

University of Antwerp
Faculty of Arts (FLW)
Faculty of Business and Economics (FBE)

**The Effect of Intralingual Subtitles on Incidental Vocabulary Acquisition
in A1-A2 Learners of Spanish as a Foreign Language for Specific Purposes**

Master thesis submitted for the Master's degree in Professional
Communication and Management at the University of Antwerp to be
defended by Elisa Mamés Antuña

Supervisor:
Prof. Dr. Lieve Vangehuchten

Co-supervisor:
Prof. Dr. Pilar Pérez Cañizares

Assessor:
Prof. Dr. Sarah Bernolet

Academic year 2025-2026



**University
of Antwerp**

**The Effect of Intralingual Subtitles on Incidental Vocabulary Acquisition in
A1-A2 Learners of Spanish as a Foreign Language for Specific Purposes**

Copyright.

© Copyright: Universiteit Antwerpen Without written permission of the supervisor(s) and the authors it is forbidden to reproduce or adapt in any form or by any means any part of this publication. Requests for obtaining the right to reproduce or utilize parts of this publication should be addressed to Universiteit Antwerpen, Master in de Professionele Communicatie & Management, Prinsstraat 13, 2000 Antwerpen (België) | mpcm@uantwerpen.be A written permission of the supervisor(s) is also required to use the (original) methods or products, and for submitting this publication in scientific contests.

Declaration of non-plagiarism .

The undersigned, Elisa Mamés Antuña, student in the Master's program in Professional Communication & Management (FBE-FLW), declares that this thesis is entirely original and has been written solely by her. For all information and ideas derived from other sources, the undersigned has explicitly and in detail cited the sources.

Brussels, 26/05/2026

Preface

Writing a master's thesis on the effect of subtitles on vocabulary acquisition in a foreign language for specific purposes is not an easy task. There are different viewpoints on how subtitles could or could not benefit beginner learners, but one thing remains common: subtitles are a tool used in educational contexts, but also during leisure time.

Being multilingual myself and having a special affection for languages and for how the acquisition of one happens, this topic spoke deeply to me. With this thesis, I did not only investigate the effect subtitles have on vocabulary acquisition, but I also thought further and deeper than what the results showed me. This thesis was a pleasure to work on and made me discover how much I appreciate learning more about the interesting topic of language acquisition.

Although writing a master's thesis usually happens alone, I would never have been able to do this without the support of all the people around me. First, I would like to thank my supervisor, Lieve Vangehuchten, and my co-supervisor, Pilar Pérez Cañizares, for their support, critical feedback, interesting conversations, and their trust in my ability to complete this thesis.

Next to my promotors, I want to thank my parents and godfather for their support throughout all my years of studying, and for always knowing that I could accomplish whatever I set my mind to.

Lastly, I want to thank my boyfriend, for his unconditional emotional support, his help, his encouragement, and for always being there for me.

Everyone believed in me throughout this journey, and for that I want to thank them from the bottom of my heart.

Table of contents

Copyright

Declaration of non-plagiarism

Preface

Abstract

Introduction 5

Chapter 1 Theoretical Framework **9**

1.1 Introduction 9

1.2 Theoretical perspectives 10

 1.2.1 Input Theory 10

 1.2.2 Dual-Coding Theory 11

 1.2.3 The Theory of Multimedia Learning 11

 1.2.4 Cognitive load theory & Linguistic Influence 11

1.3 Incidental Vocabulary Learning Through Subtitles 12

 1.3.1 Incidental language learning 12

 1.3.2 Types of subtitles: interlingual and intralingual 13

 1.3.3 Intralingual subtitles for educational purposes 14

1.4 Conceptual Model 16

Chapter 2 Research Objective and Central Research Question **18**

2.1 Research objective 18

2.2 Central research question 18

2.3 Subquestions 19

 2.3.1 Subquestion 1: Learner Background Variables 19

 2.3.2 Subquestion 2: Comprehension and short-term retention 20

Chapter 3 Methodology **21**

3.1 Research Type 21

3.2 Data Collection 21

3.3 Research Units 22

3.4 Measuring Instruments	24
3.5 Data-analysis	25
3.5.1 Phase 1: Codebook & IRR	25
3.5.2 Phase 2: Statistical analysis	26
3.5.3 Recall and Production – raw scores	27
3.5.4 Detailed on Production	28
3.5.5 Comprehension Scores	29
3.5.6 Prior Knowledge and Cross-linguistic Influence on New Vocabulary Retention	30
3.6 Ethical Considerations	31
Chapter 4 Results	32
4.1 Introduction	32
4.2 Inter Rater Reliability (IRR) of the Codebook	32
4.3 Results Recall and Production – raw scores	33
4.3.1 Recall scores from R	33
4.3.2 Production scores from R	34
4.4 Results Detailed production	36
4.5 Results Comprehension scores	38
4.6 Results Prior Knowledge and Cross-linguistic Influence	39
4.6.1 Recall and Prior Knowledge	39
4.6.2 Production and Prior Knowledge	40
4.6.3 Recall and Cross-linguistic influence	41
4.6.4 Production and Cross-linguistic influence	43
Chapter 5 Conclusion and Discussion	45
5.1 Discussion	45
5.1.1 Recall	45
5.1.2 Production	46
5.1.3 Comprehension Score	47
5.1.4 Prior Knowledge and Cross-linguistic Influence	49
5.2 Conclusion	50
5.3 Limitations	52
Appendix 57	
Appendix 1: Codebook	57

Appendix 2: Informed Consent	61
Appendix 3: Assistance with third-party research	65
<u>Bibliography</u>	<u>67</u>
<u>Use of AI Tools</u>	<u>71</u>
<u>List of tables</u>	<u>74</u>
<u>List of images</u>	<u>80</u>

Abstract

Although audiovisual input and subtitling are increasingly common in language learning contexts, relatively little is known about their effectiveness for learners at the earliest stages of proficiency. Drawing on Input Theory, Dual-Coding Theory, the Theory of Multimedia Learning, and Cognitive Load Theory, this study investigates the effect of videos with intralingual (L2) subtitles on the incidental vocabulary acquisition of A1–A2 learners of Spanish as a foreign language for specific purposes. The study involved 172 first-year students enrolled in Spanish for Business and Economics, who were divided into two conditions: videos ‘with’ subtitles and videos ‘without’ subtitles. Four analyses were conducted to address the research questions. The first analysis focused on the raw scores of recall and production questions. The second analysis examined qualitative responses of individual students for the production questions. The third analysis investigated the relationship between comprehension scores and vocabulary gains, while the fourth explored the extent to which prior knowledge and cross-linguistic influence affected the acquisition of new vocabulary. The results indicate that L2 subtitles support learners’ recognition and recall of new vocabulary items, although their effect on productive vocabulary acquisition was less significant. Comprehension scores appeared to be more strongly related to learners’ existing vocabulary knowledge than to the subtitle condition itself. Prior knowledge of Spanish and cross-linguistic influence from other Romance languages provided learners with a clear advantage in acquiring new vocabulary, whereas cross-linguistic influence from non-Romance languages did not result in greater advantages compared to the Dutch-only speaking group. The findings regarding recall, production, and comprehension are largely consistent with existing literature on subtitle-supported vocabulary learning, while the results concerning cross-linguistic influence differ from those commonly reported in previous research. Nevertheless, these findings should be interpreted in light of the study’s limitations as only a linear mixed model was used to analyse the data, although this data was non-linear.

Introduction

Rationale and positioning

The acquisition of a foreign language depends to a large extent on the availability of adequate and well-targeted input. One tool that may support this process is the use of audiovisual materials adapted to the learner's proficiency level. In present day second language (L2) education and audiovisual input has become increasingly central (Bensalem, 2018; Danan, 2004; Peters et al., 2016; Van der Kolk & Feijoo, 2024), particularly because it exposes learners to authentic or semi-authentic language use in meaningful contexts. Viewing videos in the target language can help vocabulary acquisition, listening comprehension and an increased awareness of how language is used in natural or scripted interaction. Research has suggested that audiovisual input may improve receptive vocabulary knowledge, form recognition, and speech stream decoding (Montero Pérez, 2022). It has to be noted that subtitles may support comprehension and memory coding, but not necessarily listening skills themselves (Bird & Williams, 2002).

Within audiovisual input, subtitles occupy a particularly important position. Intralingual subtitles (L2 subtitles; i.e., subtitles presented in the same language as the audio) have repeatedly been associated with positive effects on vocabulary learning and listening comprehension. Meta-analytic research demonstrates beneficial effects on receptive vocabulary acquisition and comprehension skills across a range of L2 learning contexts (Montero Pérez et al., 2014; Kurokawa et al., 2024). From a cognitive perspective, subtitles are assumed to facilitate dual-channel processing by combining auditory and written input, which may strengthen lexical encoding and retention (Paivio, 1986; Winke et al., 2013).

General social observations confirm the educational potential of subtitling. Typically, foreign-language audiovisual materials are presented with interlingual subtitles (spoken in the L2 and written in the L1), and countries where this practice is common tend to report higher levels of foreign language proficiency than countries where dubbing is the norm, such as Spain and

Italy (Almeida & Costa, 2014; Baumeister et al., 2025; Black, 2021). At the same time, media consumption habits have changed significantly. According to Baumeister et al. (2025), people spend an average of 22.5 hours per week consuming television-related content. The rise of streaming platforms such as Netflix and online media platforms such as YouTube has increased viewers' access to original-language audiovisual content with captions (intralingual) or interlingual subtitles, thereby allowing audiences, especially younger viewers, to choose their preferred viewing format more freely (European Union, 2024). Younger generations increasingly opt for original-language versions with captions or subtitles (European Union, 2024). A survey from the European Union for example, suggest that approximately 65% of European individuals aged 15–24 prefer original-language versions with subtitles (interlingual or intralingual), whereas 44% of viewers aged 55 and older still prefer dubbed versions in their respective language.

Despite growing evidence supporting subtitling, dubbing remains the preferred option in many cases (Almeida & Costa, 2014; Danan, 2004; Van der Kolk & Feijoo, 2024). This preference is partly attributed to limited knowledge of the educational benefits associated with subtitling (Vanderplank, 2016), but also deeply rooted in historical and cultural viewing habits (Almeida & Costa, 2014; Baumeister et al., 2025; Van der Kolk & Feijoo, 2024). In countries where dubbing has historically dominated audiovisual media, audiences are generally accustomed to this format due to prolonged exposure from an early age (Almeida & Costa, 2014). Consequently, the preference for dubbing over subtitling can also be understood as a result of long-standing educational and communication policies (Van der Kolk & Feijoo, 2024).

As a result, opportunities for incidental language learning through audiovisual exposure may be underutilized, particularly among young and beginner learners. Although previous research has consistently demonstrated positive effects of intralingual subtitles, less is known about their effectiveness for learners at the earliest stages of language proficiency (A1), who possess only limited prior knowledge of the target language and whose literacy skills in that language may still be developing (Bensalem, 2018; Peters & Webb, 2018). The rationale for the present study therefore lies in the growing need to identify effective, accessible, and low-

threshold methods for supporting language acquisition during its initial stages. Subtitled audiovisual input may represent such a method; however, further empirical evidence is required to determine whether intralingual subtitles are beneficial for beginner learners and under which conditions these benefits may emerge.

