation, contextualize regarding rules, principles and causes […]" (KMK 2013b). Mathematics, however, sets the strictest limits when it comes to the definition of EXPLAIN and the activities tied to said definition by requiring students to "[…] use mathematical terminology […] to communicate mathematical ideas, and link concepts and processes within and between mathematical contexts" (NSW Standards Authority 2019). The determined linguistic demand

for EXPLAIN in all four subjects reflected this. For History, the analysis of the tasks placed EXPLAIN between levels B1+ and B2 for both age groups (see Table 8.92). With the onset of limitations following the definition in the context of Geography, the demand for EXPLAIN was calculated to start at level A2 for the lower age group and B1 for the higher age group with a maximal demand at level B1+ in speaking for grade 9. The same effect could be observed for Biology, if to a lesser extent, placing EXPLAIN between levels A2+ and B1 for both age groups. For the most restrictive definition, EXPLAIN used in Mathematics, the CEFR levels arrived ranged from A2 to B1 for both grades. Consequently, when looking at the spread of each subject, it seems as if the more restrictive the definition is set to the use of language, the lower is the linguistic demand while at the same time, the cognitive demand seems to increase (compare Figure 8.20 to Figure 8.23).

EXPLAIN	Skill						max. relative frequency	
	Reading		Speaking		Writing			
History	B2	B2+	B2	B1+	B1+	B2	B2	B2
Geography	A2	B1	B1	B1+	A2+	B1	A2	B1
Biology	B1	A2+	B1	B1	A2+	B1	B1	B1
Mathematics	A2	-	B1	B1	A2	A2	B1	B1
grades	7	9	7	9	7	9	7	9

Table 8.92 - Comparison of CEFR levels per skill and max. relative frequency, per subject for EXPLAIN (grade 7 green, grade 9 blue)

The overview for EXPLAIN illustrates the increased demand from DESCRIBE. In reference to the classification of classroom discourse, it appears as if EXPLAIN allows for more free production of language than DESCRIBE, and thus requiring more language competence, regardless the subject. Likewise, the table shows that, for one exception, an increase in demand between the two age groups was detected. The differences among the subjects are equally visible.

Based on the above-discussed findings from the analysis of the corpus of tasks for all four subjects, a possible sequence of subjects to be taught in bilingual tracks could start with Mathematics in grade 7. Biology may expand this following either in the second half of grade 7 or in grade 8. Although the analysis of the selected tasks for Biology resulted in four discourse functions being assessed as too demanding for the lower age group, while Geography had only two, Biology seems suitable as the second bilingual subject because the character of language is *controlled productive*. The ability to master a subject that allows for *freely productive* language use requires an enhanced lexicon, including language of BICS-quality, which students most likely would not know to a satisfying extent in early grade 8. Hence, Geography would best be introduced following Biology in the second half of grade 8. Additionally, the onset of Geography after Biology takes the aspect of the (overly) demanding textbook material into consideration. Lastly, the tasks analyzed for History revealed that History is the most challenging subject language-wise. The results not only proved Handro's evaluation that History is

already prone to overtaxing students in the language of schooling (see chapter 4.4) but also ties in with the findings Lamsfuß-Schenk discussed determining that only tasks of AFB 1 could be carried out in the target language (cf. Lamsfuß-Schenk 2008: 22).

In summary, the analysis of the tasks and the obtained results imply that Mathematics should be the entry point for bilingual teaching, followed by Biology as another subject employing *patterned phrasing*. With the development of language skills towards CEFR level B1, subjects calling for *free production* of language could be introduced starting with Geography. History should be the last of the four subjects to be taught bilingually since not only the subject as such is language driven, but the character of language differs significantly from the other three (see chapter 4). In the following chapter, the findings of both studies will be brought together to come up with a proposal for a sequence of subjects taking both texts and tasks into consideration.

9   Sequential approach as pathway to CLIL

Synthesizing the results of both studies, i.e., the analysis of the demand imposed on receptive skills when working with textbook material as well as the demand imposed on both receptive and productive skills imposed by the tasks featured in these textbooks, yields a sequential approach to subjects. This sequential approach should mirror the progress students show in the course of their language learning, respecting the idea of Vygotsky's Zone of *Proximal Development* and *Bloom's Taxonomy*, i.e., being challenging but not overtaxing. The results discussed
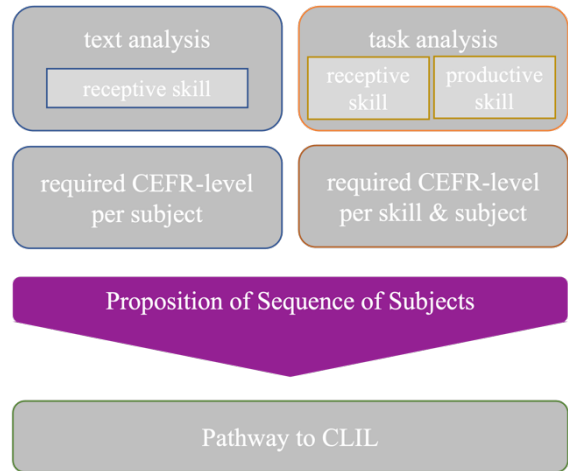
Figure 9.1 - Study Design, Proposition of Sequence

for each subject above (see chapters 7.5 and 8.4) are combined to establish such a sequence. To further support this sequence, the determined proportion of BICS and CALP in the subject-specific language of the distinct subjects is put into perspective with the findings of the two conducted studies and Mentz's categories of the character of language per subject (see chapter 4.2).

9.1   Synopsis of all results and proposed sequence

Referring to the findings discussed in chapter 8.4 and 7.5, an interconnection between lexical sophistication, expressed in the key figures of EVP and AWL, the BICS/CALP proportion, the categorization according to Mentz (see chapter 4.1), and the level of required productive skills seems to be in place. Subjects with a high to a higher intermediate proportion of CALP language displayed the comparably lowest lexical sophistication. Mathematics and Biology had the lowest CEFR levels assigned to EVP and AWL. However, both subjects featured the most significant proportion of CALP language. At the same time, both were least demanding regarding speaking and writing. Remarkably, both were also characterized as controlled or limited productive. On the other hand, Geography and History were assumed to require high intermediate to high proportions of BICS and consequently low proportions of CALP.

Both subjects required an increased to high level of lexical sophistication and were characterized as freely productive (see Figure 9.2).

	interactive freely productive	non-interactive limitedly/ controlled productive
dialogic communication horizontal knowledge BICS		History
authoritative communication vertical knowledge CALP		Geography Biology Mathematics
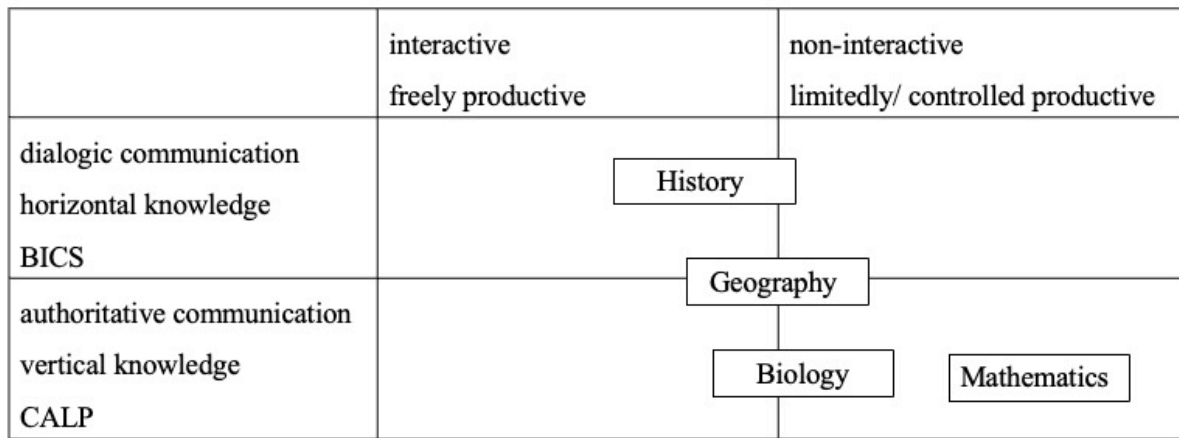
Figure 9.2 - The four subjects in focus allocated with the types of communication, kinds of knowledge and quality of language used (BICS, CALP)

Looking at the subjects and their subject-specific language once more, the reasons for these evaluations and results become apparent. Language in Mathematics and, to a lesser extent, in Biology is dominated by vocabulary for the subject-specific aspects. Some academics even argue that Mathematics is a language of its own, with vocabulary to be learned as an integral part of the subject. This holds true for either language of schooling. *Galileo Galilei* already argued in 1623 that the universe was "written in mathematical language" being the first scientist to call Mathematics an own language with the other language as a tool to talk about it (cf. Silver 2017: 364). This results in a lexically focused and constrained language which ties in with the *limited and controlled production* of having to apply certain predefined phrasings to describe, i.a., calculations and procedures. CALP-classified language may convey a certain level of difficulty and demand since it is conjured of words and phrases that stem from an academic background hardly used in everyday language. However, the results obtained during the two studies show that despite the high proportion of CALP-classified language, the limitations in phrasing and lexis decrease the linguistic demand to a level that allows for Mathematics to be considered as an appropriate and suitable starting point for teaching in a foreign language. Students would have to learn the subject-specific vocabulary and its appropriate use in either language of schooling. Moreover, communicative skills in Mathematics only play a subordinate role requiring a limited use of BICS-classified language, as the production is controlled and limited. Additionally, both studies' results suggested individually that Mathematics imposes the lowest level of linguistic demand. The material as well as the requirements determined through the analysis of the tasks implied a comparably low level of demand (see Table 7.13, Table 7.14, Figure 8.16, and Figure 8.18).

In coherence with the considerations above, Biology makes a suitable second subject to be taught bilingually since the constraints of language production are still relatively tight. At the same time, the proportion of BICS language slowly increases with the subject being open to more discussion and

subject-oriented communication (see Figure 9.2) as outlined in the curriculum (cf. Ministerium für Bildung, Wissenschaft, Weiterbildung und Kultur 2014a). Geography can be introduced as a third subject into a bilingual strand. The proportion of CALP decreases to lower intermediate levels while BICS increases to higher intermediate levels. Lexical sophistication increases to higher intermediate and high levels (see Table 7.5 and Table 7.6), causing the required CEFR levels for productive skills to increase (see Figure 8.8 and Figure 8.10), especially since Geography's language was already categorized as freely productive (cf. Mentz 2015: 259). The last subject to be introduced at a relatively late point in time of bilingual teaching would be History. This is not only justified for the unique role language plays in this subject but also for the high proportions of BICS-classified language, which in part turns to CALP, meaning that everyday language, including that of the past, may become subject-specific language when being the subject of a History lesson. The spinning jenny, e.g., is a term of everyday use during the era of industrialization but was turned into a subject-specific term used in an academic context when researching and discussing the impact of said tool. It can be concluded that the higher the proportion of BICS and the less controlled language production, the more challenging and demanding the subjects regarding lexical sophistication and productive skills are (see Figure 9.3).

	Mathematics	Biology	Geography	History
Level of Lexical Sophistication	■ (comparably low)	■ (high intermediate)	■ (high)	■ (high)
Level of Required Productive Skills	■ (comparably low)	■ (low intermediate)	■ (high intermediate)	■ (high)
Proportion of CALP	■ (high)	■ (high intermediate)	■ (low intermediate)	■ (comparably low)
Proportion of BICS	■ (comparably low)	■ (low intermediate)	■ (high intermediate)	■ (high)
Manner of Language Production	limited or controlled productive	limited or controlled productive	free productive	free productive

■ comparably low   ■ low intermediate   ■ high intermediate   ■ high

Figure 9.3 - Synopsis of results of text and task analysis, BICS and CALP, Manner of Production

In summary, Mathematics is the subject to be introduced first into bilingual strands and to be taught until the middle of the second year of instruction, most likely in grade 8 (see chapter 8.4). Hence, Biology should be introduced at this point, i.e., in the second half of the second year of instruction during grade 8. Biology should be part of bilingual strands until students have progressed to CEFR level B1. Having achieved this level, Geography should be introduced during the second half of the third year of bilingual instruction during grade 9. As the most challenging and demanding subject linguistically, History should be introduced as the last subject to a bilingual strand. The earliest point would be grade 10, which allows students to adjust to the required level of skills in History. The targeted CEFR levels here are B1 progressing to B2, levels which the studies have proven as necessary for successful participation (see Figure 9.4 - Proposed Sequence of Subjects for Bilingual Strands). Ruhm's study on

narrative competence of students at the end of grade 10 further corroborates this sequence as she determined that the majority of students displayed competences equivalent to level B1. Only a minority of the analyzed cohort of 185 students showed spoken production at an advanced level (cf. Ruhm 2014: 207). Accordingly, with History being a subject targeting and requiring high levels of narrative competence (cf. Ministerium für Bildung, Wissenschaft, Weiterbildung und Kultur 2021: 86), the sequence of subjects is feasible as proposed. By introducing History at a relatively late point in their bilingual education, students can prepare for a potential continuation of the subject as a bilingual subject during grades 11 to 13 or the prospect of studying either the language itself or a subject whose academic discourse is mainly in English. The combined findings of the text and task analysis result in a slightly changed sequence in comparison to the initial sequence derived from the text analysis alone.

subject				
targeted CEFR-level	A2/B1	B1	B1	B1/B2
grade	7	8	9	10

■ Mathematics   ■ Biology   ■ Geography   ■ History
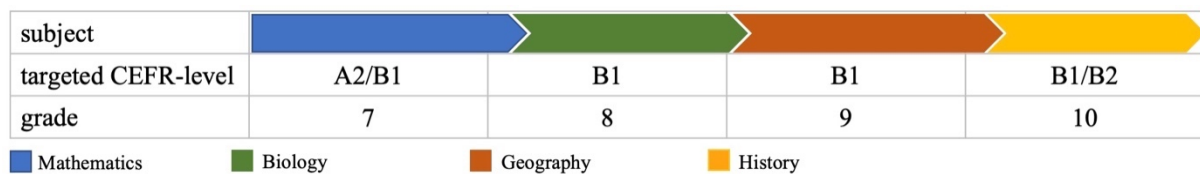
Figure 9.4 - Proposed Sequence of Subjects for Bilingual Strands

With this sequence in mind, based on the findings and the identified demand per subject, a pathway to CLIL can be derived.