Defining the scope of the study

This study focuses on the use of subtitles in audiovisual materials as a didactic tool for foreign language learning. Audiovisual content offers a particularly relevant context for examining vocabulary acquisition because it combines spoken language with visual information, thereby creating a rich and contextualized learning environment in which learners can infer meaning from multiple sources simultaneously. In addition to linguistic input, visual context, and situational cues may support learners' comprehension processes and facilitate the noticing of unfamiliar lexical items. As vocabulary knowledge forms a fundamental component of communicative competence, investigating how audiovisual input contributes to lexical development is pedagogically relevant.

Due to the limited prior knowledge of the target language for beginners, it is unclear whether written support in the same language as the audio facilitates comprehension and vocabulary acquisition or instead overloads cognitive processing. While subtitles may promote noticing and recognition of unfamiliar words by providing orthographic reinforcement alongside auditory input (Montero Pérez et al., 2014), alternatively simultaneously processing written and spoken language may impose additional cognitive demands, particularly for learners with low proficiency levels (Winke et al., 2013; Montero Pérez, 2022). Reynolds et al. (2022) further note that beginner-level learners are underrepresented in studies investigating the effects of intralingual subtitles, revealing a gap in the existing literature despite the potential pedagogical relevance for early-stage language learners.

Within this context, the study examines the effect of videos with intralingual (L2) subtitles compared to videos without subtitles on the incidental vocabulary acquisition of A1–A2 learners of Spanish as a foreign language. Incidental vocabulary learning refers to situations

in which learners are not explicitly instructed to memorize lexical items but at the same time acquire new vocabulary as a by-product of meaning-focused exposure (Bairstow & Lavour, 2015; Bensalem, 2018). For beginning learners, whose lexicon is still relatively limited, such incidental exposure may play an important role in early vocabulary development and in building familiarity with recurring lexical patterns. For beginners, who still have a rather limited vocabulary, such indirect vocabulary environments can play an important role in the initial development of vocabulary and in ensuring the repetition of syntactic patterns.

Given the broad range of possible outcomes associated with audiovisual input, this study limits its focus specifically to incidental vocabulary acquisition rather than other aspects of language development such as grammar acquisition, pronunciation, or speaking fluency. In addition, the study concentrates exclusively on intralingual subtitles and does not include other modalities, such as interlingual subtitles or dubbing. This delimitation allows for a more focused examination of the educational value of intralingual subtitles during the early stages of foreign language learning.

The study is guided by the following central research question: “What is the effect of videos with L2 subtitles compared to videos without subtitles on the incidental vocabulary acquisition of A1–A2 learners of Spanish as a foreign language for specific purposes?”

In addressing this question, this master’s thesis aims to contribute to the ongoing discussion concerning the educational value of intralingual subtitles in early-stage L2 acquisition and to provide empirically grounded insights for their potential use in both educational settings and home environments.

Chapter 1 Theoretical Framework

1.1 Introduction

Research on the effectiveness of using subtitles in the classroom context remains inconclusive, partly because studies have been conducted with highly varied learner profiles, proficiency levels, and educational objectives. Understanding how intralingual subtitles support second language acquisition requires a theoretical perspective that accounts not only for learning outcomes, but also for the cognitive processes underlying multimedia language input. Because subtitle-supported audiovisual materials simultaneously engage auditory, visual, and linguistic processing, their effectiveness depends on both input characteristics and learners' cognitive resources (Montero Pérez et al., 2014).

Research in classroom contexts further complicates the picture. Montero Pérez et al. (2014) and Montero Pérez (2022) show that subtitles can be successfully integrated into language instruction, but their effectiveness varies depending on implementation and learner proficiency. Subtitles therefore do not automatically have a uniformly positive or negative effect; rather, their impact is shaped by contextual and learner-related factors. Taken together, the literature indicates that intralingual subtitles tend to support vocabulary acquisition, yet their effectiveness is mediated by variables such as subtitle type, learner proficiency, cognitive load, and instructional context.

As a result, there is no consensus on a single “best” approach to using subtitles in language learning. In particular, limited research has focused specifically on A1 learners of Spanish as a foreign language and on incidental vocabulary acquisition immediately following exposure. This highlights the need for further research that not only measures empirical learning outcomes but also situates these outcomes within a theoretically grounded understanding of multimodal processing and cognitive constraints in early-stage L2 acquisition.

1.2 Theoretical perspectives

Because subtitle processing involves both cognitive and learner-related factors, this thesis draws on multiple complementary theoretical frameworks. These theoretical perspectives provide a conceptual foundation for understanding how learners process audiovisual input and how subtitles may influence language learning. In particular, they help explain linguistic input processing, the supportive role of multimodal information in comprehension and retention, and the impact of cognitive overload on learning. The following subsections therefore outline several key theories that together inform the analysis of subtitle-supported language learning.

1.2.1 Input Theory

Krashen's (1985) Input Hypothesis emphasizes the central role of meaningful and comprehensible input in language acquisition. It constitutes one of the main principles of the Natural Approach. This approach is based on the idea that the acquirer or learner must comprehend messages, since receptive acquisition is the foundation of productive ability. Language development occurs when learners are exposed to input slightly beyond their current proficiency level but interpretable through contextual and linguistic cues. The definition given by Abbott and Wingard (1981:15) reflects well on what is being understood with 'input':

"The input is a verbal experience, visual or auditory, which may consist of language in use, as when listening to someone during a conversation; or of deliberately selected usage, as when reading the examples preceeding an exercises [sic]."

This definition underlines how context plays a crucial role in understanding and providing intake, which Chaudron (1985) describes as the learner's processing of input between perception and acquisition. Subtitle-supported audiovisual materials may increase input comprehensibility by making the linguistic signal clearer and more accessible, therefore encouraging learners to derive meaning from contextual information and linguistic form.

1.2.2 Dual-Coding Theory

For beginners, intralingual subtitles can serve as a visual reinforcement of the auditory input, helping learners better understand what they hear while simultaneously acquiring new vocabulary. This assumption is grounded in Dual-Coding Theory (Paivio, 1986), which posits that information is processed through partially independent but interconnected verbal and non-verbal systems. When auditory input is accompanied by written text, learners can encode linguistic information through both channels, potentially resulting in more robust memory representations. In the context of subtitles, the combination of spoken language and on-screen text may facilitate comprehension and strengthen form–meaning mappings, thereby supporting incidental vocabulary acquisition.

1.2.3 The Theory of Multimedia Learning

Another theory grounded in Dual-Coding Theory is Mayer's (2002) Theory of Multimedia Learning. This theory argues that learning is optimized when learners actively select, organize, and integrate information from multiple modalities, such as words and images. In subtitle-supported videos, learners integrate auditory input with visual-verbal information in order to construct a coherent mental representation. This integration process may promote deeper processing and enhance vocabulary acquisition (Bensalem, 2018; Bird & Williams, 2002; Frumuselu et al., 2015). However, attentional allocation and the ability to handle several information streams at once are necessary for successful integration.

1.2.4 Cognitive load theory & Linguistic Influence

At the same time, Cognitive Load Theory (Sweller, 1988) highlights limitations of working memory capacity and emphasizes that instructional formats can either optimize or overload cognitive resources. Subtitles require learners to read while simultaneously listening and processing visual information, which may increase processing demands. Peters and Webb (2018) suggest that subtitles can impose additional cognitive load, particularly for learners at lower proficiency levels who may not yet process written L2 input automatically. When working memory capacity is exceeded, learning may be hindered rather than facilitated. Thus, while intralingual subtitles can support comprehension and lexical access, they may also

introduce additional demands that affect beginning learners differently than more advanced learners.

This thesis also draws on research into learner-related moderating factors, specifically cross-linguistic influence and prior language knowledge. Previous knowledge of another language can facilitate vocabulary learning through shared lexical, phonological, or morphological features, with stronger effects observed when languages are typologically related, as shown for closely related language pairs such as French and Spanish (Bartolotti, 2016; Flores-Salgado & Gutiérrez-Koyoc, 2024). Prior language knowledge is therefore expected to moderate how subtitle-supported input is processed and how effectively it contributes to vocabulary development, providing an explanatory framework for individual differences among beginning learners.

1.3 Incidental Vocabulary Learning Through Subtitles

Building on the theoretical perspectives discussed above, this section focuses on incidental vocabulary learning through subtitle-supported audiovisual materials in an educational context. More specifically, it explores how intralingual subtitles may contribute to the incidental acquisition of new vocabulary and how different types of subtitles influence this learning process in the classroom.

1.3.1 Incidental language learning

In an L2 classroom context, there are two primary ways to foster language learning: intentional and incidental learning. In intentional learning, acquisition most often refers to activities, where the focus lies on actively memorizing different facets of a language, such as grammar and vocabulary (Frumuselu et al., 2015). In contrast, incidental vocabulary learning refers to situations in which learners are not explicitly instructed to memorize specific words but may nevertheless acquire new lexical items as a by-product of meaning-focused exposure. This method represents a more unconscious way of learning a foreign language (Frumuselu et al., 2015). In the context of incidental vocabulary learning, learners may encounter new lexical items visually, aurally, or through both modalities, and may unconsciously retain some

of these items in receptive, productive, or oral form. Incidental vocabulary learning is widely acknowledged as an important means of acquiring a new language (Bensalem, 2018; Danan, 2004). For A1 learners, whose lexical knowledge remains limited, such exposure may play a crucial role in early vocabulary development.

1.3.2 Types of subtitles: interlingual and intralingual

Different types of subtitles can be implemented in a classroom context. The two most commonly used categories are interlingual subtitles (spoken in the foreign language while the subtitles are in the native language: L2 and L1) and intralingual subtitles or captions (both the spoken and written forms are in the foreign language: L2 and L2) (Montero Pérez et al., 2014). Interlingual subtitling also includes another form: reversed subtitles. In this format, the spoken language is in the native language while the subtitles are in the foreign language (L1 and L2) (Bairstow and Lavour, 2012; Montero Pérez et al., 2014).

Subtitles, regardless of their type, can support second language vocabulary acquisition, though the nature of the gains varies depending on the viewing mode (Danan, 2004; Van der Kolk & Feijoo, 2024). Vanderplank (1993) suggests that because intralingual subtitles clearly indicate the boundaries between words in the speech stream and are unaffected by accent variations and sound degradation, which can have a negative impact on auditory comprehension, they may make it easier to access the target language and improve comprehension and learning.

Empirical research investigating the impact of subtitles on language learning has shown mixed but generally positive findings. An early meta-analysis (Montero Pérez et al., 2014) demonstrated that subtitled videos, particularly intralingual subtitles, tend to have a beneficial effect on vocabulary acquisition. These findings suggest that subtitles help learners segment spoken language and relate sounds to their written forms, as predicted by Paivio's (1986) Dual-Coding Theory. Reynolds et al. (2022) similarly show that videos with subtitles generally result in greater vocabulary gains than videos without text, with intralingual subtitles having the strongest effects. Frumuselu et al. (2015) report that intralingual subtitles are more effective than interlingual subtitles for acquiring new vocabulary, likely because

subtitles in the target language create stronger alignment between spoken and written input, thereby enhancing incidental learning. More recent meta-analytic work by Kurokawa et al. (2024) supports the overall positive effect of L2 subtitles on incidental vocabulary learning, while also identifying moderating factors such as learner level and assessment method. All the results from the different studies are in line with the so-called “subtitle principle” which states that L2 captions tend to yield positive learning outcomes when auditory input is supported by written text (Kurokawa et al., 2024).

1.3.3 Intralingual subtitles for educational purposes

Subtitles in audiovisual input have been widely discussed in second language acquisition research, particularly in relation to vocabulary learning. However, their educational value appears to depend on different theoretical perspectives and contextual factors. The following section therefore outlines the main arguments in favour of subtitle use, as well as the main concerns raised in the literature.

Research on the impact of subtitles on language learning has shown generally positive findings (Montero Pérez et al., 2014), resulting in the recommendation that intralingual subtitles should be used strategically in a classroom context since they encourage reflective attention to both oral and written content (Danan, 2004). The different theoretical perspectives discussed above support this educational strategy. For example, from a second language acquisition perspective, Krashen’s Input Hypothesis (1985) emphasizes the importance of meaningful and comprehensible input. Language development occurs when learners are exposed to input that is slightly beyond their current proficiency level but still interpretable through contextual and linguistic cues. Subtitle-supported audiovisual materials may increase input comprehensibility by making the linguistic signal clearer and more accessible, as theorized by the dual-coding theory (Paivio, 1986) and the multimedia learning theory (Mayer’s, 2002).

However, some researchers advocating a “no subtitles” approach argue that learners may be encouraged to focus entirely on the auditory input when subtitles are absent, potentially leading to improved listening skills and more natural language acquisition (Krashen, 1985).

From this viewpoint, subtitles might reduce the necessity to process auditory input independently, possibly hindering the development of listening skills. This tension has contributed to an ongoing debate in the literature regarding whether subtitles function primarily as supportive foundation or as a potential distraction.

Concretely, Cognitive Load Theory highlights limitations of working memory capacity and emphasizes that instructional formats can either optimize or overload cognitive resources. In this sense, subtitles require learners to read while simultaneously listening and processing visual information, which may increase processing demands. Studies suggest that cognitive overload may occur due to the increased demands placed on working memory, as learners must simultaneously read, listen, and interpret visual information (Peters & Webb, 2018). This can affect the relationship between the independent (video condition) and dependent variables (incidental vocabulary acquisition) through cognitive load as a theoretical mediator (Baarda & Van Vianen, 2025), potentially resulting in a negative effect. Peters and Webb (2018) further suggest that subtitles can impose additional cognitive load, particularly for lower-proficiency learners who may not yet process written L2 input automatically. When working memory capacity is exceeded, learning may be hindered rather than facilitated. Thus, while intralingual subtitles can support comprehension and lexical access, they may also introduce additional demands that affect beginning learners differently from more advanced learners.

Taken together, the existing literature indicates that subtitles, particularly intralingual subtitles, tend to support vocabulary acquisition. However, the effectiveness of subtitles appears to be mediated by factors such as subtitle type, learner proficiency, cognitive load, and instructional context. Consequently, there is no clear consensus on a single “best” approach to using subtitles in language learning, which highlights the need for further research that considers both empirical outcomes and practical classroom applications. However, most of these research focus on English as a foreign language, while relatively little attention has been paid to classroom contexts involving languages other than English or to contexts in which the foreign language is taught for specific purposes, such as business language rather than everyday language use (Black, 2021; Wei & Fan, 2022). In this regard,

the present thesis aims to gain a better understanding of the potential of intralingual subtitles for incidental vocabulary acquisition among A1–A2 learners of Spanish for specific purposes such as business language, as this area has not yet been examined in previous research (Bensalem, 2018; Reynolds et al., 2022).

1.4 Conceptual Model

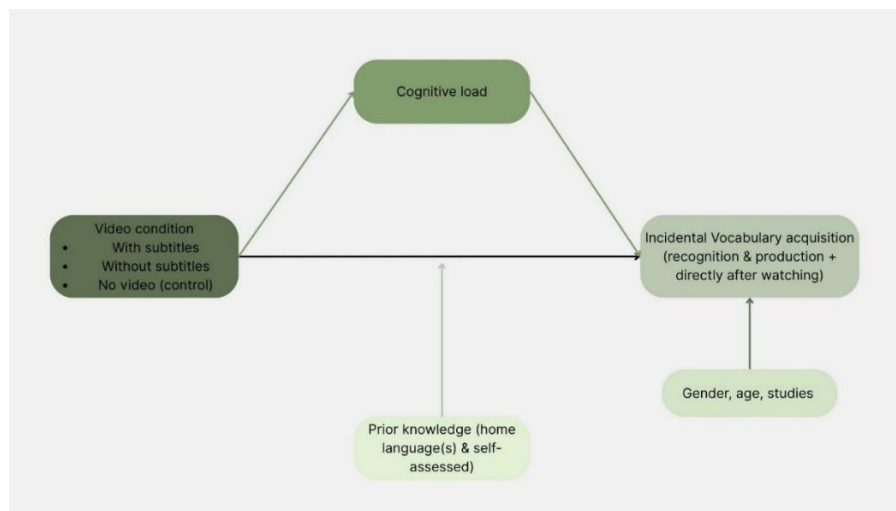


Figure 1: Conceptual Model

The main purpose of this study is to examine whether the presence of intralingual (L2) subtitles in videos affects incidental vocabulary acquisition in beginning learners of Spanish. As discussed in the theoretical framework, multiple studies support the benefits of intralingual subtitles for foreign language vocabulary acquisition (Bensalem, 2018; Montero Pérez et al., 2013; Montero Pérez et al., 2014; Reynolds et al., 2022). Consequently, the video condition (with subtitles, without subtitles, and no video) can be considered the independent variable which influences the dependent variable, namely incidental vocabulary acquisition. The dependent variable represents the outcome being measured in the study, whereas the independent variable refers to the factor that may affect this outcome (Baarda & Van Vianen, 2025).

Studies suggest that cognitive overload may occur due to increased processing demands placed on working memory, as learners must simultaneously read, listen to, and interpret visual information (Peters & Webb, 2018). This may affect the relationship between the

independent and dependent variable through cognitive load as a theoretical mediator, potentially resulting in a negative effect.

Although Kurokawa et al. (2024) support the positive influence of subtitles on vocabulary learning, they also identify prior knowledge of other languages as a moderator of this positive effect. This argument is further supported by Montero Pérez (2022) and therefore must be taken into consideration, as some students may previously have been exposed to Romance languages, such as French or Italian, to a greater extent, which could facilitate the acquisition of new Spanish vocabulary.

Another point to consider is that learners with prior knowledge of multiple languages, regardless of whether these belong to the Romance language family, tend to acquire an additional language more rapidly (Blumenfeld et al., 2016). Bartolotti and Marian (2016) also demonstrate empirical evidence that prior language knowledge influences how effectively learners acquire a new language through cross-linguistic influence. Learners who speak typologically related languages may benefit from similarities such as shared cognates, morphological patterns, and semantic associations that facilitate lexical processing. In addition, multilingual learners with structurally distant language repertoires may develop enhanced metalinguistic awareness that supports vocabulary inference and incidental vocabulary acquisition.

The control variables used in this study are qualitative in nature, as they allow variables to be classified into categories (Vandervieren, 2025). These variables are mainly used to create groups within the data and are therefore often reported rather than directly analysed (Frumuselu et al., 2015; Montero Pérez, 2022; Reynolds et al., 2022). Participants' educational background is not frequently included in such analyses, as studies often involve students from different academic years and fields of study. In contrast, the present sample is more homogeneous, as all participants are Bachelor students in Business and Economics. In addition, all participants are learning the same target language, namely Spanish. Self-assessed sociodemographics were also collected to determine whether participants had prior knowledge of Spanish or of other languages.

Chapter 2

Research Objective and Central Research Question

2.1 Research objective

This study aims to investigate whether the use of intralingual (L2) subtitles in videos significantly contributes to the incidental vocabulary acquisition of A1–A2 learners of Spanish as a foreign language for specific purposes, compared to the complete omission of subtitles. By focusing specifically on beginner learners, this research seeks to determine whether written support in the target language enhances early lexical development or whether it introduces additional processing demands that may limit learning. Understanding this issue has important implications for didactic choices in language education, particularly for learners who are still in the initial stages of language development and whose exposure to authentic L2 input is often limited.

2.2 Central research question

“What is the effect of videos with L2 subtitles compared to videos without subtitles on the incidental vocabulary acquisition of A1-A learners of Spanish as a foreign language for specific purposes?”

This question forms the central focus of the study. Prior research has shown that subtitles can enhance lexical learning in multimedia input, as they provide learners with simultaneous auditory and textual cues that facilitate noticing and lexical retention (Montero Pérez et al., 2014). Studies on multimodal input further argue that combining visual and auditory information promotes deeper engagement with linguistic forms, which is essential for vocabulary learning among beginning learners (Bensalem, 2018).

From a theoretical perspective, the potential benefits of subtitles can be interpreted through the Input Hypothesis (Krashen, 1985), Dual-Coding Theory (Paivio, 1986), and Multimedia Learning Theory (Mayer, 2002). These frameworks could provide a possible explanation on

how subtitles may support comprehension and retention by making input more accessible and by allowing learners to process information through both auditory and visual channels.

In line with the so-called “subtitle principle,” which reported that L2 captions tend to yield positive learning outcomes when auditory input is supported by written text (Kurokawa et al., 2024). Based on these findings, L2 subtitles may contribute to incidental vocabulary acquisition compared to viewing videos without subtitles.

However, Cognitive Load Theory (Sweller, 1988) suggests that these benefits may depend on learners’ ability to manage the simultaneous processing of auditory and written input, which may be particularly demanding for lower-proficiency learners such as A1–A2 students.

Hence, the effect of subtitles on incidental vocabulary acquisition vary depending on learners’ cognitive resources and proficiency level. This study therefore investigates whether the positive effects suggested by multimedia and dual-coding theories outweigh the potential cognitive challenges identified by Cognitive Load Theory in the specific context of beginner learners of Spanish for specific purposes.

2.3 Subquestions

2.3.1 Subquestion 1: Learner Background Variables

“To what extent do sociodemographic variables (e.g., home language(s), and self-reported language knowledge) influence the effect of L2 subtitles on incidental vocabulary acquisition?”