9.2   Pathway to CLIL

Schools in Rhineland-Palatinate employ mostly bilingual teaching offering several content subjects, i.a., the four subjects analyzed for this study. The foreign language used most often is English. The teaching material available caters to this need. The order in which these subjects are offered strongly depends on the availability of teaching staff on the one hand and, on the other hand, follows the order in which the potential subjects are introduced through the general curriculum in the various grades. History and Biology, e.g., are introduced from grade 7
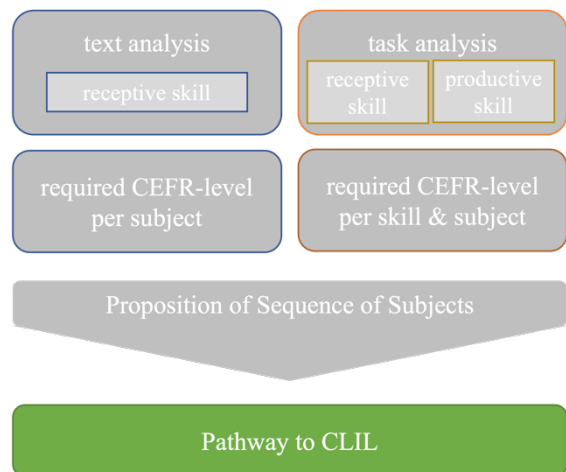
Figure 9.5 - Study Design, Pathway to CLIL

onwards in the regular language of schooling, i.e., German. This sequence was adopted for bilingual teaching as well; a survey of the subjects offered in Rhineland-Palatinate proved this (see Chapter 12.1). Furthermore, the roots of bilingual teaching in Germany seem to justify the early introduction of bilingual History despite the identified demands the subject imposes on students' language competences. After World War 2, bilingual teaching along the French-German border was intended to foster a better understanding among the people who had maintained a historically strained relationship up until that

point. As a consequence, History was proclaimed the prototypical subject for bilingual teaching to be taught initially in French, also, because it was defined as one of the core subjects by the KMK (cf. 2013a). The studies above (see chapters 7.5 and 8.4) showed that the currently applied sequence and choice of subjects do not reflect the level of language competence students possibly display at the point of introduction. Accordingly, History was ruled out as a beneficial subject for bilingual teaching in grades 7 and 8 based on the overtaxing high linguistic demand imposed by the subject.

Only when the language requirements of the distinct subjects are identified does language work become possible in the content subjects taught in the foreign language. While bilingual teaching uses language to convey information and knowledge about the content (see Figure 3.1), CLIL uses the content subjects to teach both then content AND the language, i.e., CLIL uses a bi-directional approach (see Figure 9.6). Other than bilingual teaching, CLIL has a dual focus on content and language teaching (cf. Coyle, Hood & Marsh 2010: 1). At times, the terms bilingual and CLIL are used as synonyms in Germany (cf. KMK 2013a: 4), when in reality in bilingual teaching, language work at best consists of vocabulary work or scaffolding for phrasing, i.e., useful chunks are provided. However, CLIL has to be more than a patch whenever a lexical or grammatical item is missing. Rather, CLIL has to be a systematic approach that intertwines with the curricula of content subjects and English classes. The language has to be used to teach the content subject while the content subject is used to teach the language. CLIL has to be seen as "[…] teaching WITH and THROUGH language" (emphasis as in the original Eurydice Report 2006: 8). This also ties in with VanPatten et al.'s studies that determined that structured input activities regarding grammar were far more beneficial when "presented before an internalized system begins to develop" (Gass 2003: 231). In other words, with CLIL presenting students with grammatical aspects embedded in the content without having introduced said feature beforehand, students are more likely to pick up both the grammar and its meaning than through traditional language teaching. This way, the abstract context in which grammar teaching is usually embedded becomes an authentic context, in which the use of a grammatical structure fosters the understanding of the meaning, both on the content and the language side.
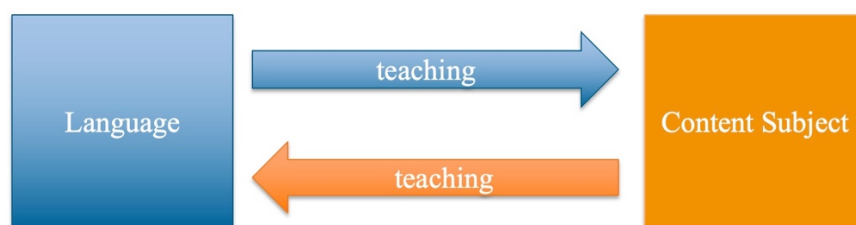
Figure 9.6 - CLIL schema for the use of language and content combined

In accordance with the determined sequence, "real" CLIL teaching in grade 7 should start with Mathematics. Language in this subject has its very own character (see chapters 4.2 and 9.1), being almost yet another language in itself. The English curriculum sets a target for language production that corresponds with the limited or controlled language production required in Mathematics. According to

the English curriculum, language production is based on "classroom phrases" as a basic competence (Ministerium für Bildung 2021: 35). In the studies, spoken production in Mathematics in the lower grades was likewise limited to CEFR level A2+, i.e., a short sequence of sentences using basic expressions (cf. Council of Europe 2018: 71). Similarly, the curriculum limits reading skills to single sentences and very brief texts (cf. Ministerium für Bildung 2021: 39), equivalent to the criteria set for CEFR level A2 (cf. Council of Europe 2018: 63). The analysis identified level A1+ as sufficient for Mathematics at this early stage (see Figure 8.19). Furthermore, the likelihood of cross-language influence is limited. Although learners in early stages show comparably high influence of their L1 in L2 acquisition, this can be reduced in Mathematics as this subject requires learners to learn subject-specific language independent of both L1 and L2 (cf. Kroll & Sunderman 2003: 108). The targets set by the English curriculum can be met with Mathematics, as shown through the analysis conducted beforehand.

Language work in Mathematics could entail, in addition to subject-related vocabulary, the teaching of the use of the simple present as an expression of either eternal truths, an expression of habits or general time, or as the tense being used in instructions (cf. Swan 2015: sec. 463). The former English curriculum lists the simple present as the tense-aspect combination taught first in English (cf. Ministerium für Bildung 2000: 19). Likewise, the current version names the ability to name and describe things and people as well as expressing present aspects as a target for grades 5 and 6 (cf. Ministerium für Bildung 2021: 25). With the help of the content subject matter, e.g., the eternal truth that a square always has four right angles or other mathematical rules, such uses of the tense could be taught independently of grammatical rules but with the aim of students cognitively understanding the reasons and cases in which such a tense has to be used. Students would probably understand the meaning of grammar more easily, which ties in with Langacker's idea of cognitive grammar (cf. Niemeier 2017: 47). With the authenticity provided by teaching and learning in the content subject in the foreign language, students use grammar while at the same time comprehending as to why they use a particular tense or aspect as it is tied to their own activities while carrying out a task or talking about it after having completed the task. The simple past, e.g., could be taught or reinforced when talking about the iterations taken to solve a calculation. This again corresponds to the English curriculum, which targets the ability to express activities and actions of the past (cf. Ministerium für Bildung 2021: 45) and criteria of the CEFR at level A2 which requires short sequences of sentences and phrases with connectors (cf. Council of Europe 2018: 71). The character of limited and controlled language production within the constraints of the subject-related language further supports Mathematics as a suitable first CLIL subject. The proper vocabulary to be used for this subject would have to be learned in either language of schooling.

With Biology as a second subject introduced to a CLIL-strand, the use of language would expand. The English curriculum states that students are to learn to express hypotheses in grades 7 and 8 (cf. Ministerium für Bildung 2021: 45). According to the *English Grammar Profile* (EGP) provided by Cambridge University, hypotheses expressed as conditional sentences type 1 range from CEFR level A2 to B1. Level A2 applies to conditional sentences using the simple present in the main and subordinate

clauses (cf. Cambridge UP 2015: sec. CONDITIONAL). If hypotheses are expressed using the will-future, the CEFR level increases to B1 (ibid.). Either variation of conditionals would suffice to express assumptions on, e.g., experiments. Likewise, the levels assigned by the EGP are on par with the levels identified for Biology as a subject and the discourse function HYPOTHESIZE (see Table 8.59).

Experiments conducted in Biology offer rich opportunities to introduce and apply conditional sentences. Students would have to use both conditional type 1 and 2, allowing them to understand the meaning as potentially possible (type 1) or as hypothetical space in which the predicted or assumed outcome is doubted as likely by the speaker (cf. Niemeier 2017: 148). Furthermore, talking about experiments allows for the use of conditional with a future meaning and the likelihood of the realization, which also fosters an understanding of the grammatical aspect in close connection to the actual use rather than learning grammatical rules. Likewise, the progressive aspect for both currently ongoing activities, e.g., describing an ongoing experiment such as the measuring of the pulse (cf. Mathews & Olmesdahl 2010: 54), or backgrounding activities, could be integrated into CLIL Biology classes. Using the progressive in parallel to ongoing actions and/or processes could help to understand the idea of "zooming in" on activities (cf. Niemeier 2017: 107) and "how the speaker sees the event" (Swan 2015: sec. 470.2). The frequently used discourse function EXPLAIN required enhanced productive skills. According to the curriculum, students apply pre-practiced language patterns within the familiar context of the subject and use linear sequences of sentences to explain or describe matters of increased complexity (cf. Ministerium für Bildung 2021: 37) which corresponds to levels A2 and B1 in the CEFR (cf. Council of Europe 2018: 69ff.). These pre-practiced patterns fit the idea of still limited and controlled language production in Biology (see Figure 9.2) and are reflected in the determined levels of linguistic requirement for the subject and its analyzed material (see Table 7.5 - Table 7.8 and Figure 8.12 and Figure 8.14). Another grammatical aspect that could be included in CLIL Biology is the passive. The EGP places the passive with "by to add information" (Cambridge UP 2015: sec. PASSIVE) at CEFR level A2, passives formed with the infinitive, simple past, or present progressive are at level B1. Both levels are on par with the identified levels required in early bilingual Biology classes (see Figure 8.13). Likewise, the English curriculum aims at the introduction of the passive in grades 7 and 8 (cf. Ministerium für Bildung 2021: 45). Especially experiments allow for a hands-on explanation of energy flow between the agent and the patient, as the energy flow can potentially be observed while happening. This way, the learners realize the importance of either part (cf. Niemeier 2017: 161), following again the ideas of cognitive grammar rather than those of rule-based grammar.

CLIL theory (see chapter 3.1) discusses the four dimensions in CLIL teaching, content, cognition, communication, and culture. The cultural aspect is one of the four C's to which teaching various subjects in a foreign language caters. While most prominently the term "culture" invokes the notion of knowing about different countries in which the target language is spoken, culture can also mean that students get to know the culture of a distinct academic or scientific community. In Biology, English is widely used

in the academic discourse. Hence, learning the subject-specific language through CLIL teaching has a cultural aspect in itself as students familiarize themselves with the discipline's conventions.

Geography was identified as a suitable third subject to be introduced during the third year of bilingual instruction, which would be grades 9 and 10 in this study. To fully benefit from the potential this subject holds for the CLIL approach, students would have to display competence levels at the onset of level B1, enabling them to handle increased complexity (see Figure 8.11). Likewise, the English curriculum, as well as the identified required CEFR levels, call for increased fluency and level of detail (cf. Ministerium für Bildung 2021: 36f., 39; Council of Europe 2018: 63). Furthermore, Geography allows for comparisons, e.g., information on different countries' GDP, which are also mentioned in the English curriculum as the ability to talk about aspects of increased complexity while using appropriate connectors (cf. Ministerium für Bildung 2021: 61). In Geography, based on the material analyzed, students voice their own opinions and takes on certain given aspects more often than in the previous two subjects. Accordingly, the required skill of spoken production increases to what corresponds with level B1 at which students "can express opinions on subjects [and] give simple reasons to justify a viewpoint on a familiar topic." (Council of Europe 2018: 72). This matches the targeted level of skills for grades 9 and 10 in the English curriculum (cf. Ministerium für Bildung 2021: 61). Accordingly, CLIL Geography introduced as a third subject is not only in line with the English curriculum but also offers opportunities to teach grammatical aspects along with lexical and content-oriented knowledge. In the material analyzed, interviews, newspapers, or eyewitness reports were used frequently to convey somewhat authentic impressions of subject-related events such as the tsunami in 2005 or the 2010 Haiti earthquake (cf. Löbmann 2014: 15, 17). These offer the opportunity to explain and practice indirect speech, which is mentioned in the English curriculum for grades 9 and 10 (cf. Ministerium für Bildung 2021: 67). The identified level of demand at B1 for the skills engaged (see Figure 8.10) matches the assessment by the EGP for reported speech, starting at A2 and increasing to B1 (cf. Cambridge UP 2015: se. REPORTED SPEECH). The material provides a certain degree of authenticity allowing to comprehend the shifts in tenses and personal pronouns.