As this thesis investigates incidental vocabulary acquisition, the inclusion of this subquestion is grounded in the relevance of lexical learning processes and individual learner variability. Prior linguistic knowledge and background characteristics can substantially influence how learners process subtitled input (Montero Pérez et al., 2014). Research on second language vocabulary acquisition further suggests that existing language knowledge and linguistic environment shape the ease with which learners acquire new vocabulary (Bartolotti & Marian, 2016; Bensalem, 2018; Blumenfeld et al., 2016).

This subquestion is most directly linked to the main research question, as it examines how sociodemographic variables influence the effect of L2 subtitles on vocabulary development.

These background factors therefore provide essential context for interpreting vocabulary gains and may help explain differences between learners exposed to the same instructional input.

2.3.2 Subquestion 2: Comprehension and short-term retention

“Do background variables, such as comprehension scores, explain differences in short-term vocabulary retention?”

Recent evidence suggests that individual cognitive learner characteristics may moderate the effectiveness of subtitles, indicating that comprehension ability could influence how effectively learners notice and retain new lexical items (Kurokawa et al., 2024). Montero Pérez et al. (2014) report that learners with stronger video comprehension benefit more from subtitled input because they are better able to integrate auditory and textual information. Similarly, Bairstow (2015) notes that subtitles impose processing demands and that learners with stronger comprehension skills are more capable of efficiently integrating information from multiple modalities.

This subquestion therefore examines whether short-term vocabulary retention can be explained not only by subtitle condition, but also by learners’ ability to comprehend and process the audiovisual input effectively.

Chapter 3 Methodology

3.1 Research Type

This study adopts a quantitative experimental research design, as its aim is to examine the effect of intralingual (L2) subtitles in videos on the incidental vocabulary acquisition of A1-A2 learners of Spanish as a foreign language for specific purposes. The experimental nature of the study lies in the comparison between different video conditions, namely videos with intralingual (L2) subtitles, videos without subtitles, and a control condition without video exposure.

3.2 Data Collection

The data used in this study were collected by Professor Vangehuchten and Professor Basanta during the academic year 2024–2025. Data collection was based on five short videos (approximately three minutes each) taken from the coursebook *Meta Profesional Plus 1* (Díaz Gutiérrez et al., 2023), which were shown during class time: two videos in the first semester and three in the second semester. Showing each video together with its corresponding test took approximately 15 minutes of class time.

Regarding the meta-data, these were collected during the first two testing moments in the first semester through a brief survey. The participants consisted of students of A1-A2 Spanish for Business and Economics from the Faculty of Business and Economics at the University of Antwerp. Both the tests and the survey were administered through Qualtrics, and the data were exported and collected in an Excel file.

After data collection, the participant data were consolidated and cleaned in Excel, retaining only the relevant responses. Students who had been assigned to two different conditions were removed from the dataset, and responses from students for whom no meta-data were registered were likewise excluded from the study. This decision was made because the absence of meta-data made it impossible to account for the potential influence of learner-

related background variables on the results. Consequently, only students for whom complete meta-data were available were retained in the final dataset for analysis.

3.3 Research Units

Participants were A1–A2 learners of Spanish as a foreign language for specific purposes in Business and Economics from the Faculty of Business and Economics at the University of Antwerp. The students were divided into two conditions. Group 1 (L2 subtitles) watched videos in Spanish with Spanish subtitles, whereas Group 2 (without subtitles) watched the same videos but without subtitles.

A total of 172 participants took part in the study (mean age: 19). They were randomly assigned to the two conditions: Group 1 ($n = 88$) and Group 2 ($n = 84$). All students followed the same standard Spanish curriculum and were exposed to the same vocabulary content. The difference between the groups lies in the additional video exposure: Group 1 viewed videos with subtitles, whereas Group 2 viewed the videos without subtitles.

Tables 1 to 4 provide an overview of the demographic characteristics of the participants and the distribution across conditions. Table 1 presents the gender distribution of the sample, showing a relatively balanced representation of male and female participants, with a small number of participants for whom gender was not reported. Table 2 outlines the participants' fields of study, with Business Management representing the largest group, followed by Social and Economic Sciences and Commercial Engineering. Table 3 shows the distribution of participants across the two experimental conditions, indicating an almost equal division between the condition with subtitles and the condition without subtitles. Finally, Table 4 combines gender and condition, illustrating how the gender distribution was spread across both experimental groups. Overall, the tables indicate a relatively balanced sample across conditions while also providing insight into the demographic composition of the participants.

Table 1: Gender distribution

Gender	N	%
--------	---	---

Female (F)	75	43.6
Male (M)	83	48.3
Not mentioned (NA)	14	7.0

Table 2: Distribution by study programme

Studies	N	%
Business Management	82	47.7
Economic Policy	8	4.7
Commercial Engineer	36	20.9
Political Sciences	1	0.6
Social and Economic Sciences	41	24.4
Communication Sciences	2	1.2
NA	1	0.6

Table 3: Distribution by condition

Condition	N	%
With subtitles	88	51.2
Without subtitles	84	48.8

Table 4: Gender Distribution per Condition

Condition \ Gender	Gender		
	Female	Male	NA
With	32 (36.4%)	49 (55.7%)	7 (7.9%)
Without	43 (51.2%)	34 (40.5%)	7 (8.2%)

3.4 Measuring Instruments

Short videos in Spanish (approximately three minutes), adapted to the A1–A2 level, were used and included a number of unknown lexical items (ranging from 12 in the first video, 29 for the second video, 30 for the third video, 32 for the fourth video, and 37 for the fifth video). The videos were selected based on their suitability for beginner learners. Group 1 and Group 2 each watched the same video once, with Group 1 viewing the video with intralingual (L2) subtitles and Group 2 watching it without subtitles.

The videos were selected from *Las etapas de Hanna*, part of *Meta Profesional Plus 1* (Díaz Gutiérrez et al., 2023), which is also the coursebook used in class. By cross-referencing the vocabulary used in the videos with the lexical lists from the coursebook, only genuinely new words were selected for testing. This allows for a controlled measurement of incidental vocabulary acquisition, as prior classroom exposure to these lexical items can be excluded.

After viewing the video, participants in Group 1 and Group 2 completed a brief survey on their linguistic background and demographic characteristics during the first two testing moments, followed by a test on vocabulary retention composed of two parts: a first part assessing incidental word recall and a second part assessing word production (see Appendix 2) of the newly introduced vocabulary.

For the receptive part of the test, students selected the correct translation from three different options. For the productive part, they filled in missing words, where the first letter of the target word was provided together with the translation.

Table 5: example of one question

Example of a recall question: Yo _____ mucho a Berlín. (reizen; to travel) apellido (1) ; viajo (2) ;congreso (3)	Example of a production question: Hannah estudia P_____ (reclame; advertising) (publicidad)
---	--

The students also completed comprehension questions. In the first test, these questions consisted of open-ended answers in English, whereas in the following four tests they consisted of true/false questions.

3.5 Data-analysis

3.5.1 Phase 1: Codebook & IRR

For the data-analysis of the production part, the observations from Group 1 and Group 2 ($n = 172$) were first categorized and adapted for analysis in RStudio. Since the productive vocabulary questions consisted of open-ended answers, a codebook (see Appendix 1) was developed specifically for this part of the test. A codebook was not necessary for the recall and comprehension questions, as these had fixed-answer formats.

The codebook was developed through an iterative process in which five categories of answer possibilities were established: error-free, almost correct, synonym, incorrect and no answer. The category ‘synonym’ was included for a few reasons: 1) students with prior knowledge may also have a broader vocabulary repertoire, and giving a synonym could therefore show more mastery of the language, and 2) a synonym is not always an incorrect answer depending on the question. The student’s answer being a synonym could, depending on the context used, be seen as equal to the correct answer. Meaning that placing it under “incorrect” would not always be appropriate. One thing that has to be noted is that for production the questions were a C-test type, meaning that the first letter of the correct answer was given. Thus,

producing a synonym indicated that the student did not know the specific target word, but was still able to provide an appropriate alternative within the context of the phrase. Responses that were phonetically similar to the target word but not fully correct, for instance, may reflect partial lexical acquisition and therefore provide additional insight into the underlying learning process, an aspect that the sum of the scores alone cannot capture. Hence why these types of answers would be inserted into the 'almost correct' category, because they have some kind of resemblance with the correct answer, showing that the students know the word but don't know how to write it yet.

As mentioned in the first phase, before consolidating the Excel file into a format fit for RStudio, the finalized codebook was applied to the dataset. A binary coding system, also called dummy variables, was then used. Dummy variables are used to make categorical variables usable in quantitative research (De Maeyer et al., 2012). Usually, they are used to include categorical variables in statistical analyses as independent variables (De Maeyer et al., 2012). In the case of this thesis, the five response categories function as dummy variables for the dependent variable vocabulary acquisition, because the purpose is to examine the level of accuracy in the productive acquisition of the new lexicon.

Once the codebook was approved, the first binary codification of the dataset was performed. Subsequently, a new Excel file was created to generate a randomized sample of 30% of the data ($n = 50$) for a second coder, in order to assess inter-rater reliability (IRR) using Cohen's Kappa (κ) in SPSS 31. This is in line with methodological recommendations suggesting that at least 20% of the data should be used for reliability checks in large datasets (Lombard, Snyder-Duch & Bracken, 2002). Cohen's Kappa (κ) was used, as it is considered an appropriate statistic for categorical coding. The IRR was only necessary for the production questions, as these were the only questions requiring categorization into the five coding categories.

3.5.2 Phase 2: Statistical analysis

In the second phase, the adapted codification of the full dataset was statistically analysed using the software R-studio (version 2026.01.0-392). Four parts of the data were analysed: the first one looked at raw scores for recall and production score, the second at specific modalities of the production answers, the third part visualised the effect of prior knowledge

and cross-linguistic influence, and the last part looked specifically at comprehension scores. Each part used a linear regression models (LMM), taking into account the longitudinal nature of the data whenever necessary. LMM are very typical to use in regression analysis as they help to discern variables that have an influence on the effect of other variables on the dependent variable (De Maeyer et al., 2012).

A linear mixed model was fitted in RStudio for several reasons. First, the data contain repeated measures, as each student completed five vocabulary tests throughout the academic year. Second, it is expected that some students may differ in their initial vocabulary knowledge through prior knowledge of Spanish, meaning that individual starting levels may vary considerably. Third, the dataset may contain an unequal number of observations, since some students may have missed individual test administrations. Linear mixed models are robust to such missing data and unbalanced designs, unlike traditional repeated-measures ANOVA, which requires complete datasets and assumes equal group sizes. Finally, the primary focus of the analysis is to estimate a linear time effect while repeated-measures ANOVA treats them as categories.

3.5.3 Recall and Production – raw scores

In this analysis, the model included time (test number), condition (intralingual subtitles versus no subtitles), and the interaction between both as fixed effects. These variables were examined for the answers on both the recall and production questions. Random intercepts for students were included to account for individual differences in baseline vocabulary knowledge.