The intercultural component is equally addressed through CLIL Geography lessons as per definition the subject deals with other countries, their cultures, and their living conditions and prompts for comparisons of the own to the foreign culture (cf. Ministerium für Bildung 2021: 70; cf. Ahrend 2013: 111). Lastly, Geography as a subject could also support the understanding of the sequence and interconnection of tenses when discussing, e.g., the genesis of the earth. This way, students can lean on concrete subject-related knowledge to come to an understanding of the meaning and use of the various forms of past tenses. When talking about the Big Bang theory, e.g., students would be completely aware of the temporal distance between the moment of the Big Bang and their discussing it. This could serve as an explanation to understand the meaning of the past tense, precisely the simple past, to express said difference and the developments thereafter as a "sequence of completed events following each other"

(Niemeier 2017: 93). Likewise, the intention of the past progressive could once again be reinforced when discussing co-occurring events in this context.

However, the use of tenses and their order is something for which History as a CLIL subject is predestined. With a late introduction to the canon of CLIL subjects, CLIL History can benefit from the fact that all English tenses have been introduced by the second half of grade 9 and in grade 10 (cf. Ministerium für Bildung 2021: 67). Teaching about historical facts offers a perfect opportunity to clarify and reinforce the use and sequence of tenses. Students get the chance to cognitively understand why, e.g., the past perfect is used to express an "earlier past" (cf. Swan 2015: sec. 423). The EGP also places the use of this tense-aspect combination at levels B1 and B2 (cf. Cambridge UP 2015: sec. PAST). Furthermore, History provides authentic examples of which students know the past and bounded character of situations, the specific event time, and students' temporal distance to these events (cf. Niemeier 2017: 93), and can come to an understanding of grammar and its meaning based on this rather than being limited to the knowledge of how to apply grammatical rules. Another factor in favor of a late introduction of History to a CLIL-strand is the high proportion of BICS-classified language used and to be used in the subject (see Figure 9.3). As was shown by the analysis, this high proportion imposed very high linguistic demands on students (see chapters 7.5.2 and 8.4.1). With a required CEFR level at B1+ and B2 (see Figure 8.5 and Figure 8.7), students in the higher grades achieve these levels as per the English curriculum. Accordingly, students are increasingly able to engage in longer conversations and arguments, to partake in everyday communicative situations, and to adopt and advocate a given point of view (cf. Ministerium für Bildung 2021: 61). These aims correlate with the characterization of language production as being free and spontaneous (see Figure 4.1 - The four subjects in focus allocated with the types of communication, kinds of knowledge and quality of language used (BICS, CALP) as well as with the competences described in the CEFR (cf. Council of Europe 2018: 69 ff.) for spoken production at levels B1+ and B2. But not only does the level of productive skills determined from the analysis indicate a preferably late introduction, the yielded results from the text analysis also call for an introduction at a point at which students can read longer, lengthy, increasingly complex texts (see chapter 7.5.2). As per the English curriculum, this ability or to achieve this ability is the mandate of English teaching in grades 9 and 10 (cf. Ministerium für Bildung 2021: 62). Again, with the character and demand of the texts the History textbook featured (see Table 7.1 to Table 7.4), this subject offers the opportunity to work towards the targeted level of skills in an authentic and reality-based context.

In summary, the proposed sequence of subjects facilitates the teaching of both content and language as equal entities in compliance with the English curriculum, its prescribed sequence of grammatical aspects and the incremental progress of language skills. Furthermore, this sequence of subjects can be regarded as a spiraled curriculum (*Spiralcurriculum*). From the language teaching perspective, language-oriented contents are repeated and reinforced continually along the lines of increasingly challenging subject-oriented contents. CLIL teaching following the proposed sequence of subjects offers the opportunity to practice the use of the language and, foremost, to foster the

understanding of the meaning of grammar in an authentic context without overtaxing students and, thus, giving them the opportunity to participate successfully in these classes

10 Résumé

This thesis set out to systematically analyze teaching material - in specific textbooks - used in bilingual classes in Rhineland-Palatinate to unveil the linguistic demand the material imposes on students. The findings, their implications, and the potential way forward to CLIL teaching are summarized in this chapter.

10.1 Summary

Based on the research design (see chapter 5.2) and the analyses conducted, the results answer the research questions raised in chapter 5.1. The summary, however, deviates from the initial order to allow for a bottom-up approach to provide the answers:

II What is the (average) level of language proficiency required by textbook material in bilingual teaching in Rhineland-Palatinate for the distinct grades?

To determine the average required language proficiency for the textbooks used in the four subjects, History, Geography, Biology, and Mathematics, the analysis employed an automated tool, *Text Inspector*, to obtain information on various key figures which indicate the linguistic demand the texts imposed on students (see chapter 6.3). Preliminary considerations on the intention and character of textbook material already showed significant differences among the four subjects (see chapter 4.1). The selected material (see chapter 6.2) and its treatment corroborated those preliminary findings and revealed further differences concerning the linguistic demand. Based on the key figure results from the automated analysis, the required language proficiency in reading for *Exploring History 1 & 2* arrived at CEFR levels B1 to C1 for the lower grades and CEFR levels B1 to C2 for the higher grades (see chapters 7.1 and 7.5.2). The results obtained for *Diercke Geography Volumes 1 & 2* placed the demand between levels B1+ and C2 for the lower grades. For the higher grades, the demand increased to levels C1 to C2, with single key figures even arriving at the CEFR's academic level D1 (see chapters 7.2 and 7.5.3). The selected material for Biology proved to be less demanding. For grades 7 and 8, the CEFR levels ranged from B1 to B2, while for grades 9 and 10, the demand increased slightly, ranging from B1 to C1 (see chapters 7.3 and 7.5.4). Lastly, the Mathematics material required a marginally adjusted analysis since the selection followed a different process (see chapter 6.2.4). Accordingly, the analysis was not conducted on selected subchapters but a corpus concerned with a thematic field, either Algebra or Geometry. For the lower grades, the determined CEFR levels were between A2 and B1 for both thematic fields combined. For the higher grades, the average demand arrived at level B1+ (see chapters 7.4 and 7.5.5). The average CEFR levels per subject showed significant differences, which had to be considered in answering Research Question IV.

Résumé

Teaching, however, entails more than textbook material. Although texts are central to teaching (see chapter 4.1), for some subjects more than others, classroom discourse is equally important in teaching. Accordingly, the following research question mirrors this aspect by asking:

III What is the (average) level of proficiency required by tasks and their discourse functions within the textbook material for the distinct grades?

The intention of this research question was again to determine the required levels of skill to successfully or much rather satisfyingly carry out the tasks from a language point of view. The required level of language skill has to be appropriate for the targeted age group to ensure successful participation in bilingual classes (see chapter 2.2), acknowledging the interdependence of language proficiency, tasks, and content performance (see chapter 4). With the chosen research methodology (see chapter 8.2), it was essential to verify the results through peer review (see chapter 5.3). Following the coding manual (see chapter 8.2.2) and considering all possible CEFR levels per skill and discourse function before the assessment of the actual task (see chapter 8.2.1), the results for this study were divided into the three skills employed per discourse function and the average required level of skill per subject and age group. This summary starts with the latter.

For History, the required average level for reading in the lower classes arrived at CEFR level B1+. The same level also applied to writing while speaking and the maximum relative frequency of all levels assigned for the material analyzed for the lower age group were at B1 (see chapter 8.3.1, Figure 8.5). For a detailed overview of all discourse functions, see Figure 8.4. The results for the older age group, students in grades 9 and 10, increased for reading and the overall maximum relative frequency to level B2. Writing remained at level B1+ while speaking remained at level B1 (see Figure 8.7). Again, Figure 8.6 provides a detailed overview of all discourse functions.

In Geography, the required language proficiency to carry out the tasks for the lower classes arrived at CEFR level of A2. Reading and writing were slightly more demanding at level A2+, while writing was at level A2 (see chapter 8.3.2, Figure 8.9). The results for the tasks carried out in the higher grades increased to CEFR level B1 for the maximum relative frequency, reading, and speaking. The CEFR level for writing increased only slightly to level A2+ (see chapter 8.3.2, Figure 8.11). A detailed overview of the results for all discourse functions is given in Figures 8.8 and 8.10.

Interestingly, the results for the task analysis in Biology resembled those obtained for Geography for the lower grades. The relative maximum frequency arrived at level B1, while the averages for all three skills were at level A2(+) (see chapter 8.3.3 and Figure 8.13). Other than in Geography, those CEFR levels remained the same for the tasks aimed at the higher grades, except for reading at B1 (see Figure 8.15). Figure 8.12 and Figure 8.14 show the individual results for all discourse functions per age group.

The results for Mathematics revealed that, despite the perception that Mathematics is a challenging subject, the required language proficiency was not demanding. The maximum relative

frequency for the lower age group arrived at CEFR level A1. Both reading and writing required an average level of A1+ while speaking was more demanding at level A2 (see chapter 8.3.4 and Figure 8.17). The levels determined for the lower grades increased only slightly with the results obtained from the material analysis for the higher grades. The maximum relative frequency increased slightly to A2, along with reading and speaking (A2+). The averagely required level of skill for writing remained at level A1+ (see Figure 8.19). For Mathematics, Figure 8.16 and Figure 8.18 provide an overview of the distinct discourse functions and their CEFR level results.

The analysis of the discourse functions further revealed that their definitions vary throughout the subjects. The demand also differed. This difference became evident for two discourse functions used in all four subjects. The assessment of DESCRIBE and EXPLAIN resulted in significantly different CEFR levels for the three skills in each subject. As shown in Table 8.91 and Table 8.92, History imposed the highest demand for discourse functions and age groups. Mathematics, however, had the lowest CEFR levels (see chapter 8.4.5). The implications of those results influenced the answer to Research Question I.

IV Does the required level of skill within bilingual teaching correspond with the levels targeted in the Rhineland-Palatinate curriculum for English? And if not, could bilingual students' proficiency still be sufficient at the given grades to successfully participate in bilingual classes of the selected subjects?

The overarching indirect question, along with the concrete question of linguistic demand imposed by bilingual teaching, was whether students in bilingual tracks are sufficiently skilled, English language wise, to participate successfully in those classes. Accordingly, the benchmark of educational success was discussed in chapter 2.2 to answer this question. Apart from marks being awarded by teachers for assessments, another measure to determine student performance, especially in the languages, is the CEFR and its six different levels (see chapter 2.2). This scale was the primary benchmark in this thesis. A comparison of the CEFR levels determined for the texts and tasks to the targeted level defined by the KMK was called for to answer this research question. Additionally, the findings of the DESI study (see chapter 3.2), concluding that bilingual students' language proficiency, especially those with a selection process, can be considered advanced by approximately one level. That is to say that grade 7 students in regular tracks should achieve CEFR level A2 while students in bilingual tracks would at the same time be more advanced, heading towards level B1. Likewise, students in the higher grades should achieve levels B1 and in parts B2 at the end of grade 10. Students in bilingual tracks probably achieve this target earlier but most likely remain at level B2 since the B2 band is quite expansive (see chapters 4.3 and 6.4.2). To account for the span of those two levels, the lower and the higher end of the targeted levels margined the red-colored zones shown in every figure, visualizing the results for either the key figures or subchapters for texts, the individual discourse functions, or averages per skill. These visualizations

Résumé

(e.g., Figure 7.2 or Figure 7.10, Figure 8.8 or Figure 8.14) demonstrate the relation between the determined demand and the targeted CEFR level.

Accordingly, for the results of the text analyses per subject, the comparison of the results for History showed that for both textbooks and age groups, the demand imposed by the texts was too high (see chapter 7.5.2). For grades 7 and 8, the demand exceeded the targeted levels by up to two levels (see Figure 7.1 to Figure 7.6). For grades 9 and 10, the difference between the demanded and the targeted skill level was comparable. Although the progress students potentially make in their language skills was considered, the required levels exceeded the targeted by up to two levels (see, e.g., Figure 7.8 to Figure 7.12). Those results indicate that successful participation is unlikely in the lower grades. In the higher grades, intense scaffolding and support would be required to ensure educational success.

The texts of *Diercke Geography Volume 1* were also comparably demanding (see chapter 7.5.3). The chapters analyzed for grades 7 and 8 showed that the demand was at least one CEFR level higher than the targeted level (see Figure 7.13 to Figure 7.18). A single key figure result of *Diercke Geography Volume 2* exceeded the regular CEFR scale, demanding a skill level allocated to academic language. Apart from this single figure, the texts for the higher age group demanded skills one to two levels higher than those targeted (see Figure 7.19 to Figure 7.24). Just like with History, the demand resulting from the text complexity very likely jeopardizes educational success in the lower age group and could only be achieved in the higher age group with language-oriented teacher input.

This picture changed with the results for the analysis of *Discover Biology Volume 1* (see chapter 7.5.4). The texts analyzed for grades 7 and 8 were mainly in the targeted range requiring level B2 in only single subchapter. Hence, the targeted levels A2 to B1 were mostly met for the upper margin (see Figure 7.25 to Figure 7.30). The same levels were identified for the demand of the chapters analyzed for the second age group (see Figure 7.31 to Figure 7.36). Since the levels match the targeted levels, this material was considered entirely appropriate for the targeted age group with a good chance of educational success.

Lastly, the texts for Mathematics, which at the same time were also the tasks analyzed, resulted in very low demand imposed on students' reading skills (see chapter 7.5.5). The individual key figures determined for the thematic fields in the lower age group exceeded the targeted level only once (see Figure 7.37 to Figure 7.40). The same applied to the higher age group (see Figure 7.41 to Figure 7.44). Judging from linguistic demand alone, Mathematics is the subject in which successful participation in the early grades is most likely.

Concerning the results for the text analyses, the required skill level matched the targeted skill level for the lower grades only for Biology (with single exceptions) and Mathematics. The texts in Geography demanded more than the targeted levels aim at; this exceeding demand was comparable to the History texts. For the higher grades, the texts in Biology and Mathematics were on par with the targeted levels. The texts in Geography and History were still too demanding; however, the extent to which the demand exceeded the targeted level was not as high and could be managed with appropriate

teacher input. Based on those results, successful participation in bilingual classes in the lower grades is most likely possible in Mathematics and Biology, while the other two subjects are not suitable for the lower but the higher grades. This is supported by Figure 7.45 and Figure 7.47, giving the average CEFR level per key figure and subject. Here, the majority of assessments is not within the red-colored zone indicating the targeted CEFR level.

The second part of this thesis analyzed the linguistic demand imposed by the tasks and their individual discourse functions. Again, visualizations illustrating the obtained results against a red-colored zone representing the targeted levels as per the KMK show possible discrepancies. The results and conclusion from this part are a more substantial argument regarding the research question as tasks and discourse are at the heart of participation. Since the averagely demanded skill levels have already been discussed, the focus here shifts to the discourse functions.

The analysis for History showed that even those discourse functions allocated in AFB 1, concerned mainly with reproduction, are too demanding – if even only slightly (see Figure 8.4). Discourse functions placed towards the center or the upper right quadrant (high cognitive and high linguistic complexity) of the CLIL matrix for History were too demanding requiring skill levels of mainly B2 (see chapter 8.4.1 and Figure 8.20). This demand, however, did not increase considerably with the tasks for the higher grades making successful participation in History classes in grades 9 and 10 more likely (see Figure 8.6). The discourse functions of AFB 3 still pose a challenge that could easily overtax students' language proficiency.

In contrast to the relatively high demand imposed by the texts, the tasks for Geography were not as demanding, even for grades 7 and 8 (see chapter 8.4.2 and Figure 8.8). In this age group, only three discourse functions of AFB 3 were too demanding (see Figure 8.21). The demand leveled with tasks and their demand for the higher grades (see Figure 8.10). This assessment allows for the assumption that successful participation in bilingual Geography would already be possible in the lower grades if students meet or even exceed the targeted CEFR levels set by the KMK and the demand imposed by the textbook material was adjusted.

For Biology, a comparable conclusion is appropriate (see chapter 8.4.3). The analysis revealed that, in general, bilingual Biology classes could allow for successful participation already in the early grades. However, three discourse functions of AFB 3 were too demanding (see Figure 8.12 and Figure 8.22). With students' assumed progress in language skills, those discourse functions arrive exactly at the targeted levels for grades 9 and 10, and thus, success in bilingual Biology as well as subjects requiring the skills of these discourse functions is likely (see Figure 8.14).

Despite being perceived as cognitively challenging, the results for Mathematics showed that the language part of the discourse functions was not (see chapter 8.4.4). All discourse functions and their demand per skill were already within the targeted range of CEFR levels in grades 7 and 8 (see Figure 8.16). Although Mathematics does not employ AFBs, the allocation of discourse functions on the CLIL matrix showed that only three discourse functions impose high linguistic demand, while all other

discourse functions, at most, require a medium linguistic demand (see Figure 8.23). This assessment did not change for grades 9 and 10 (see Figure 8.18). Accordingly, Mathematics probably offers the best opportunity for educational success in the lower age group, while in the higher age group, the linguistic demand appears too low to really challenge students sufficiently to ensure progress.

In agreement with the results of the text analyses, the results of the task analyses support the initial idea that in grades 7 and 8, educational success is most likely in Mathematics and Biology. With respect to the manner of language production in Biology (see 4.2 and Figure 4.1), this subject is regarded as more promising than Geography. In Geography, the determined demand for the tasks would allow for bilingual students to participate successfully. However, the manner of language production and the demand for the texts leads to the assumption that an early onset of bilingual Geography overtaxes students in grades 7 and 8. Accordingly, History is ruled out as a subject letting bilingual students in the lower age groups to participate successfully. This exclusion concerns both aspects, content knowledge and language acquisition. The targeted language skills, respecting the advancement bilingual students show, increase with grades 9 and 10 to such an extent that educational success in Geography and History becomes more likely in this age group, i.e., the targeted skill level increasingly matches the levels demanded by both texts and tasks.

I Is it possible to derive a sequence of subjects to be introduced in bilingual strands that reflects the level of proficiency students achieve to promote both subject and language learning as equal entities and thus be a pathway to CLIL?

The findings that answered the previously discussed research questions RQII-IV allow to derive and build a sequence of subjects that accounts for the targeted level of skills and the development students make throughout the four school years for which texts and tasks were analyzed.

Accordingly, Mathematics should be introduced first. The results from the text and task analyses corroborate this approach (see chapters 7.5.6 and 8.4.5). Mathematics require low levels of lexical sophistication, language production is limited and controlled, and the proportion of BICS is low. The high level of CALP does not oppose the early introduction of the subject into bilingual tracks (see chapter 8.4.4). With Biology as a second subject, the sequence respects the manner of language production still being limited and controlled and the slightly increased levels of lexical sophistication. The slight increase in the proportion of BICS corresponds with the development assumed in language proficiency. Likewise, the only slight decrease in CALP still does not oppose the introduction of bilingual Biology in the early grades, as those subject-specific terms are either translated or defined in the textbooks. Increased language skills are necessary to transition to free language production in classroom discourse. Hence, Geography should only be introduced as a third subject, considering the language level students are at. The earliest onset is possible when all students master CEFR level B1 without limitations (see Figure 9.4). Accordingly, History, with its high demands stemming from text complexity and language skills, the high proportion of BICS, the unique character of texts, and the required free production of speech, should only be introduced at a very late point.

Résumé

This sequence allows for CLIL teaching. Following the sequence with which, e.g., grammatical phenomena are introduced in English classes, the lexical diversity, based on the findings obtained from the English textbooks, increases, and the Flesch-Reading-Ease indices develop over time, teaching content subjects in a foreign language offer manifold opportunities to advance both language and content subject competences and thus, move away from uni-directional bilingual teaching to bi-directional CLIL teaching (see Figure 9.6).

10.2 Limitations and future opportunities

The studies presented in this thesis strove for a comprehensive approach to determine the linguistic demand imposed by the selected bilingual subjects. This aspiration is evident in the volume of data and coverage of subjects and their material. However, the studies were limited, focusing on material only. Another limiting factor was the concentration on Rhineland-Palatinate and the realization of bilingual teaching at their secondary schools. The global pandemic starting in early 2020 hindered the research further. Initially, an exploratory part in schools was planned for but had to be canceled.

The analyzed material, moreover, represents only a fraction of material used for bilingual teaching as teachers often create their own material for the lack of appropriate and ready-made fitting material (see chapter 3.2). Another aspect that could not be respected within the scope of this thesis was the effect potential teacher input may have on the partly very demanding texts and tasks analyzed. Consequently, the presented results have to be understood as a starting point and information for publishers and teachers to either revise the material or offer support for students. Furthermore, the impossibility of researching actual classroom discourse limits the assessment of the discourse to the expected, minimally necessary student performance as perceived reasonable by the researchers.

The studies further intended to strictly separate content and language to avoid interference. This separation, however, was not entirely possible since language comprises the content. Nevertheless, the task analysis mainly remained uninfluenced by the text complexity analyzed in the first part. The required text was isolated and assessed independently for each task to ensure maximum separation of content and language.

Although teaching material has not yet made its way into mainstream CLIL research, it is a field offering vast opportunities for the future. The identified limitations, e.g., could serve as a base for further research. Applying the sequence and the drafted pathway to CLIL, material could be developed and tested in schools. Classroom discourse based on the material analyzed could be researched to verify the findings of this thesis' second study. Furthermore, the canon of subjects could be enhanced, including subjects that not only operate through spoken discourse but also allow for physical responses, such as Physical Education, Music or Arts Education. Further, studies could expand to further school types offering bilingual tracks in Rhineland-Palatinate or even to other federal states.

Basically, this thesis can be seen as a contribution to research in its infancy. Although CLIL and bilingual teaching are very prominent research fields, the research on purely language demands and

material, let alone those combined, is still scarce. The results, however, indicate that there is more to do in the future to develop education in a foreign language towards CLIL – the silver bullet of language education.

10.3   Conclusion

This thesis contributes to a field of research that has not yet gained sufficient attention in CLIL research concerning both material and language competences and proficiency of bilingual students (see chapter 4.4). Furthermore, it provides information and insights from which multiple bilingual and CLIL education stakeholders could benefit.

Based on this thesis' findings, Fein's findings on CLIL teacher education (cf. 2021), and potential future research as outlined in chapter 10.2, the KMK could adopt the identified sequence of subjects and establish this as a national standard. Likewise, in the light of the findings revealing exceeding demand in the core subjects, specialized teacher training should become mandatory to enable teachers to recognize those challenges and react appropriately. Publishers could use the findings yielded from this thesis to revise existing material to adjust the demand. At the same time, they could also enhance their portfolio to other subjects and add language work aspects in the CLIL sense. Equipped with materials of such quality, trained teachers could focus more on the challenges and potential hurdles embedded in the material and on language work than having to worry about material creation. Lastly, the research field has new input to enhance the often-biased discussion on student performance. A shift towards student potential and material is inevitable.

Respecting this potential encompassed with this thesis, which is not to say that this field of research is saturated with this single contribution, a pathway to "true" CLIL teaching in Rhineland-Palatinate can be paved. Adjusting material to the targeted and actual potential of students in the selected grades, following the outlined sequence of subjects in the bilingual strands, and having skilled teachers in schools could allow students to benefit from bilingual education optimally. In a bilingual setting like outlined, employing CLIL rather than bilingual teaching helps to tap into the full potential of teaching content subjects in a foreign language and – ultimately – to fulfill the mandate of teaching in Germany.

11 References

Ahrend, Rosita (ed.). 2013. *Diercke geography for bilingual classes. vol. 2: Textbook*. Druck A3. Vol. 2. Braunschweig: Westermann Schulbuchverlag.

Albrecht, Volker & Maik Böing. 2010. Wider die gängige monolinguale Praxis?! - Mehrperspektivität und kulturelle Skripte als Wegbereiter zur Zweisprachigkeit im bilingualen Geographieunterricht. In Doff, Sabine (ed.), *Bilingualer Sachfachunterricht in der Sekundarstufe. Eine Einführung*. (Narr Studienbücher), 58–71. Tübingen: Narr Francke Attempto Verlag.

Around the world. 1, Schülerbuch. 2014. 1. Aufl. Berlin: Cornelsen.

Around the world. 2, Schülerbuch. 2014a. 1. Aufl., 4. Dr. Berlin: Cornelsen.

Baetens Beardsmore, Hugo. 1993. *European models of bilingual education* (Multilingual Matters 92). Clevedon [England] ; Philadelphia: Multilingual Matters.

Baker, Colin. 2011. *Foundations of bilingual education and bilingualism* (Bilingual Education & Bilingualism 79). 5th ed. Bristol, UK ; Tonawanda, NY: Multilingual Matters.

Baker, David. 2008. *Elevate 1: Mathematics 11-14, Level 5 - 6*. Cheltenham: Nelson Thornes.

Baker, David & Paul Hogan. 2008 *Elevate 1: Mathematics 11-14, Levels 3-4*. Cheltenham: Nelson Thornes.

Baker, David & Paul Hogan. 2008a. *Elevate 1: Mathematics 11-14, levels 4 - 5*. Cheltenham: Nelson Thornes.

Beacco, Jean-Claude. 2007. Items for a description of linguistic competence in the language of schooling necessary for learning/teaching history (end of obligatory education). An approach with reference points. Council of Europe Publishing, Language Policy Unit. www.coe.int/t/dg4/linguistic/Source/Source2012_Sem/semSept/SemScol12_10Linneweber.pptx.pdf. (9 December, 2019).

Beck, Bärbel & Eckhard Klieme. 2007. Einleitung. In Bärbel Beck & Eckhard Klieme (eds.), *Sprachliche Kompetenzen: Konzepte und Messung: DESI-Studie (Deutsch Englisch Schülerleistungen International)*.

Bernstein, Basil. 1999. Vertical and Horizontal Discourse: An essay. *British Journal of Sociology of Education* 20(2). 157–173. https://doi.org/10.1080/01425699995380.

Bildungsserver RLP. 2016. Bilinguale Züge. Pädagogisches Landesinstitut Rheinland-Pfalz. https://fremdsprachen.bildung-rp.de/bilingualer-unterricht.html. (13 November, 2019).

Bildungsserver RLP. 2017. Bilinguale Züge. *Fremdsprachen*. https://fremdsprachen.bildung-rp.de/bilingualer-unterricht.html.

Bildungsserver RLP (ed.). 2022. Unterrichtsfächer- und Themen. https://bildung-rp.de/unterricht/faecher-themen.html. (8 March, 2022).

BNC Consortium (ed.). 2007. The British National Corpus, XML Edition. Oxford Text Archive. http://hdl.handle.net/20.500.12024/2554.