The dependent variable was the raw vocabulary score, representing the number of correctly produced target words per test. Although this variable is right-skewed and consisted of discrete integer values, an approximation with a linear mixed model based on the principles of linear regression was the most appropriate analytical choice for this thesis (De Maeyer et al., 2012). In addition, the interaction between time and score was not linear, while still remaining acceptable with the principles of linear regression. The main aim of this model was to determine whether there were significant differences in vocabulary development through

time effect between students who watched videos with intralingual subtitles (Group 1) and those who watched the same videos without subtitles (Group 2).

For this reason, the null model was first based solely on the raw scores of both question types. Given the longitudinal and nested structure of the data, with repeated measurements per student across five testing phases, a multilevel approach was adopted by design, following the recommendations of De Maeyer et al. (2012). An intercept-only null model was first fitted to quantify the degree of clustering, expressed as the intraclass correlation coefficient (ICC). This provides a baseline against which to interpret how much variance the fixed effects explain. The substantive model of interest includes time, condition, and their interaction as fixed effects, as this directly addresses the main research hypothesis. For transparency, intermediate models adding time and condition separately are also reported.

This is a two-level model, as the production and recall items per test were aggregated into a single raw score rather than analysed individually. This simplification was made to facilitate the analysis. In its full form, the data has a three-level structure: individual items are nested within test occasions, which are in turn nested within students ($n = 172$).

3.5.4 Detailed on Production

A second analysis was conducted using the same linear mixed model (LMM) structure, this time on a sum score of the individual binary items. Although a general linear mixed model (GLMM) with binomial family would be statistically more appropriate, a simple LMM structure is still meaningful in the scope of this research. The same rationale as for the raw score analyses was applied.

Rather than analysing total vocabulary scores, this model focused on the type of response produced by each student for the production of individual lexical items, triggered by the first letter that was given. Each response was coded into categories such as error-free, almost correct, synonyms, incorrect, or no response. For the purposes of the statistical analysis, these categories were converted into dummy variables (0/1), indicating whether a particular response type occurred or not (see codebook, appendix 1). Afterwards, for the facilitation of

this master's thesis, each category was also assigned its own score, where 1 represented error-free responses and 5 represented no response.

This focus allows for a more fine-grained examination of vocabulary acquisition. Whereas the previous linear mixed model evaluated overall vocabulary growth, this model examined how the quality of individual word production changed over time.

3.5.5 Comprehension Scores

For the fourth analysis, the same methodological principles as in the previous three analyses are applied. This analysis introduces comprehension scores as an additional variable in order to examine vocabulary retention at the level of individual students. It should be noted that the comprehension scores from Tests 2 to 5 were based on true/false questions and were therefore already binary in nature. Test 1, however, consisted of open-ended responses and therefore required manual binary coding. A broad acceptability criterion was applied, meaning that any response demonstrating a correct understanding of the content was coded as 1, regardless of the language in which the response was written, while incorrect or missing responses were coded as 0.

The aim of this analysis is to address the second sub-question of this thesis, namely how comprehension scores may account for differences in short-term vocabulary retention. Recent evidence suggests that individual learner characteristics can moderate the effectiveness of subtitles, implying that comprehension ability may play a role in the extent to which learners notice and retain new lexical items (Kurokawa et al., 2024). Similarly, Montero Pérez et al. (2014) report that learners with higher levels of video comprehension derive greater benefit from subtitled input, as they are better able to integrate auditory and textual information.

Based on this theoretical and empirical background, the present model was implemented to examine whether comprehension performance is associated with differences in vocabulary retention outcomes within the dataset.

3.5.6 Prior Knowledge and Cross-linguistic Influence on New Vocabulary Retention

For the third analysis, the sociodemographics collected from each student were combined with the primary dataset, namely the raw scores. Within the meta-data, two variables were of particular interest: prior knowledge of Spanish and cross-linguistic influence derived from the students' different known (native) language(s). These variables are based on self-assessment by the students rather than on a standardized assessment of their proficiency in the reported languages.

These two variables were operationalised differently. Prior knowledge of Spanish was coded as a binary variable, distinguishing students who had had any meaningful contact with Spanish through formal education or exchange from those without such prior exposure. Brief or incidental contact, such as using a language-learning application for a few months, was coded as absent (0). This binary variable was included as a covariate in the first analysis of raw scores.

For the CLI analysis, categorical responses were recoded into four mutually exclusive groups. The first group consists of students with some form of prior knowledge of Spanish, either through family exposure, informal contact such as songs or language-learning applications, or formal education. The second group includes students without prior knowledge of Spanish but with a Romance language background, meaning they use French, Portuguese, or another Romance language in interactions with family or peers, either as a native language or as an additional language. The third group consists of students without prior knowledge of Spanish who are bi- or multilingual in non-Romance languages. The fourth and final group includes students who primarily speak Dutch and who do not belong to any of the previous categories, serving as the reference category in the model.

Groups 2 and 3 reflect cross-linguistic influence, referring to the role of previously acquired languages in facilitating the acquisition of new vocabulary, whereas Group 1 reflects direct prior knowledge of the target language. The distinction between Romance and non-Romance cross-linguistic influence is based on research suggesting that typological proximity between languages strengthens cross-linguistic transfer effects, with Romance languages being structurally closer to Spanish than non-Romance languages (Bartolotti, 2016; Flores-Salgado

& Gutiérrez-Koyoc, 2024). For the analysis, the Spanish L1 group ($n = 5$) was excluded from the CLI model, as a group of this size does not meet the minimum threshold for reliable dummy variable estimation in an LMM (approximately 30 observations per level). These students were retained in all other analyses.

This modelling approach was employed to address the sub-question concerning the influence of learner background variables on students' production and retention of new vocabulary. Using the same linear mixed-effects modelling framework as in previous analyses, these additional predictors were incorporated to compare students with prior knowledge of Spanish or Romance-language cross-linguistic influence to those without such advantages.

3.6 Ethical Considerations

The survey and tests for Group 1 and Group 2 were administered via Qualtrics. The first question in the survey also served as the Informed Consent Form (see Appendix 2). Permission for this research was granted by the ethical commission on September 24th, 2024. The collected data will be stored on UA servers for a period of ten years. The researchers will treat the data confidentially and use it solely for scientific purposes.

Chapter 4 Results

4.1 Introduction

Before presenting the results of each analysis, Table 6 provides an overview of the descriptive statistics for all outcome measures. All analyses are based on N = 172 students, following the exclusion of participants with missing metadata and non-binary participants due to insufficient cell size.

Table 6: Overview of the Descriptives for all outcome measures

Measure	N Students	Observations	Mean	SD	Min	Max
Recall	172	591	3.91	1.28	0	5
Production (raw)	172	590	1.59	1.43	0	5
Production (detailed)	172	2940	3.32	1.63	1	5
Comprehension	172	590	3.48	1.42	0	7

4.2 Inter Rater Reliability (IRR) of the Codebook

To ensure the reliability of the coding process, inter-rater reliability (IRR) was measured using Cohen's Kappa. This codebook was only used for the productive part of this analysis, and not for the recall or comprehension questions. Since each of the five tests consisted of five questions that were scored across five coding categories (error-free, almost correct, incorrect, and nothing filled in) through a binary coding system (0 or 1), Cohen's Kappa was calculated separately for each question and category rather than across all tests at once. Combining all items into one overall calculation could have merged different coding decisions and therefore given a less accurate representation of agreement.

For the global IRR, a conservative approach was used. The lowest Kappa value found across all categories within each individual test was taken as the IRR estimate for that test, based on the idea that the overall reliability of a test cannot be stronger than its weakest point. Applying this method across the five tests resulted in a global IRR of $\kappa = .481$. Although this formally falls within the moderate agreement range, this score was almost entirely influenced by one outlying item, namely Question 3 of the “half correct” category in Test 2 ($\kappa = .481$). All other items showed substantial to near-perfect agreement between raters.

More specifically, Tests 1, 3, 4, and 5 had minimum Kappa values of .857, .892, .846, and .876 respectively, while the large majority of individual items reached $\kappa = 1.00$. Despite the single lower score in Test 2, inter-rater agreement across the full dataset was therefore consistently high. This indicates that the coding scheme was applied reliably and that the binary classifications reflected a shared and stable interpretation of the scoring criteria.

4.3 Results Recall and Production – raw scores

4.3.1 Recall scores from R

A linear mixed model was fitted to examine the effects of test and condition on recall raw scores, with random intercepts for students. Table 7 presents the fixed effects estimates.

Table 7: Fixed Effects Estimates of Recall

Predictor	β	SE	t	p
Intercept	4.31	0.15	28.72	< .001
Test	-0.17	0.05	-3.68	< .001
Condition	0.65	0.21	3.05	.002
Test x Condition	-.021	0.06	-3.37	< .001

Results revealed a significant main effect of test ($\beta = -0.17$, $t(477) = -3.68$, $p < .001$), indicating that recall scores decreased over time in the condition with subtitles. A significant main effect of condition was also found ($\beta = 0.65$, $t(563) = 3.05$, $p = .002$), with the 'Without' condition showing higher scores at baseline. The interaction between test and condition was likewise

significant ($\beta = -0.21$, $t(473) = -3.37$, $p < .001$), suggesting that the rate of decline over time is stronger in the absence of subtitles.

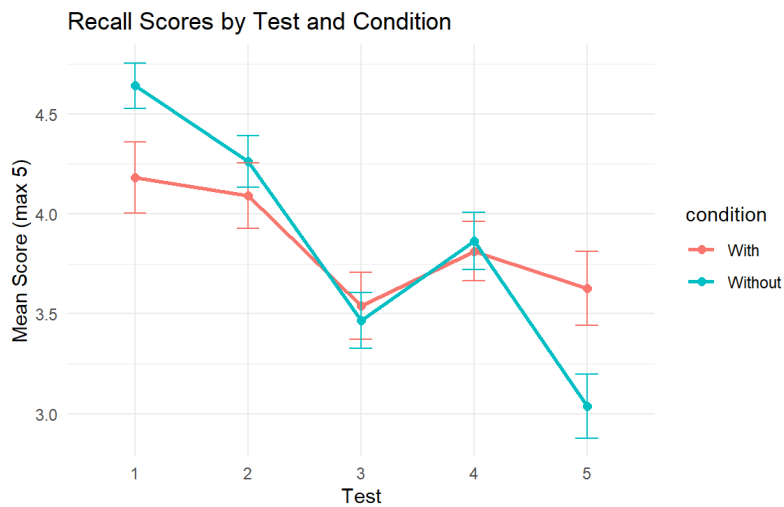


Figure 1: Recall Scores by Test and Condition

Figure 1 displays mean recall scores per test and condition over time. Both conditions start at a relatively high level and follow a broadly similar downward trend across the five tests, with a slight recovery at test 4 before declining again at test 5. The 'Without' condition begins marginally higher at test 1 but ends notably lower at test 5 (approximately 3.05 vs. 3.65), suggesting a steeper rate of decline, consistent with the significant test $\times$ condition interaction found in the model. The lines run closely together across the middle time points, with the divergence becoming most visible at test 5.