References

Bohn, Matthias & Sabine Doff. 2010. Biologie bilingual: Die Perspektive aus der Unterrichtspraxis. In Doff, Sabine (ed.), *Bilingualer Sachfachunterricht in der Sekundarstufe. Eine Einführung.* (Narr Studienbücher), 72–88. Tübingen: Narr Francke Attempto.

Bonnet, Andreas. 2004. *Chemie im bilingualen Unterricht: Kompetenzerwerb durch Interaktion* (Studien Zur Bildungsgangforschung Bd. 4). Opladen: Leske und Budrich.

Bonnet, Andreas. 2015. Sachfachlicher Kompetenzerwerb in naturwissenschaftlichen CLIL Kontexten. In Bernd Rüschoff, Julian Sudhoff & Dieter Wolff (eds.), *CLIL Revisited, Eine kritische Analyse zu gegenwärtigen Stand des bilingualen Sachfachunterrichts* (Forum Angewandte Linguistik), vol. 54, 165–183. Frankfurt am Main: Peter Lang.

Breidbach, Stephan. 2015. Zum Stand der Entwicklung einer CLIL-Didaktik. In Bernd Rüschoff, Julian Sudhoff & Dieter Wolff (eds.), *CLIL Revisited, Eine kritische Analyse zu gegenwärtigen Stand des bilingualen Sachfachunterrichts* (Forum Angewandte Linguistik), 205–224. Frankfurt am Main: Peter Lang Edition.

Bundesrepublik Deutschland. 2020. *Grundgesetz der Bundesrepublik Deutschland*.

Burwitz-Melzer, Eva & Ivo Steininger. 2016. Inhaltsanalyse. In Daniela Caspari (ed.), *Forschungsmethoden in der Fremdsprachendidaktik*, 256–268.

Butzkamm, Wolfgang. 2002. *Psycholinguistik des Fremdsprachenunterrichts: von der Muttersprache zur Fremdsprache* (UTB 1505). 3., neubearb. Aufl. Tübingen Basel: Francke.

Cambridge UP. 2015. English Grammar Profile. Reference. *English Profile - The CEFR for English*. www.englishprofile.org.

Cambridge UP. 2021. English Vocabulary Profile. Reference. *English Profile - The CEFR for English*. www.englishprofile.org.

Cammarata, Laurent. 2016. Foreign Language Education and the Development of Inquiry-Driven Language Programs: Key Challenges and Curricular Planning Strategies. In Laurent Cammarata (ed.), *Content-based foreign language teaching. Curriculum and Pedagogy for Developing Advanced Thinking and Literacy Skills*, 123–146. New York: Routledge.

Capewell, David. 2002. *Framework maths. Year 7C, Year 7C,*. Oxford: Oxford University Press.

Casson, John. 2001. Edexcel GCSE mathematics. Oxford: Heinemann Educational. https://archive.org/details/edexcelgcsemathe0000unse_x6t1. (11 May, 2022).

CGP (ed.). 2014. *CGP Mathematics for Key Stage Three*. Broughton-in-Furness, Cumbria, UK: CGP.

Chall, Jeanne S. & Edgar Dale. 1995. *Readability revisited: the new Dale-Chall readability formula*. Cambridge, Mass: Brookline Books.

Coetzee-Lachmann, Debbie. 2007. *Assessment of subject-specific task performance of bilingual geography learners: Analysing aspects of subjetc-specific written discourse*. Osnabrück: Osnabrück. https://osnadocs.ub.uni-osnabrueck.de/bitstream/urn:nbn:de:gbv:700-2009030617/2/E-Diss864_thesis.pdf. (6 July, 2021).

References

Cohen, Louis, Lawrence Manion & Keith Morrison. 2011. *Research methods in education*. 7th ed. London ; New York: Routledge.

Commission of the European Communities. 2003. Promoting Language Learning and Linguistic Diversity: An Action Plan 2004 - 2006. https://ec.europa.eu/transparency/regdoc/rep/1/2003/EN/1-2003-449-EN-F1-1.Pdf. (17 September, 2019).

Council of Europe (ed.). 2010. *Common European framework of reference for languages: learning, teaching, assessment*. 10th print. Cambridge: Cambridge Univ. Press.

Council of Europe (ed.). 2018. *COMMON EUROPEAN FRAMEWORK OF REFERENCE FOR LANGUAGES: LEARNING, TEACHING, ASSESSMENT COMPANION VOLUME WITH NEW DESCRIPTORS*. Strasbourg.

Council of Europe (ed.). 2020. *Common European framework of reference for languages: learning, teaching, assessment ; companion volume*. Strasbourg: Council of Europe Publishing.

Coyle, Do, Philip Hood & David Marsh. 2010. *CLIL: content and language integrated learning*. Cambridge, UK ; New York: Cambridge University Press.

Coyle, Do. 2011. Post-method pedagogies: Using a second or other language as learning tool in CLIL settings. In *Content and foreign language integrated learning: contributions to multilingualism in European contexts* (Linguistic Insights : Studies in Language and Communication), vol. 108, 49–73. Frankfurt am Main: Peter Lang.

Coxhead, Averil. 1998. *An Academic Word List*. Wellington, N.Z.: School of Linguistics and Applied Language Studies, Victoria University of Wellington.

Cummins, Jim. 1999. BICS and CALP: Clarifying the Distinction. University of Toronto. https://files.eric.ed.gov/fulltext/ED438551.pdf. (3 December, 2020).

Dalton-Puffer, Christiane. 2003. Learner Attitudes and L2 Pronunciation in Austria. *World Englishes* 16(1). 115–128. https://doi.org/10.1111/1467-971X.00052.

Dalton-Puffer, Christiane. 2007. *Discourse in Content and Language Integrated Learning (CLIL) Classrooms* (Language Learning & Language Teaching). Vol. 20. Amsterdam: John Benjamins Publishing Company. https://doi.org/10.1075/lllt.20.

Davies, Mark. 2004. expressly. *British National Corpus (from Oxford University Press)*. https://www.english-corpora.org/bnc/. (16 August, 2022).

Demarmels, Sascha. Was misst man eigentlich, wenn man Verständlichkeit misst? In *Messen in der Linguistik: Beiträge zu den 5. Tagen der Schweizer Linguistik.*, 105–121.

Department for Education (ed.). 2013a. Mathematics programmes of study: key stages 1 and 2. National curriculum in England. https://www.gov.uk/government/publications/national-curriculum-in-england-mathematics-programmes-of-study. (4 September, 2021).

References

Department for Education (ed.). 2013b. National curriculum in England: mathematics programme of study - key stage 3. https://www.gov.uk/government/publications/national-curriculum-in-england-mathematics-programmes-of-study. (4 September, 2020).

Deutsche Kinder- und Jugendstiftung (ed.). 2014. *Schulerfolg gestalten: ein Leitfaden für Schulen und ihre Partner*. 1. Aufl. Seelze: Kallmeyer in Verbindung mit Klett.

Diehr, Bärbel. 2012. What's in a name? Terminologische, typologische und programmatische Überlegungen zum Verhältnis der Sprachen im Bilingualen Unterricht. In Bärbel Diehr & Lars Schmelter (eds.), *Bilingualen Unterricht weiterdenken: Programme, Positionen, Perspektiven*, 17–36. Frankfurt am Main: Peter Lang.

DIPF & Leibniz-Institut für Bildungsforschung und Bildungsinformation (eds.). 2020. *Bildung in Deutschland: ein indikatorengestützter Bericht mit einer Analyse zu Bildung in einer digitalisierten Welt. 2020*. Bielefeld: wbv.

Doll, Johannes. 2002. *Geschichtstexte verstehen in einer anderen Sprache: eine vergleichende Untersuchung der Leseleistung von Schülern aus deutsch-brasilianischen Begegnungsschulen und deutschen Gymnasien* (Europäische Hochschulschriften. Reihe 11, Pädagogik Bd. 869 = Publications Universitaires Européennes. Série XI, Pédagogie ; v. 869; Bd. 869 = European University studies. Series XI, Education ; 869). Frankfurt am Main ; New York: P. Lang.

Dörnyei, Zoltán. 2007. *Research methods in applied linguistics: quantitative, qualitative, and mixed methodologies* (Oxford Applied Linguistics). Oxford ; New York, N.Y: Oxford University Press.

Dossey, John A., Katherine Halvorsen & Sharon McCrone. 2012. *Mathematics education in the United States 2012: a capsule summary fact book: written for the Twelfth International Congress on Mathematical Education (ICME-12), Seoul, Korea, July 2012*. Reston, VA: National Council of Teachers of Mathematics.

DuBay, William H. 2004. The Principles of Readability. Impact Information. http://www.impact-information.com. (4 December, 2020).

Duran, P. 2004. Developmental Trends in Lexical Diversity. *Applied Linguistics* 25(2). 220–242. https://doi.org/10.1093/applin/25.2.220.

Education Council. 2001. *Report from the Education Council to the European Council "The concrete future onjectives of education and training systems."* Stockholm. (13 June, 2022).

Ellis, Nick C. & Peter Robinson. 2008. An Introduction to Cognitive Linguistics, Second Language Acquisition, and Language Instruction. In Nick C. Ellis & Peter Robsinson (eds.), *Handbook of Cognitive Linguistics and Second Language Acquisition*, 3–24. London: Routledge/Taylor and Francis Group.

English G 21, A. 3, Schülerbuch. 2010. 1. Aufl. Berlin: Cornelsen.

English G 21, A. 5, [Hauptbd.]. 2010a. 1. Aufl., 1. Dr. Berlin: Cornelsen.

English G 21, A. 2, Hauptbd.. 2011. 1. Aufl., 5. Dr. Berlin: Cornelsen.

English G 21, A. 4, [Hauptbd.]. 2011a. 1. Aufl., 3. Dr. Berlin: Cornelsen.

References

English G 21, A. 6, Lehrerfassung. 2011b. 1. Aufl., 1. Dr. Berlin: Cornelsen.

European Centre of Mondern Languages. 2022.
https://www.ecml.at/Aboutus/MemberStateRepresentation/tabid/1255/language/en-GB/Default.aspx. (23 March 2022).

European Commision. 2006. *Content and Language Integrated Learning (CLIL) at school in Europe* (Survey. Eurydice (Brussels)). Brüssel: Eurydice.

European Commission. 2014. *COMMUNICATION FROM THE COMMISSION GENERAL REPORT ON THE ACTIVITIES OF THE EUROPEAN UNION.* Brussels: European Commission. https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:52014DC0012&from=CS. (27 July, 2021).

Eurydice Report. 2017. *Key Data on Teaching Languages at School in Europe.* European Commission.

Fein, Felicitas. 2021. *Educating the Educators - CLIL Teacher Competences and Their Promotion in German Teacher Training: A Case Study.* Koblenz: University Koblenz-Landau.

Flesch, Rudolph. 1948. A new readability yardstick. *Journal of Applied Psychology* 32(3). 221–233. https://doi.org/10.1037/h0057532.

Flesch, Rudolph. 1949. *The Art of Readable Writing.* New York: Harper & Brothers Publishers.

Flesch, Rudolf. 1981. *How to write plain English: a book for lawyers and consumers.* New York: Barnes & Noble Books.

Gass, Susan M. 2003. Input and Interaction. In Catherine J. Doughty & Michael H. Long (eds.), *The Handbook of Second Language Acquisition*, 224–255. Oxford, UK: Blackwell Publishing Ltd. https://doi.org/10.1002/9780470756492.ch9.

Gass, Susan M. & Alison Mackey. 2020. Input, Interaction, and Output in L2 Acquisition. In Bill VanPatten, Gregory D. Keating & Stefanie Wulff (eds.), *Theories in Second Language Acquisition. An Introduction* (Second Language Acquisition Research), 192–222. 3rd Edition. New York, NY: Routledge.

Gautschi, Peter. 2015. *Guter Geschichtsunterricht: Grundlagen, Erkenntnisse, Hinweise* (Geschichtsunterricht erforschen Band 1). 3. durchges. und korr. Aufl. Schwalbach/Ts: Wochenschau.

Gauvain, Mary. 2020. Vygotsky's Sociocultural Theory. In *Encyclopedia of Infant and Early Childhood Development*, 446–454. Elsevier. https://doi.org/10.1016/B978-0-12-809324-5.23569-4.

Geisler. 2012. Operatoren nach Anforderungsbereichen Geographie. Klett Verlag. (5 April, 2022).

Goertz, H. -J. 1995. *Umgang mit Geschichte. Eine Einführung in die Geschichtstheorie.* Reinbeck: Rowolth.

Golay, David. 2005. *Das bilinguale Sachfach Geographie: eine empirische Untersuchung zum sachfachlichen Lernzuwachs im bilingual deutsch-französischen Geographieunterricht in der Sekundarstufe I ; (mit unterrichtsmethodischen Empfehlungen und erprobten Materialien für die*

References

English G 21, A. 6, Lehrerfassung. 2011b. 1. Aufl., 1. Dr. Berlin: Cornelsen.

European Centre of Mondern Languages. 2022.
https://www.ecml.at/Aboutus/MemberStateRepresentation/tabid/1255/language/en-GB/Default.aspx. (23 March 2022).

European Commision. 2006. *Content and Language Integrated Learning (CLIL) at school in Europe* (Survey. Eurydice (Brussels)). Brüssel: Eurydice.

European Commission. 2014. *COMMUNICATION FROM THE COMMISSION GENERAL REPORT ON THE ACTIVITIES OF THE EUROPEAN UNION.* Brussels: European Commission. https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:52014DC0012&from=CS. (27 July, 2021).

Eurydice Report. 2017. *Key Data on Teaching Languages at School in Europe.* European Commission.