Something that needs to be noted for this plot, and the next ones, is that time and results are not linear, yet this research based its model on the principles of linear regression model for the feasibility of this master's thesis scope. The decline isn't constant, but differs from test to test, this is the reason why students' result aren't linear positive nor negative, but vary depending on the test.

4.3.2 Production scores from R

A linear mixed model was fitted to examine the effects of test and condition on production raw scores, with random intercepts for students. Table 8 presents the fixed effects estimates.

Table 8: Fixed Effects Estimates of Production

Predictor	β	SE	t	p
Intercept	2.07	0.17	12.44	< .001
Test	-0.14	0.04	-3.13	.002
Condition	-0.36	0.24	-1.53	.127
Test x Condition	0.02	0.06	0.37	.709

Results showed a significant main effect of test ($\beta = -0.14$, $t(466) = -3.13$, $p = .002$), indicating that production scores declined over time across both conditions. The main effect of condition did not reach significance ($\beta = 0.36$, $t(469) = -1.53$, $p = .127$), suggesting no overall difference in production scores between the two conditions. The interaction between test and condition was likewise non-significant ($\beta = 0.02$, $t(463) = 0.37$, $p = .709$), suggesting that the rate of change over time did not differ between conditions.

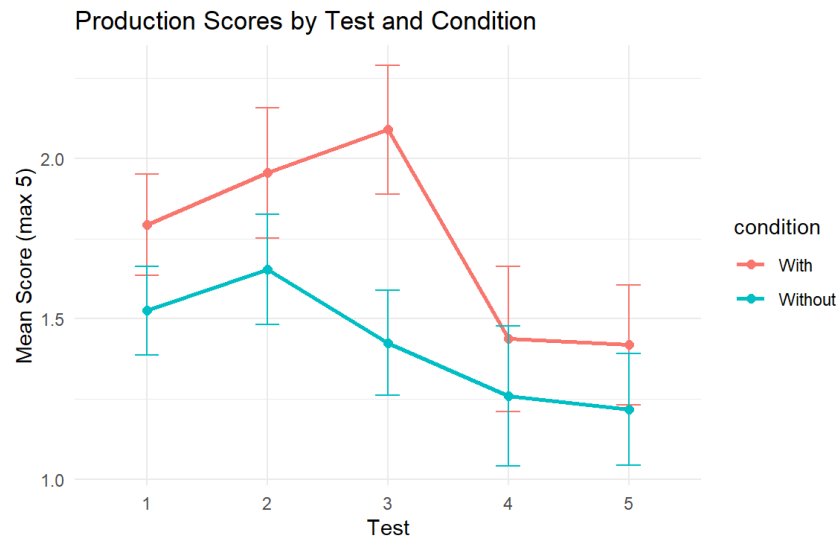


Figure 2: Production Scores by Test and Condition

Figure 2 displays mean production scores per test and condition. The 'With' condition starts higher and peaks at test 3 before dropping notably at tests 4 and 5. The 'Without' condition remains comparatively lower and more stable throughout, showing a gentler decline. Despite this visual difference, neither the main effect of condition nor the interaction term reached

significance in the model, likely reflecting the considerable overlap in confidence intervals across time points.

4.4 Results Detailed production

A linear mixed model was fitted to examine the effects of test and condition on detailed production scores, with random intercepts and slopes for test per student, and random intercepts per item to account for item-level variability. Descriptive statistics indicated a mean score of 3.32 (SD = 1.63) across 172 students and 4325 observations. Table 9 presents the fixed effects estimates.

Table 9: Fixed Effects Estimates of Detailed Production

Predictor	β	SE	t	p
Intercept	3.04	0.18	16.80	< .001
Test	0.09	0.04	2.43	.016
Condition	0.23	0.23	1.14	.256
Test x Condition	-0.02	0.05	-0.37	.713

Results showed a significant main effect of test ($\beta = 0.09$, $t(117) = 2.43$, $p = .016$), indicating that detailed production scores increased slightly over time regardless of condition. Neither the main effect of condition ($\beta = 0.23$, $t(160) = 1.14$, $p = .256$) nor the interaction between test and condition ($\beta = -0.02$, $t(111) = -0.37$, $p = .713$) reached significance, suggesting that the two conditions did not differ in detailed production scores or in their trajectory over time.

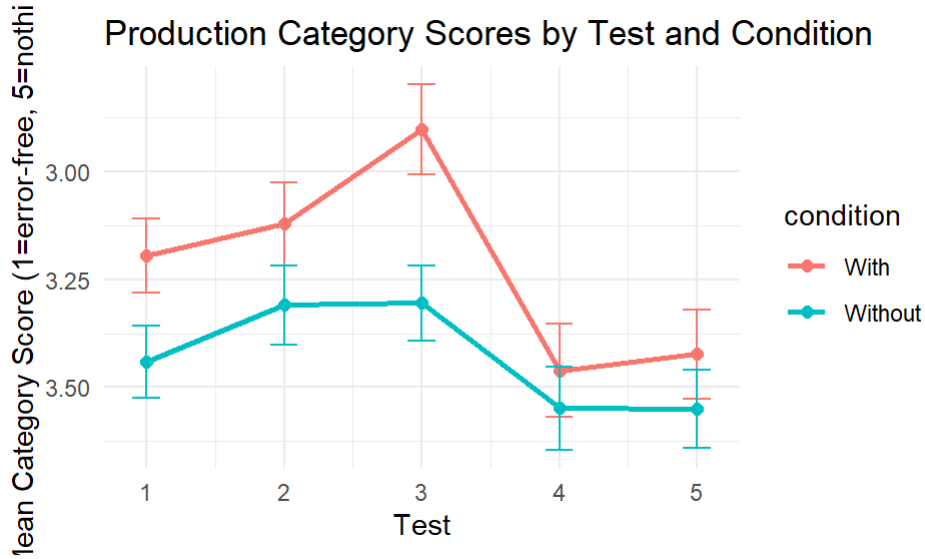


Figure 3: Production Category Scores by Test and Condition

Figure 3 displays mean detailed production category scores per test and condition, where lower values indicate higher accuracy (1 = error-free, 5 = no response). The 'With' condition starts higher in accuracy, peaks at test 3, then drops notably at tests 4 and 5. The 'Without' condition remains comparatively more stable across time points. Despite this visual difference, neither the interaction nor the main effect of condition reached significance in the model, consistent with the considerable overlap in confidence intervals across time points.

A generalised linear mixed model was fitted separately for almost correct and error-free response categories. Table 10 presents the fixed effects estimates for both outcomes.

Table 10: Fixed Effects of GLMM on 'Almost Correct' & 'Error-free'

Predictor	Almost Correct				Error-free			
	β	SE	z	p	β	SE	z	p
Intercept	-1.96	0.28	-6.93	< .001	-0.74	0.24	-3.14	.002
Test	-0.14	0.06	-2.11	.035	-0.17	0.05	-3.53	<.001
Condition	0.08	0.26	0.31	.760	-0.42	0.27	-1.56	.119

TestX	-0.02	0.09	-0.26	.792	0.03	0.07	0.39	.696
Condition								

For almost correct responses, results showed a significant main effect of test ($\beta = -0.14$, $z = -2.11$, $p = .035$), indicating that the likelihood of producing an almost correct response decreased slightly over time. Neither the main effect of condition ($\beta = 0.08$, $z = 0.31$, $p = .760$) nor the test $\times$ condition interaction ($\beta = -0.02$, $z = -0.26$, $p = .792$) reached significance, suggesting that the two conditions did not differ in the rate of almost correct responses or in their change over time.

For error-free responses, the main effect of test was likewise significant ($\beta = -0.17$, $z = -3.53$, $p < .001$), indicating a consistent decline in fully correct responses over time. The main effect of condition ($\beta = -0.42$, $z = -1.56$, $p = .119$) and the test $\times$ condition interaction ($\beta = 0.03$, $z = 0.39$, $p = .696$) were both non-significant, indicating no reliable difference between conditions in error-free production at baseline or in their trajectory over time.

4.5 Results Comprehension scores

A linear mixed model was fitted to examine the effects of test (time) and condition on comprehension scores, with random intercepts for students to account for individual differences. Table 11 presents the fixed effects estimates.

Table 11: Fixed Effects Estimates of Comprehension Scores

Predictor	β	SE	t	p
Intercept	3.18	0.17	18.35	< .001
Test	0.14	0.04	2.61	.009
Condition	-0.06	0.25	-0.24	.813
Test x Condition	-0.05	0.08	-0.66	.508

Results showed a significant main effect of test ($\beta = 0.14$, $t(481) = 2.61$, $p = .009$), indicating that comprehension scores increased over time regardless of condition. The main effect of

condition was not significant ($\beta = -0.06$, $t(577) = -0.24$, $p = .813$), suggesting no overall difference in comprehension scores between the two conditions. The interaction between test and condition was likewise non-significant ($\beta = -0.05$, $t(477) = -0.66$, $p = .508$), indicating that the rate of change over time did not differ between conditions.



Figure 4: Comprehension Scores by Test and Condition

Figure 4 displays the mean comprehension scores per test and condition over time. Both conditions show a generally increasing trend across tests, with a dip at test 3 before recovering and reaching their highest scores at test 5, confirming the significant effect of time found in the model. The two conditions follow largely parallel trajectories, consistent with the non-significant interaction term, though the 'With' condition ends higher at test 5.

4.6 Results Prior Knowledge and Cross-linguistic Influence

4.6.1 Recall and Prior Knowledge

A linear mixed model was fitted to examine the effect of test, condition, and prior knowledge on recall scores across all five tests, with random intercepts per student to account for individual differences. Table 12 presents the fixed effects estimates.

Table 12: Fixed Effects Estimates of Recall & Prior Knowledge as moderator

Predictor	β	SE	t	p
-----------	---------	----	---	---

Intercept	4.78	0.16	30.37	< .001
Test	-0.38	0.04	-8.70	< .001
Condition	-0.55	0.21	-2.59	.010
Prior Knowledge	0.51	0.15	3.48	< .001
Test x Condition	0.21	0.06	3.30	.001

Results revealed a significant main effect of test ($\beta = -0.38$, $t(471) = -8.70$, $p < .001$), indicating that recall scores declined consistently over time regardless of condition. The main effect of condition was also significant ($\beta = -0.55$, $t(552) = -2.59$, $p = .010$), with the 'Without' condition showing higher scores at baseline. Prior knowledge emerged as a significant predictor of recall performance ($\beta = 0.51$, $t(130) = 3.48$, $p < .001$), indicating that students who reported some prior knowledge of Spanish scored consistently higher on recall than those without. The test and condition interaction remained significant ($\beta = 0.21$, $t(474) = 3.30$, $p = .001$), demonstrating that even after taking prior knowledge into account, the difference in recall trajectories between conditions remains. The model explained 13.7% of variance through fixed effects alone (marginal $R^2 = .137$), rising to 33.8% when individual differences were included (conditional $R^2 = .338$). The intraclass correlation coefficient (ICC = .233) indicates that approximately 23% of the total variance in recall scores may be attributed to differences between students.