Fein, Felicitas. 2021. *Educating the Educators - CLIL Teacher Competences and Their Promotion in German Teacher Training: A Case Study.* Koblenz: University Koblenz-Landau.

Flesch, Rudolph. 1948. A new readability yardstick. *Journal of Applied Psychology* 32(3). 221–233. https://doi.org/10.1037/h0057532.

Flesch, Rudolph. 1949. *The Art of Readable Writing.* New York: Harper & Brothers Publishers.

Flesch, Rudolf. 1981. *How to write plain English: a book for lawyers and consumers.* New York: Barnes & Noble Books.

Gass, Susan M. 2003. Input and Interaction. In Catherine J. Doughty & Michael H. Long (eds.), *The Handbook of Second Language Acquisition*, 224–255. Oxford, UK: Blackwell Publishing Ltd. https://doi.org/10.1002/9780470756492.ch9.

Gass, Susan M. & Alison Mackey. 2020. Input, Interaction, and Output in L2 Acquisition. In Bill VanPatten, Gregory D. Keating & Stefanie Wulff (eds.), *Theories in Second Language Acquisition. An Introduction* (Second Language Acquisition Research), 192–222. 3rd Edition. New York, NY: Routledge.

Gautschi, Peter. 2015. *Guter Geschichtsunterricht: Grundlagen, Erkenntnisse, Hinweise* (Geschichtsunterricht erforschen Band 1). 3. durchges. und korr. Aufl. Schwalbach/Ts: Wochenschau.

Gauvain, Mary. 2020. Vygotsky's Sociocultural Theory. In *Encyclopedia of Infant and Early Childhood Development*, 446–454. Elsevier. https://doi.org/10.1016/B978-0-12-809324-5.23569-4.

Geisler. 2012. Operatoren nach Anforderungsbereichen Geographie. Klett Verlag. (5 April, 2022).

Goertz, H. -J. 1995. *Umgang mit Geschichte. Eine Einführung in die Geschichtstheorie.* Reinbeck: Rowolth.

Golay, David. 2005. *Das bilinguale Sachfach Geographie: eine empirische Untersuchung zum sachfachlichen Lernzuwachs im bilingual deutsch-französischen Geographieunterricht in der Sekundarstufe I ; (mit unterrichtsmethodischen Empfehlungen und erprobten Materialien für die*

References

Praxis) (Geographiedidaktische Forschungen 39). Nürnberg: Selbstverl. des Hochschulverbandes für Geographie und ihre Didaktik.

Green, Anthony, Hanan Khalifa & Cyril J Weir. 2013. Examining textual features of reading texts - a practical approach. *Cambridge English: Research Papers*. Cambridge: Cambridge UP.

Institut zur Qualitätsentwicklung im Bildungswesen. 2019. Grundstock von Operatoren - Mathematik. KMK. https://www.iqb.hu-berlin.de/abitur/dokumente/mathematik. (21 September, 2021).

Hancewicz, Euthecia. 2005. Discourse in the Mathematics Classroom. In ASCD- Association for supervision and curriculum development (ed.), *Literacy Strategies for Improving Mathematics Instruction*, 72–86. Alexandria, Va: Association for Supervision and Curriculum Development.

Handro, Saskia. 2013. Sprache und historisches Lernen. Dimensionen eines Schlüsselproblems des Geschichtsunterrichts. In Michael Becker-Mrotzek, Karen Schramm, Eike Thürmann & Helmut J. Vollmer (eds.), *Sprache im Fach: Sprachlichkeit und fachliches Lernen* (Fachdidaktische Forschungen Band 3), 317–334. Münster New York München Berlin: Waxmann.

Harsch, Claudia. 2017. Proficiency. *ELT Journal* 71(2). 250–253. https://doi.org/10.1093/elt/ccw067.

Hasberg, Wolfgang. 2004. Historisches Lernen im bilingualen Geschichtsunterricht (?). In Andreas Bonnet & Stephan Breidbach (eds.), *Didaktiken im Dialog: Konzepte des Lehrens und Wege des Lernens im bilingualen Sachfachunterricht* (Mehrsprachigkeit in Schule Und Unterricht Bd. 2), 221–236. Frankfurt am Main ; New York: Lang.

Heine, Lena. 2010. Fremdsprache und konzeptuelle Repräsentation: bilingualer Unterricht aus kognitiver Perspektive. In Sabine Doff (ed.), *Bilingualer Sachfachunterricht in der Sekundarstufe. Eine Einführung* (Narr Studienbücher), 199–212. Tübingen: Narr.

Jarvis, Scott. 2013. Chapter 1. Defining and measuring lexical diversity. In Scott Jarvis & Michael Daller (eds.), *Studies in Bilingualism*, vol. 47, 13–44. Amsterdam: John Benjamins Publishing Company. https://doi.org/10.1075/sibil.47.03ch1.

Johansson, Victoria. 2008. Lexical Diversity and lexical density in speech and writing: a developmental perspective. *Working Papers* (Working Papers, Lund University) 53. 61–79.

Kapust, Waltraud. 1998. *Universality in noun classification*. San Jose: San Jose State University.

Kilgo, Margret. 2021. Definitions of Verbs in the Tested Student Expectations Mathematics - Grades 3-11. https://www.newsummerfieldids.org/cms/lib5/TX01917270/Centricity/Domain/35/Math%20-%20-verbs.pdf. (24 June, 2021).

Klieme, Eckhard, Wolfgang Eichler, Andreas Helmke, Rainer H. Lehmann, Günter Nold, Hans-Günter Rolff, Konrad Schröder, Günter Thomé & Heiner Willenberg. 2006. Unterricht und Kompetenzerwerb in Deutsch und Englisch. Zentrale Befunde der Studie Deutsch Englisch Schülerleistungen International (DESI). Deutsches Institut für Internationale Pädagogische Forschung. https://www.dipf.de/de/forschung/pdf-forschung/llib/desi-zentrale-befunde. (9 June, 2022).

KMK. 2003. *Bildungsstandards für die erste Fremdsprache (Englisch/Französisch) für den Mittleren Schulabschluss*.

KMK. 2013. Operatoren für das Fach Mathematik an den Deutschen Schulen im Ausland.

KMK. 2013a. *Konzepte für den bilingualen Unterricht –Erfahrungsbericht und Vorschläge zur Weiterentwicklung".* *Beschluss der Kultusministerkonferenz vom 17.10.2013.* Kultusministerkonferenz. www.kmk.org. (17 September, 2019).

KMK. 2013b. Operatoren für die naturwissenschaftlichen Fächer Biologie, Chemie und Physik an den Deutschen Schulen im Ausland.

KMK. 2014. Bildungsstandards für die fortgeführte Fremdsprache (Englisch/Französisch) für die Allgemeine Hochschulreife: Beschluss der Kultusministerkonferenz vom 18.10.2012. *Sammlung der Beschlüsse der Ständigen Kultusministerkonferenz*. Köln: Carl Link.

KoA-Kompendium. 2013. *KoA - Keiner Ohne Abschluss*. https://koa.rlp.de. (13 June, 2022).

Körner, Henning (ed.). 2016. *Mathematik neue Wege. 7: Arbeitsbuch für Gymnasien 7*. Braunschweig: Bildungshaus Schulbuchverlage.

Körner, Henning (ed.). 2016a. *Mathematik neue Wege. 9: Arbeitsbuch für Gymnasien 9*. Braunschweig: Bildungshaus Schulbuchverlage.

Krechel, Hans-Ludwig. 2010. Lern- und Arbeitstechniken im bilingualen Sachfachunterricht. In Sabine Doff (ed.), *Bilingualer Sachfachunterricht in der Sekundarstufe. Eine Einführung* (Narr Studienbücher), 154–168. Tübingen: Narr.

Kremer, Vanessa. 2017. *Die Bedeutung von Sprache im Mathematikunterricht. Eine empirische Untersuchung anhand der schriftlichen Subtraktion und der Bearbeitung von Textaufgaben in den Jahrgangsstufen 4 bis 12*. Wuppertal: Bergische Universität Wuppertal, Dissertation.

Kröger, Rolf J. (ed.). 2009. *Exploring history: for bilingual classes. 2, [Schülerbd.]: ...* Dr. A 2. Braunschweig: Westermann.

Kröger, Rolf J. & Christa Lohmann (eds.). 2007. *Exploring history: for bilingual classes. 1, [Schülerbd.]: ...* Dr. A, [Nachdr.]. Braunschweig: Westermann.

Kroll, Judith F. & Gretchen Sunderman. 2003. Cognitive Processes in Second Language Learners and Bilinguals: The Development of Lexical and Conceptual Representations. In Catherine J. Doughty & Michael H. Long (eds.), *The Handbook of Second Language Acquisition*, 104–129. Oxford, UK: Blackwell Publishing Ltd.

Kuckartz, Udo. 2012. *Qualitative Inhaltsanalyse: Methoden, Praxis, Computerunterstützung* (Grundlagentexte Methoden). Weinheim Basel: Beltz Juventa.

Kuckuck, Miriam & Jantje Röder. 2017. Förderung des Leseverstehens im Geographieunterricht durch reziprokes Lesen am Beispiel eines Textes zur Entwicklungszusammenarbeit. In Alexandra Budke & Miriam Kuckuck (eds.), *Sprache im Geographieunterricht: bilinguale und sprachsensible Materialien und Methoden*, 231–244. Münster: Waxmann.

References

Kultusministerkonferenz. 2005. *Beschlüsse der Kultusministerkonferenz Einheitliche Prüfungsanforderungen in der Abiturprüfung Geschichte Beschluss der Kultusministerkonferenz vom 01.12.1989 i. d. F. vom 10.02.2005.* https://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/1989/1989_12_01-EPA-Geschichte.pdf. (23 March, 2022).

Landesamt für Statistik RLP. 2018. *Statistische Analysen. Kennzahlen zur Bildung in Rheinland-Pfalz 2017*. Bad Ems: Statistisches Landesamt Rheinland-Pfalz.

Lamsfuß-Schenk, Stephanie. 2008. *Fremdverstehen im bilingualen Geschichtsunterricht: eine Fallstudie* (Mehrsprachigkeit in Schule Und Unterricht Bd. 8). Frankfurt am Main ; New York: P. Lang.

Lamsfuß-Schenk, Stephanie. 2015. Sachfachlicher Kompetenzerwerb in gesellschaftswissenschaftlichen CLIL-Kontexten. In Bernd Rüschoff, Julian Sudhoff & Dieter Wolff (eds.), *CLIL Revisited: eine kritische Analyse zum gegenwärtigen Stand des bilingualen Sachfachunterrichts* (Forum Angewandte Linguistik Band 54), 151–164. Frankfurt am Main: Peter Lang Edition.

Landesregierung Rheinland-Pfalz. 2015. *Verfassung für Rheinland-Pfalz*.

Leaton Gray, Sandra, David Scott & Peeter Mehisto. 2018. Educated Side by Side: The Role of Language in the European Schools. In Sandra Leaton Gray, David Scott & Peeter Mehisto (eds.), *Curriculum Reform in the European Schools*, 49–73. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-71464-6_3.

Leisen, Josef. 2015. Zur Integration von Sachfach und Sprache im CLIL-Unterricht. In *CLIL Revisited. Eine kritische Analyse zum gegenwärtigen Stand des bilingualen Sachfachunterrichts* (Forum Angewandte Linguistik), vol. 54, 225–244. Frankfurt am Main: Peter Lang.

Lidwell, William. 2009. FEATURE: More with less. *Interactions* 16(6). 72–75. https://doi.org/10.1145/1620693.1620711.

Linneweber-Lammerskitten, Helmut. 2010. Items for a description of linguistic competence in the language of schooling necessary for learning/teaching mathematics (at the end of compulsory education). An approach with reference points. Council of Europe Publishing, Language Policy Unit. www.coe.int/t/dg4/linguistic/Source/Source2012_Sem/semSept/SemScol12_10Linneweber.pptx.pdf. (9 December, 2019).

Lipski-Buchholz, Kathrin. 2018. *Bilingualer Mathematikunterricht - Motivation der Schülerinnen und Schüler für Fremdsprache und Mathematik*. Braunschweig: Technische Universität Braunschweig Dissertation.

Little, David. 2010. The linguistic and educational integration of children and adolescents from migrant backgrounds. Council of Europe Publishing, Language Policy Unit. https://rm.coe.int/the-

References

linguistic-and-educational-integration-of-children-and-adolescents/16805a0d1b. (19 September, 2019).

Llinares, Ana & Tom Morton. 2017. *Applied linguistics perspectives on CLIL* (Language Learning & Language Teaching 47). Amsterdam ; Philadelphia: John Benjamins Publishing Company.

Llinares, Ana, Tom Morton & Rachel Whittaker. 2012. *The roles of language in CLIL* (Cambridge Language Teaching Library). Cambridge ; New York: Cambridge University Press.

Llinares, Ana & Rachel Whittaker. 2007. Talking and writing in the social sciences in a foreign language: a linguistic analysis of secondary school learners of geography and history. *Models and Practice in CLIL* (Volumen Monográfico) 83–94.

Llinares, Ana & Rachel Whittaker. 2009. CLIL in Social Science Classrooms: Analysis of Spoken and Written Productions. In Y. Ruiz de Zarobe & R. Jiménez Catalán (ed..), *Content and Language Integrated Learning: Evidence from Research in Europe*. 215-234. Bristol, Blue Ridge Summit: Multilingual Matters.

Llinares, Ana & Rachel Whittaker. 2010. Writing and speaking in the history class: A comparative analysis of CLIL and first language contexts. In *Language Use and Language Learning in CLIL Classrooms* (AILA Applied Linguistics Series; v. 7), vol. 7, 125–144. Amsterdam ; Philadelphia: John Benjamins Publishing Company.