4.6.2 Production and Prior Knowledge

A linear mixed model was fitted to examine the effect of test, condition, and prior knowledge on production raw scores across all five tests, with random intercepts per student. Table 13 presents the fixed effects estimates.

Table 13: Fixed Effects Estimates of Production & Prior Knowledge as moderator

Predictor	β	SE	t	p
Intercept	1.51	0.18	8.42	< .001

Test	-0.12	0.04	-2.81	.005
Condition	0.45	0.24	1.91	.057
Prior Knowledge	0.65	0.20	3.22	.002
Test x Condition	-0.02	0.06	-0.39	.695

Results showed a significant main effect of test ($\beta = -0.12$, $t(460) = -2.81$, $p = .005$), indicating that production scores declined over time. The main effect of condition did not reach significance ($\beta = 0.45$, $t(423) = 1.91$, $p = .057$), suggesting no reliable overall difference in production scores between conditions. Prior knowledge was a significant predictor of production performance ($\beta = 0.65$, $t(160) = 3.22$, $p = .002$), with students reporting some prior knowledge of Spanish producing consistently higher scores. The test $\times$ condition interaction was non-significant ($\beta = -0.02$, $t(462) = -0.39$, $p = .695$), indicating that the two conditions did not differ in their rate of change over time. The marginal R^2 (.065) indicates that the fixed effects explain a modest proportion of variance, while the conditional R^2 (.528) and ICC (.495) together suggest that individual differences between students account for a substantial share of the total variance in production scores — considerably more so than in recall.

4.6.3 Recall and Cross-linguistic influence

A linear mixed model was fitted to examine the effects of test, condition, and cross-linguistic influence group on recall scores across all five tests, with random intercepts per student to account for individual differences. In this model, the Dutch group serves as the reference category, meaning that the coefficients for Romance, and Multilingual reflect the difference in scores relative to Dutch-only speakers. Table X presents the fixed effects estimates.

Table 14: Fixed Effects Estimates of recall and CLI

Predictor	β	SE	t	p
Intercept	4.95	0.16	30.50	< .001
Test	-0.37	0.04	-8.33	< .001

Condition	-0.69	0.21	-3.23	.001
CLI: Romance	0.37	0.24	1.56	.122
CLI: Multilingual	-0.05	0.15	-0.35	.731
Test x Condition	0.21	0.06	3.27	.001

Results revealed a significant main effect of test ($\beta = -0.37$, $t(434) = -8.33$, $p < .001$), indicating a consistent decline in recall scores over time. The main effect of condition was significant ($\beta = -0.69$, $t(515) = -3.23$, $p = .001$), and the test $\times$ condition interaction was likewise significant ($\beta = 0.21$, $t(444) = 3.27$, $p = .001$), indicating that the rate of decline differed between conditions, consistent with findings from the earlier recall model. The overall effect of CLI group did not reach significance ($\chi^2(3) = 5.82$, $p = .121$), and none of the individual group contrasts differed significantly from the Dutch reference category. The model explained 11.9% of variance through fixed effects (marginal $R^2 = .119$), rising to 32.9% when accounting for individual differences (conditional $R^2 = .329$). The ICC (.238) indicates that approximately 24% of the total variance in recall scores is attributable to differences between students.

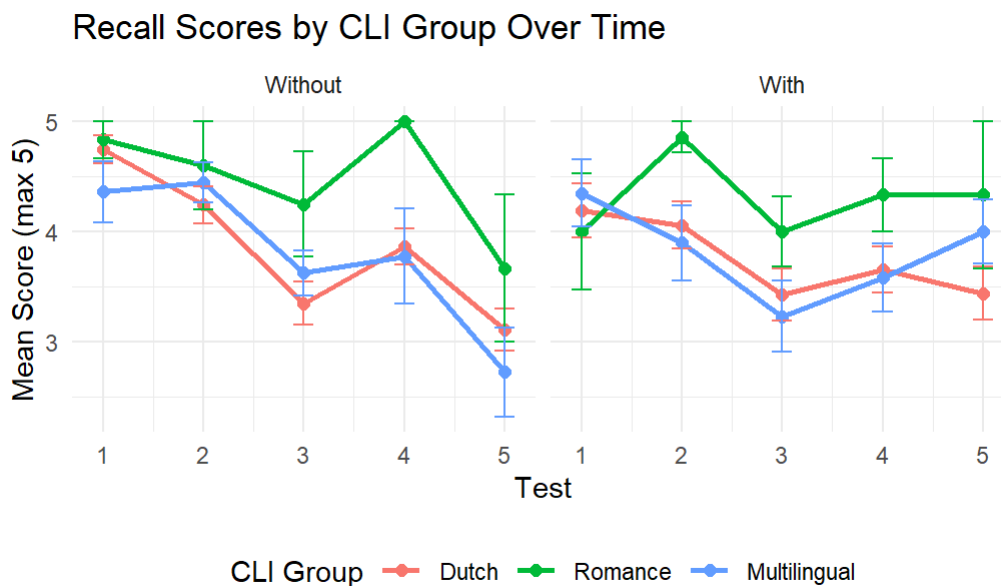


Figure 5: Recall Scores by CLI Group Over Time

Figure 5 displays mean recall scores per test, condition, and CLI group over time. Across both conditions, all three groups follow a broadly similar downward trend, with a partial recovery

at test 4 before declining again at test 5. The Romance group consistently scores at or above the Dutch reference group in both conditions, while the Multilingual group tracks closely alongside Dutch throughout, though it dips somewhat lower at test 5 in the 'Without' condition.

4.6.4 Production and Cross-linguistic influence

A linear mixed model was fitted to examine the effects of test, condition, and cross-linguistic influence group on production scores across all five tests, with random intercepts per student. In this model, the Dutch group serves as the reference category, meaning that the coefficients for Romance, and Multilingual reflect the difference in scores relative to Dutch-only speakers. Table 15 presents the fixed effects estimates.

Table 15: Fixed Effects Estimates of Production & CLI

Predictor	β	SE	t	p
Intercept	1.78	0.19	9.51	< .001
Test	-0.13	0.04	-3.04	.003
Condition	0.16	0.24	0.66	.509
CLI: Romance	0.37	0.32	1.15	.252
CLI: Multilingual	-0.06	0.20	-0.29	.773
Test x Condition	-0.001	0.06	-0.02	.984

Results showed a significant main effect of test ($\beta = -0.13$, $t(427) = -3.04$, $p = .003$), reflecting a gradual decline in production scores over time. Neither the main effect of condition ($\beta = 0.16$, $p = .509$) nor the test $\times$ condition interaction ($p = .984$) reached significance, consistent with the earlier production models. Neither the Romance ($p = .252$) nor the Multilingual group ($p = .773$) differed significantly from Dutch. The marginal R^2 (.029) indicates that fixed effects explain a modest proportion of variance, while the conditional R^2 (.508) and ICC (.493) together confirm that individual differences between students account for a large share of

total variance in production scores, mirroring the pattern observed in the prior knowledge model.

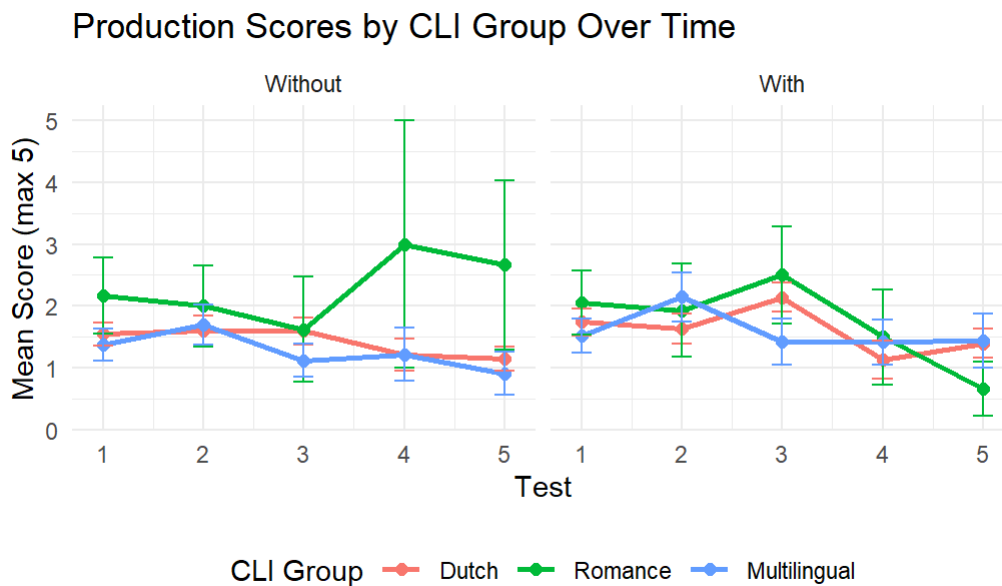


Figure 6: Production Scores by CLI Group Over Time

Figure 6 displays mean production scores per test, condition, and CLI group over time. The Romance group shows the most variable trajectory, particularly in the 'Without' condition, where it starts higher than the other groups and peaks notably at test 4 before declining at test 5; the wide confidence intervals reflect the small group size and should be interpreted with caution. In the 'With' condition, the Romance group similarly starts higher but converges toward the other groups by test 5. The Dutch and Multilingual groups follow closely overlapping trajectories in both conditions, remaining relatively low and stable across all five tests, consistent with non-significant group contrasts for these groups.

5.1 Discussion

5.1.1 Recall

The results showed a significant effect of condition on recall scores, indicating that the subtitle condition influenced learners' receptive vocabulary performance. More specifically, students in the intralingual subtitle condition demonstrated a more stable development over time than students who watched the videos without subtitles. This finding is broadly in line with previous research reporting positive effects of L2 subtitles on vocabulary recognition and recall (Montero Pérez et al., 2014; Reynolds et al., 2022; Kurokawa et al., 2024). These studies suggest that subtitles support noticing of lexical items and facilitate stronger form–meaning connections by combining spoken and written input.

At the same time, it should be noted that the interpretation of this pattern requires some caution, since the group without subtitles started with a higher baseline score than the subtitle group. This initial difference may indicate that the subtitle group experienced a greater processing burden at the beginning of the intervention. From the perspective of Cognitive Load Theory (Sweller, 1988), beginning learners may initially struggle to divide attention between auditory input, visual context, and written subtitles simultaneously. Peters and Webb (2018) similarly argue that subtitles may create additional processing demands, particularly for lower-level learners whose decoding skills in the target language are still developing.