Löbmann, Heinrich (ed.). 2014. *Diercke geography for bilingual classes. Vol. 1, [Textbook]*. Druck A7. Braunschweig: Westermann Schulbuchverlag.

Lötscher, Andreas. 2006. Die Formen der Sprache und die Prozesse des Verstehens. Textverstehen aus grammatischer Sicht. In *Text – Verstehen. Grammatik und darüber hinaus.* (Institut Für Deutsche Sprache. Jahrbuch), vol. 2005, 19–45. Berlin: De Gruyter.

Malvern, David & Brian Richards. 2002. Investigating accommodation in language proficiency interviews using a new measure of lexical diversity. *Language Testing* 19(1). 85–104. https://doi.org/10.1191/0265532202lt221oa.

Malvern, David, Brian J. Richards, Ngoni Chipere & Pilar Durán. 2004. *Lexical diversity and language development: quantification and assessment*. Houndmills, Basingstoke, Hampshire ; New York: Palgrave Macmillan.

Marx, Nicole & Grit Mehlhorn. 2016. Analyse von Lernersprache. In Daniela Caspari (ed.), *Forschungsmethoden in der Fremdsprachendidaktik*, 297–306.

Mäsch, Nando. 1994. Das deutsche Modell bilingualer Erziehung in den Sekundarstufen I und II. In *De vreemde taal als instructietaal: problemen, prioriteiten en perspectieven.*, 27–38. Utrecht.

Maset, Michael. 2015. *Bilingualer Geschichtsunterricht: Didaktik und Praxis* (Geschichte im Unterricht Band 9). 1. Auflage. Stuttgart: Verlag W. Kohlhammer.

Mathews, Horst-Dieter & Simon Olmesdahl. 2010. *Discover Biology. 1, [Hauptbd.]*. 1. Aufl. Berlin: Cornelsen.

References

Mayring, Philipp. 2008. *Qualitative Inhaltsanalyse: Grundlagen und Techniken* (Beltz Pädagogik). 10., neu ausgestattete Aufl., Dr. nach Typoskr. Weinheim Basel: Beltz.

McCarthy, Philip M. & Scott Jarvis. 2010. MTLD, vocd-D, and HD-D: A validation study of sophisticated approaches to lexical diversity assessment. *Behavior Research Methods* 42(2). 381–392. https://doi.org/10.3758/BRM.42.2.381.

Mentz, Oliver. 2010. Alle Fächer eignen sich- oder doch nicht? Überlegungen zu einem bilingualen Fächerkanon. In Sabine Doff (ed.), *Bilingualer Sachfachunterricht in der Sekundarstufe. Eine Einführung* (Narr Studienbücher), 29–43. Tübingen: Narr Francke Attempto Verlag.

Mentz, Oliver. 2015. Zur Fächerproblematik im CLIL-Unterricht. In Bernd Rüschoff, Julian Sudhoff & Dieter Wolff (eds.), *CLIL Revisited: eine kritische Analyse zum gegenwärtigen Stand des bilingualen Sachfachunterrichts* (Forum Angewandte Linguistik Band 54), 245–266. Frankfurt am Main: Peter Lang Edition.

Merino, Jan Ander. 2016. Non-linguistic content in CLIL: Is learning diminished? In David Lasagabaster & Aintzane Doiz (eds.), *CLIL experiences in secondary and tertiary education: in search of good practices* (Linguistic Insights, Studies in Language and Communication Volume 216), 17–44. Bern ; New York, NY: Peter Lang.

Metsisto, Diana. 2005. Reading in the Mathematics Classroom. In ASCD - Association for Supervision and Curriculum Development (ed.), *Literacy Strategies for Improving Mathematics Instruction*. 9–23. Alexandria, Va: Association for Supervision and Curriculum Development.

Michalak, Magdalena, Valerie Lemke & Marius Goeke. 2015. *Sprache im Fachunterricht: eine Einführung in Deutsch als Zweitsprache und sprachbewussten Unterricht* (Narr Studienbücher). Tübingen: Narr.

Milton, James & Thomaï Alexiou. 2009. Vocabulary Size and the Common European Framework of Reference for Languages. In Brian Richards, Michael H. Daller, David D. Malvern, Paul Meara, James Milton & Jeanine Treffers-Daller (eds.), *Vocabulary Studies in First and Second Language Acquisition*. 194–211. London: Palgrave Macmillan UK. https://doi.org/10.1057/9780230242258_12.

Ministerium für Bildung (ed.). 2021. Lehrplan Englisch, Sekundarstufe 1, Klassen 5-10. Rheinland-Pfalz.

Ministerium für Bildung, Wissenschaft und Weiterbildung (ed.). 2000. Lehrplan Englisch, Sekundarstufe 1, Klassen 5-9/10 für Hauptschule, Realschule, Gymnasium, Regional Schule, Gesamtschule.Rheinland-Pfalz.

Ministerium für Bildung, Wissenschaft, Weiterbildung und Kultur. 2007. Rahmenlehrplan Mathematik, Klassenstufen 5-9/10. Rheinland-Pfalz.

Ministerium für Bildung, Wissenschaft, Weiterbildung und Kultur (ed.). 2011. Verwaltungsvorschrift für bilinguale Züge an Gymnasien.

<div align="center">References</div>

Ministerium für Bildung, Wissenschaft, Weiterbildung und Kultur (ed.). 2014a. Lehrpläne für die naturwissenschaftlichen Fächer für die weiterführenden Schule in Rheinland-Pfalz. Biologie, Chemie, Physik, Klassenstufen 7 bis 9/10. Rheinland-Pfalz.

Ministerium für Bildung, Wissenschaft, Weiterbildung und Kultur. 2014b. *Abiturzeugnisse an Gymnasien, beruflichen Gymnasien, Kollegs, Abendgymnasien und Integrierten Gesamtschulen Verwaltungsvorschrift des Ministeriums für Bildung, Wissenschaft, Weiterbildung und Kultur vom 14. Januar 2014 (9413 C / 944 A – 51 410/34/35).* http://landesrecht.rlp.de/jportal/?quelle=jlink&docid=VVRP-VVRP000004288&psml=bsrlpprod.psml. (29 June, 2021).

Ministerium für Bildung, Wissenschaft, Weiterbildung und Kultur. 2018. *Schulordnung für die öffentlichen Realschulen plus, Integrierten Gesamtschulen, Gymnasien, Kollegs und Abendgymnasien (Übergreifende Schulordnung) Vom 12. Juni 2009.* Vol. 223-1–35. (5 November, 2021).

Ministerium für Bildung, Wissenschaft, Weiterbildung und Kultur. 2021. Lehrplan für die Gesellschaftswissenschaftlichen Fächer - Erdkunde, Geschichte, Sozialkunde. Rheinland-Pfalz. https://lehrplaene.bildung-rp.de. (11 November, 2021).

Ministerium für Schule und Bildung des Landes Nordrhein-Westfalen. 2015. Biologie bilingual (English). Übersicht über die Operatoren.

Ministerium für Schule und Bildung des Landes Nordrhein-Westfalen. 2015a. Geographie bilingual (English). Übersicht über die Operatoren.

Moe, Eli, Marita Härmälä, Paula Lee Kristmanson, José Pascoal & Meiluté Ramoniené. 2015. *Language skills for successful language learning. CEFR-linked descriptors for mathematics and history/civics.* Strasbourg: Council of Europe Publishing.

Mortimer, Eduardo Fleury & Philip Scott. 2003. *Meaning making in secondary science classrooms.* Maidenhead ; Philadelphia: Open University Press.

Müller-Schneck, Elke. 2005. *Bilingualer Geschichtsunterricht: Theorie, Praxis, Perspektiven* (Mehrsprachigkeit in Schule Und Unterricht Bd. 3). Frankfurt am Main ; New York: Lang.

Newby, David. 2013. Grammar. (Ed.) Michael Byram & Adelheid Hu. *Routledge Encyclopedia of Language Teaching and Learning.* London and New York: Routledge.

Newmann, Fred M. 1990. Higher order thinking in teaching social studies: a rationale for the assessment of classroom thoughtfulness. *Journal of Curriculum Studies* 22(1). 41–56. https://doi.org/10.1080/0022027900220103.

New York State Education Department. 2017. Glossary of Verbs Associated with the New York State Next Generation Mathematics Learning Standards.

Niemeier, Susanne. 2017. *Task-based grammar teaching of English: where cognitive grammar and task-based language teaching meet* (Narr Studienbücher). Tübingen: Narr Francke Attempto.

References

Nikula, Tarja. 2010. Effects of CLIL on teacher's classroom language use. In *Language Use and Language Learning in CLIL Classrooms* (AILA Applied Linguistics Series; v. 7), vol. 7, 105–124. Amsterdam ; Philadelphia: John Benjamins Publishing Company.

Nold, Günter & Henning Rossa. 2007. Leseverstehen. In Bärbel Beck & Eckhard Klieme (eds.), *Sprachliche Kompetenzen. Konzepte und Messungen. DESI-Studie*, 197–211. Weinheim Basel: Beltz Verlag.

NSW Standards Authority (ed.). 2019. NSW Syllabus for the Australian Curriculum. Mathematics K-10 Syllabus. NSW Education Standards Authority. www.educationstandards.nsw.edu.au. (24 June, 2021).

OECD. 2004. *The PISA 2003 Assessment Framework: Mathematics, Reading, Science and Problem Solving Knowledge and Skills* (PISA). OECD. https://doi.org/10.1787/9789264101739-en.

Oleschko, Sven & Moraitis, Anastasia. 2013. Die Sprache im Schulbuch. Erste Überlegungen zur Entwicklung von Geschichts- und Politikschulbüchern unter Berücksichtigung sprachlicher Besonderheiten. Zeitschrift Bildungsforschung e.V. https://doi.org/10.25656/01:8306.

Olthoff, Sarah. 2018. Sprachliche Kompetenzen im biologischen Fachunterricht. Eine Analyse der niedersächsischen Kerncurricula. *Zeitschrift für Didaktik der Biologie (ZDB) - Biologie Lehren und Lernen*. Zeitschrift für Didaktik der Biologie (ZDB) - Biologie Lehren und Lernen 1-19 Seiten. https://doi.org/10.4119/ZDB-1629.

O'Keefe, A. & G. Mark. 2017. The English Grammar Profile of learner competence: Methodology and key findings. *International Journal of Corpus Linguistics* 224(4). 45–489.

Ortega, Lourdes. 2009. *Understanding Second Language Acquisition*. London: Routledge.

Ortega, Lourdes. 2014. Second language learning explained? SLA across 10 contemporary theories. In Bill VanPatten, Gregory D. Keating & Stefanie Wulff (eds.), *Theories in second language acquisition: an introduction*, 245–272. New York: Routledge.

Owen, Nathaniel, Prithvi Shrestha & Stephen Bax. 2021. *Researching Lexical Thresholds and Lexical Profiles Across the Common European Framework of Reference for Languages (CEFR) Levels Assessed in the Aptis Test*. British Council Assessment Research Awards and Grants. https://www.britishcouncil.org/exam/aptis/research/publications/arags/researching-lexical-thresholds-and-lexical-profiles-across. (19 August, 2020).

Oxford University Press. inverse. *Oxford Advanced Learner's Dictionary*. Oxford University Press. https://www.oxfordlearnersdictionaries.com. (15 July, 2021).

Pandel, Hans-Jürgen. 2012. *Quelleninterpretation: die schriftliche Quelle im Geschichtsunterricht* (Methoden historischen Lernens). 4. Aufl. Schwalbach/Ts: Wochenschau Verl.

Parsons, Richard. 2000. *GCSE Mathematics: the workbook*. Kirkby-in-Furness: Coordination Group.

Piaget, Jean. 2010. *The language and thought of the child*. (Trans.) Marjorie Gabain & Ruth Gabain. Third edition, reprinted. London New York: Routledge Classics.

References

Piesche, Nicole. 2016. *CLIL im naturwissenschaftlichen Unterricht - Auswirkungen auf den Wissenszuwachs und die Motivation im Sachfach. Ergebnisse eines randomisierten kontrollierten Feldexperiments an Realschulen*. Ludwigsburg: Pädagogische Hochschule, Kumultativ.

Piske, Thorsten, Bernd Rüschoff, Julian Sudhoff & Dieter Wolff. 2015. Zum Erwerb der CLIL-Fremdsprache. In *CLIL Revisited. Eine kritische Analyse zum gegenwärtigen Stand des bilingualen Sachfachunterrichtes.* (Forum Angewandte Linguistik), vol. 54, 101–126. Frankfurt am Main: Peter Lang.

Rheinland-Pfalz. 2009. *Schulordnung für die öffentlichen Realschulen plus, Integrierten Gesamtschulen, Gymnasien, Kollegs und Abendgymnasien (Übergreifende Schulordnung). SchulO RP 2009*.

Rolletschek, Helga. 2021b. Effects of Bilingual Biology Teaching at Middle Schools. *Language Teaching Research Quarterly* 23. 23–44. https://doi.org/10.32038/ltrq.2021.23.04.

Ruhm, Hannah. 2014. *Narrative Kompetenz in der Fremdsprache Englisch: eine empirische Studie zur Ausprägung mündlicher Erzählfertigkeiten am Ende der Sekundarstufe I* (Fremdsprachendidaktik, Inhalts- Und Lernerorientiert Band 27). Frankfurt am Main ; New York: Peter Lang Edition.

Rumlich, Dominik. 2016. *Evaluating bilingual education in Germany: CLIL students' general English proficiency, EFL self-concept and interest* (Mehrsprachigkeit in Schule Und Unterricht Band 15). Frankfurt am Main ; New York: Peter Lang.

Rüesch, Peter. 2001. Unter welchen Bedingungen sind Kinder schulisch erfolgreich? In Stefan Mächler (ed.), *Schulerfolg: kein Zufall. Ein Ideenbuch zur Schulentwicklung im multikulturellen Umfeld*, 11–18. Zürich: Lehrmittelverl. des Kantons Zürich.

Schlutow, Martin. 2016. *Geschichte bilingual unterrichten: didaktische Grundlagen und methodische Zugänge* (Kleine Reihe Geschichte - Didaktik und Methodik). Schwalbach/Ts: Wochenschau-Verlag.

Schramm, Karen & Petra Knorr. 2016. Triangulation. In Daniela Caspari (ed.), *Forschungsmethoden in der Fremdsprachendidaktik*, 90–97.

Schröder, Marco. 2022. Email: Anzahl Lehrkräfte mit bilingualer Zusatzausbildung in RLP, Statistisches Landesamt Bad Ems.

Schwarz, Monika. 1996. *Einführung in die kognitive Linguistik*. Tübingen: Francke.

Selkirk, Keith. 1991. *Longman mathematics handbook: the language and concepts of mathematics explained*. Harlow: Longman [u.a.].

Silver, Daniel. 2017. The New Language of Mathematics. *American Scientist* 105(6). 364. https://doi.org/10.1511/2017.105.6.364.

Sonoma County Office of Education (ed.). 2021. Building Options for Discourse - Student & Teachers Responses. Sonoma County Office of Education. https://www.scoe.org/files/building-options-for-discourse.pdf. (24 June, 2021).

Statistisches Landesamt Rheinland-Pfalz. 2020a. *Statistische Berichte 2020. Allgemeinbildende Schulen im Schuljahr 2019/2020. Teil II: Lehrkräfte, pädagogische Fachkräfte*. Statistischer Bericht. Bad Ems: Statisitisches Landesamt Rheinland-Pfalz.

Statistisches Landesamt Rheinland-Pfalz. 2020b. *Statistische Berichte 2020. Studienseminare. Lehrerausbildung für allgemeinbildende und berufsbildende Schulen*. Statistischer Bericht. Bad Ems: Statisitisches Landesamt Rheinland-Pfalz.

Studienseminar Koblenz. 2013. Merkblatt zur Ausbildung für den bilingualen Unterricht.

Sudhoff, Julian. 2015. Zur Materialentwicklung im bilingualen Sachfachunterricht. In Bernd Rüschoff, Julian Sudhoff & Dieter Wolff (eds.), *CLIL Revisited - Eine kritische Analyse zum gegenwärtigen Stand des bilingualen Sachfachunterrichts* (Forum Angewandte Linguistik 54). 267–288. Frankfurt am Main: Peter Lang.

Swan, Michael. 2015. *Practical English usage: easier, faster, reference*. 3. ed., [Nachdr.]. Oxford: Oxford Univ. Press.

Textinspector.com. 2020. Textinspector.com. Online Analysis Tool for Language. *About Us*. https://textinspector.com/help/who-are-we/. (10 December, 2020).

The Editors of the Encyclopedia Britannica. 2021. Marshall Plan | Summary & Significance | Britannica. https://www.britannica.com/event/Marshall-Plan. (4 April, 2022).

Treffers-Daller, Jeanine, Patrick Parslow & Shirley Williams. 2018. Back to Basics: How Measures of Lexical Diversity Can Help Discriminate between CEFR Levels. *Applied Linguistics* 39(3). 302–327. https://doi.org/10.1093/applin/amw009.

Trenkic, Danijela & Meesha Warmington. 2019. Language and literacy skills of home and international university students: How different are they, and does it matter? *Bilingualism: Language and Cognition* 22(2). 349–365. https://doi.org/10.1017/S136672891700075X.

UNESCO Institute for Statistics. 2012. *International standard classification of education: ISCED 2011*. Montreal, Quebec: UNESCO Institute for Statistics. http://www.uis.unesco.org/Education/Documents/isced-2011-en.pdf. (9 June, 2022).

United Nations General Assembly. 1948. Universal Declaration of Human Rights. https://www.un.org/en/about-us/universal-declaration-of-human-rights. (27 July, 2022).

VanderStoep, Scott W. & Deirdre D. Johnston. 2009. *Research methods for everyday life: blending qualitative and quantitative approaches* (Research Methods for the Social Sciences). 1st ed. San Francisco, CA: Jossey-Bass.

Viebrock, Britta. 2007. *Bilingualer Erdkundeunterricht: subjektive didaktische Theorien von Lehrerinnen und Lehrern* (Mehrsprachigkeit in Schule Und Unterricht Bd. 4). Frankfurt am Main ; New York: Lang.

Vollmer, Helmut J. 2010. Fachkompetenz als fachbasierte Diskursfähigkeit am Beispiel Geographie. In Sabine Doff (ed.), *Bilingualer Sachfachunterricht in der Sekundarstufe. Eine Einführung* (Narr Studienbücher). 242–257. Tübingen: Narr Francke Attempto.

References

Vollmer, Helmut J. 2010a. Items for a description of linguistic competence in the language of schooling necessary for learning/teaching sciences (at the end of compulsory education). An approach with reference points. Council of Europe Publishing, Language Policy Unit.

Wannagat, Ulrich. 2010. *Bilingualer Geschichtsunterricht im internationalen Fokus: eine Vergleichsstudie zur Unterrichtspraxis in Deutschland und Hongkong* (Mehrsprachigkeit in Schule Und Unterricht Band 11). Frankfurt am Main: Peter Lang.

Wannagat, Ulrich. 2013. Sprachlernprozesse im bilingualen Geschichtsunterricht. In Stephan Breidbach & Britta Viebrock (eds.), *Content and Language Integrated Learning (CLIL) in Europe. Research Perspectives on Policy and Practice* (Mehrsprachigkeit in Schule Und Unterricht), vol. 14. 203–220. Frankfurt am Main: Peter Lang.

Weeke, Annegret (ed.). 2011. *Invitation to History. 1: From the American Revolution to the First World War*. 1. Aufl., 5. Dr. Berlin: Cornelsen.

Weeke, Annegret (ed.). 2011a. *Invitation to History. 2: From the End of the First World War to the Age of Globalization*. 1. Aufl., 2. Dr. Berlin: Cornelsen.

Wesche-Bingham, Marjorie. 2002. Early French Immersion: How has the original Canadian Model stood the test of time? In Petra Burmeister, Henning Wode, Thorsten Piske & Andreas Rohde (eds.), *An integrated view of language development: papers in honor of Henning Wode*, 357–380. Trier: WVT, Wiss. Verl. Trier.

Wittgenstein, Ludwig. 2016. *Logisch-philosophische Abhandlung: = Tractatus logico-philosophicus* (Edition Suhrkamp 12). 36. Auflage. Frankfurt am Main: Suhrkamp.

Wolff, Dieter & Julian Sudhoff. 2015. Zur Definition des Bilingualen Lehrens und Lernens. In Bernd Rüschoff, Julian Sudhoff & Dieter Wolff (eds.), *CLIL Revisited. Eine kritische Analyse zum gegenwertigen Stand des bilingualen Sachfachunterrichts* (Forum Angewandte Linguistik), vol. 54. 9–42. Frankfurt am Main: Peter Lang.

Wunderlich, Sarah. 2018. *History Classes as a Prime Example for Content Language Integrated Learning (CLIL) - A Critical Assessment*. Koblenz: University of Koblenz and Landau, Masterthesis.

Ziafar, Meisam & Ehsan Namaziandost. 2019. A Formulaic Approach to Propositional Density and Readability. *International Journal of Innovation and Research in Educational Sciences* (6, 6). 816–821.

Ziepprecht, Kathrin, Julia Schwanewedel, Patricia Heitmann, Malte Jansen, Hans Ernst Fischer, Alexander Kauertz, Iwen Kobow, Jürgen Mayer, Elke Sumfleth & Maik Walpuski. 2017. Modellierung naturwissenschaftlicher Kommunikationskompetenz – ein fächerübergreifendes Modell zur Evaluation der Bildungsstandards. *Zeitschrift für Didaktik der Naturwissenschaften* 23(1). 113–125. https://doi.org/10.1007/s40573-017-0061-8.

Zydatiß, Wolfgang. 2013. Generalisierbare sprachlich-diskursive Kompetenzen im bilingualen Unterricht (und darüber hinaus). In Stephan Breidbach, Britta Viebrock & Dieter Wolff (eds.),

Content and language integrated learning (CLIL) in Europe: research perspectives on policy and practice (Mehrsprachigkeit in Schule Und Unterricht), vol. 14, 315–332. Frankfurt am Main: Peter Lang.

12 Appendix

The information referenced to throughout the thesis as being in the appendix can be found on the CD-ROM attached to this thesis.

12.1 Research results of most frequently offered subjects in bilingual strands

This part gives an overview of the results of the preliminary query on schools in Rhineland-Palatinate with bilingual strands and – as far as accessible online – which subjects are offered in those strands.

12.2 Research results of most frequently used textbooks in bilingual strands

Here, further information can be found on the most frequently used bilingual textbooks in Rhineland-Palatinate. The results stem from the same query conducted for 12.1, information was not given for all bilingual schools online.

12.3 Numerical values based on Text Inspector results

To ease the reading process within chapter 7, the overview tables with the numerical values obtained from *Text Inspector* were taken out of the main document. The corresponding CEFR level remained in chapter 7 and are explained with the numerical values in the text.

12.4 Topic mapping of Mathematics material to German curriculum

As described in chapter 7.2.4 and 8.2.1, the selection process for Mathematics material involved a topic mapping to ensure the material selected matched the topics covered the German curriculum.

12.5 Data treatment: substitution of technical terms in Mathematics

As described in chapter 7.4, all texts were treated, i.e. technical or subject-specific terms were substituted with the explanations and translations provided by the textbooks. Since for Mathematics, a comprehensive textbook was not available, the substitutions were compiled from different sources. This compilation is given in this part of the appendix.

12.6 Mapping tables of discourse functions to potential CEFR level

This part of the appendix contains the mapping tables created as a preparatory step to the task analysis. The listing of those tables follows the order of subjects adopted throughout the complete thesis.

    12.6.1 History

    12.6.2 Geography

    12.6.3 Biology

    12.6.4 Mathematics

12.7 Decision tree as coding manual for task evaluation

The overview over the complete decision tree is printed in a larger format and attached to the endsheet.

12.8 Overview of task analysis results

In this file, all results of the task analysis are collected. The individual analytical steps are left out and just the final results after the peer coding process are summarized and further analyzed for the maximum relative frequencies.

12.9 Curriculumvitae

Persönliche Daten

Geburtsdatum	02.04.1981 in Lahnstein
Staatsangehörigkeit	deutsch
Familienstand	verheiratet, 2 Kinder

Studium

07/2019 – heute	Promotion im Fach Anglistik, Teildisziplin Didaktik/Linguistik: „Sequence of Subjects in German Bilingual Strands – A Pathway to CLIL" (Abschluss in 2023 geplant)
10/2013 – 03/2018	Studium der Anglistik und Geschichtswissenschaft für das Lehramt an Gymnasien, Masterabschluss (Note 1,3)
09/2003 – 11/2005	VWA-Studium der internationalen BWL, Abschluss als Wirtschaftsassistentin International Business Administration VWA

Berufstätigkeit

05/2023 – heute	LfbA für Sprachpraxis und Didaktik am Campus Koblenz, Institut für Anglistik
07/2019 – 05/2023	Universität Koblenz-Landau, Campus Koblenz, Institut für Anglistik, Projekt MoSAiK, Teilprojekt 4, CLIL-Zertifikat Institut für Anglistik, Lehre in der Sprachpraxis
04/2018 – 06/2019	Universität Koblenz- Landau, Campus Koblenz, Zentrum für Lehrerbildung, Gesamtkoordinatorin Projekt MoSAiK
09/2008 – 03/2014	Lufthansa Technical Training, Beratung zur Ausgestaltung von nicht-technischen Trainings und deren Vertrieb Key Account Management
11/2005 – 08/2008	Lufthansa Systems AG, Product Management Online Vertrieb Airline Ticket Sales: Product Management Revenue Integrity Airline Sales

Berufsausbildung

09/2003 – 11/2005	Deutsche Lufthansa AG, Ausbildung zur Luftverkehrskauffrau, VWA-Studium der internationalen BWL, Abschluss als Wirtschaftsassistentin International Business Administration VWA
08/2000 – 07/2003	Lufthansa Systems AG, Ausbildung zur Fachinformatikerin der Anwendungsentwicklung

Schulbildung

1991 – 2000	Bischöfliches Cusanus Gymnasium Koblenz, Abitur (Note 2,2)
1987 – 1991	Grundschule Rhens

12.10 Honor pledge

Hiermit versichere ich, dass ich die als Dissertation vorgelegte Abhandlung in keinen anderen Verfahren zur Erlangung des Doktorgrades oder als Prüfungsarbeit für eine akademische oder staatliche Prüfung eingereicht habe. Darüber hinaus versichere ich, dass ich die Arbeit selbstständig verfasst, keine anderen als die von mir angegebenen Quellen und Hilfsmittel benutzt und die den Werken wörtlich oder inhaltlich entnommenen Stellen kenntlich gemacht habe.

Koblenz, im Juli 2023

_____

Sarah Wunderlich