However, despite this lower starting point, the subtitle group showed a less steep decline across the semester and ended with comparatively stronger recall performance. This may suggest that learners gradually adapted to the multimodal format of subtitle-supported input. Once the initial processing demands were reduced, subtitles may have functioned as supportive foundation that enhanced lexical recognition. This interpretation is consistent with Dual-Coding Theory (Paivio, 1986) and Multimedia Learning Theory (Mayer, 2002), which

propose that information presented through multiple channels can strengthen memory when learners are able to integrate the input successfully. Bird and Williams (2002) likewise found that bimodal input (audio + text) enhanced explicit memory for unfamiliar words compared with unimodal presentation.

In other words, it cannot be fully ruled out that the two groups were not fully comparable at the start, and that the apparent benefit of subtitles in the later tests is at least partly the result of these initial differences rather than the subtitle condition itself. This means we cannot say for sure, based only on the raw scores, that subtitles caused the difference, and it shows why these results should be interpreted together with the models that control for background variables.

5.1.2 Production

Production was analysed in two ways: through raw production scores and through a more detailed qualitative coding of response types. In contrast to recall, neither analysis showed a significant effect of condition. This suggests that exposure to intralingual subtitles did not lead to statistically significant differences in learners' productive vocabulary performance when compared with the no-subtitle condition.

This finding is not entirely unexpected, as productive vocabulary knowledge is generally considered more demanding than receptive vocabulary knowledge. Whereas recall tasks mainly require recognition of previously encountered items, production tasks require learners to retrieve and reproduce lexical forms accurately. Productive knowledge therefore involves a deeper level of lexical mastery, including form recall and orthographic accuracy (Nation, 2001). Previous subtitle research has likewise shown that receptive gains tend to be stronger and emerge earlier than productive gains, particularly after limited exposure (Montero Pérez et al., 2014). Similarly, Kurokawa et al. (2024) reported stronger effects for recognition measures than for deeper retrieval tasks across studies included in their meta-analysis.

Although the differences were not statistically significant, descriptive patterns showed that the subtitle group tended to start slightly higher and remained somewhat stronger across time than the no-subtitle group. However, the pattern needs to be interpreted with caution

since the absence of significance may reflect limited statistical power due to small sample size, and non-linear developmental trajectories could also account for the pattern. Peters and Webb (2018) also found that L2 captions could improve written form recall more than L1 subtitles, although gains remained modest and were strongly moderated by learners' vocabulary size and frequency of occurrence. Other previous research has similarly found that subtitles often produce stronger gains for receptive vocabulary than for productive vocabulary, especially after short exposure periods (Montero Pérez et al., 2014).

Both groups also showed declining production scores over time, similar to recall. However, the unknown effect for this decline may have been even stronger for production tasks, as these required more effort than multiple-choice recall questions. Students had to actively write an answer rather than recognize one from three options. Even though the first letter was provided, the task still required more cognitive effort and time. Under conditions of fatigue or low motivation at the end of the second semester, learners may therefore have responded more quickly, left items blank, or produced less accurate answers.

The detailed production analysis also showed limited change in the quality of responses over time. This may indicate that although learners were exposed to the target vocabulary, one short viewing moment was not always sufficient for deeper lexical mastery. Vocabulary production often requires repeated encounters, active retrieval, and deliberate practice (Montero Pérez et al., 2018; Wei & Fan, 2022). Incidental exposure through subtitles may therefore support the early stages of acquiring new vocabulary such as recall, but may be less sufficient on its own for productive command. Although the current data do not allow firm conclusion about productive gains, future research with longer or more intensive subtitle exposure may be better placed to detect such effects in beginner learners.

5.1.3 Comprehension Score

For comprehension, the main effect of time was significant, indicating that comprehension scores increased across the semester regardless of condition. This suggests that learners became better at understanding the audiovisual input over time. Such improvement is likely related to general language development during the academic year, including increased

vocabulary knowledge, greater familiarity with Spanish sentence patterns, and continued classroom instruction.

The main effect of test was relevant, whereas the condition was not significant at baseline, meaning that subtitles did not lead to an overall statistically significant advantage in comprehension scores. This is consistent with the findings of Montero Pérez et al. (2014), who showed that vocabulary knowledge was a stronger predictor of comprehension success and that there were no significant differences between subtitle conditions and no-subtitle watching on video comprehension measures.

The interaction between time and condition was also non-significant, indicating that the development of comprehension did not statistically differ between groups. Descriptively, however, the group without subtitles started at a slightly higher baseline, while the subtitle group tended to perform somewhat better in later tests. This descriptive pattern resembles the recall findings and may reflect an initial adjustment period. At first, subtitles may have divided attention and created additional processing demands. Once learners became more accustomed to subtitle-supported input, subtitles may have shifted from being a source of overload to becoming a useful support tool, though this remains speculative given the absence of a significant interaction effect.

This interpretation is consistent with Cognitive Load Theory (Sweller, 1988) and Multimedia Learning Theory (Mayer, 2002). At the beginning stages of learning, simultaneous reading and listening may challenge working memory. Over time, however, learners may become more efficient at integrating spoken language, written text, and visual context. This interpretation is also consistent with Bairstow and Lavour (2012), who found that subtitles benefited lower-level learners' comprehension, while more advanced learners sometimes experienced subtitles as distracting. Since the present participants were beginner learners, subtitles may have become increasingly useful once they grew more familiar with the viewing format.

The highest comprehension scores were found on the final test. This may reflect increased general proficiency rather than the subtitle condition alone. Previous studies have shown that learners with broader vocabulary knowledge tend to perform better on listening and

comprehension tasks because they can allocate fewer cognitive resources to decoding individual words and more to overall meaning construction.

In sum, the comprehension findings suggest that subtitles did not produce a statistically significant advantage in understanding, either overall or in their development over time. Descriptive trends may hint at a gradual shift, but these should be interpreted with caution.

5.1.4 Prior Knowledge and Cross-linguistic Influence

The analyses concerning learner background variables provide important insight into the role of individual differences in incidental vocabulary learning.

First, prior knowledge of Spanish significantly affected both recall and production performance. Students who reported previous exposure to Spanish scored consistently higher on recall than those without prior knowledge, and this advantage was even more pronounced for production. This finding is in line with broader literature where learners with previous contact with the language are likely to possess stronger phonological familiarity, partial lexical knowledge, or greater confidence when producing words (Blumenfeld et al., 2106). Even limited prior exposure may therefore create an advantage, particularly in productive tasks that require active recall rather than recognition.

Second, the effect of cross-linguistic influence was more nuanced. The overall CLI group effect was non-significant for recall and productive tasks, meaning that language background did not reliably differentiate recall nor production performance between groups as a whole. Neither the Romance nor the Multilingual group differed significantly from the Dutch reference group in either recall or production. This suggests that typological proximity and multilingualism alone do not reliably differentiate vocabulary outcomes in this sample, which doesn't completely support the literature on positive transfer (Bartolotti & Marian, 2016; Flores-Salgado & Gutiérrez-Koyoc, 2024).

Subtitle effects, are robust to adjustment for language background. This may indicate that subtitles benefit learners more generally, while prior linguistic advantages independently contribute to higher scores. Language background therefore previous knowledge of Spanish appears to explain an important part of the variation in vocabulary outcomes, sometimes

more strongly than the subtitle manipulation itself, which isn't surprising as previous research have suggested this (Montero Pérez et al., 2014; Montero Pérez, 2022).

These findings highlight that subtitle effectiveness cannot be interpreted in isolation from learner characteristics. Prior knowledge of Spanish appears to explain a substantial part of variation in vocabulary outcomes, sometimes more than the condition itself. However, cross-linguistic influence from broader language background did not show any significance in this sample. One possible explanation is that the students of the Dutch reference group had educational exposure to French, a Romance language, which may have reduced the expected contrast with the Romance group.

5.2 Conclusion

This study investigated the effects of intralingual subtitles on the incidental vocabulary acquisition of A1–A2 learners of Spanish as a foreign language for specific purposes, with a particular focus on recall and production of newly encountered vocabulary items. The results indicated a modest benefit of intralingual subtitles for receptive vocabulary development, while no statistically significant effect was found for productive vocabulary. This discrepancy is consistent with the broader literature, in which receptive gains tend to emerge more readily than productive gains following limited audiovisual exposure. The greater cognitive demands associated with production tasks, requiring active retrieval and orthographic accuracy rather than mere recognition, likely contribute to this asymmetry. This pattern was further corroborated by a qualitative coding of response types, which similarly yielded no significant change over time. These findings suggest that incidental exposure through intralingual subtitles may support the early stages of lexical recognition under the current conditions, though this cannot be suggested for productive gains. Future research could explore whether more intensive or repeated exposure yield different outcomes for productive vocabulary in beginner learners.

Regarding comprehension, the analysis revealed a significant main effect of time but not of condition, indicating that comprehension scores improved across the semester regardless of subtitle presence. This pattern suggests that general proficiency development, rather than

subtitle exposure per se, was the primary driver of improved comprehension. This aligns with previous research demonstrating that vocabulary breadth is a stronger predictor of comprehension outcomes than subtitle condition alone.

Concerning learner background variables, prior knowledge of Spanish emerged as a stronger predictor of both recall and production performance than the subtitle condition itself. Cross-linguistic influence, by contrast, did not yield statistically significant differences between language background groups, which is a finding that does not fully align with the positive transfer effects reported in the broader literature, and one that warrants further investigation. Taken together, these findings suggest that individual learner characteristics, particularly prior knowledge of Spanish, may account for a substantial proportion of the variance in vocabulary outcomes, at times outweighing the effect of the students' condition. This underlines that subtitle effectiveness cannot be assessed in isolation from the learner profile.

All analyses, including recall, production (both raw scores and qualitative coding), comprehension, and cross-linguistic influence, were conducted using Linear Mixed Models (LMM), which are well-suited to account for individual developmental trajectories and the influence of background variables across repeated measurements. Prior knowledge was additionally included as a covariate in the recall and production raw score models.

Overall, the findings suggest that intralingual subtitles are not detrimental and may offer incremental support for receptive vocabulary development over time. Nonetheless, they should not be regarded as a substitute for deliberate vocabulary instruction, particularly when productive mastery is the pedagogical goal. In multilingual educational contexts such as Belgium, where learners bring diverse linguistic backgrounds to the classroom, prior knowledge of Spanish may shape vocabulary acquisition outcomes more than the condition. However broader linguistic background and typological proximity did not reliably differentiate outcomes in this sample, possibly due to the educational French exposure shared across groups.

5.3 Limitations

This study presents several limitations that should be considered when interpreting the findings.

First, students were given “jokers” throughout the semester that allowed them to miss two classes without penalty. As the tests were administered during class time, participation was therefore not obligatory. This contributed to an unequal number of observations across testing moments and may have affected the longitudinal analyses. Future research could reduce this limitation by making participation obligatory or by scheduling testing moments separately from regular attendance policies.

Second, several factors may help explain the lower scores observed at the end of the semester compared to the beginning. These include potential differences in test difficulty across measurement moments and variation in group composition due to missing observations. Regarding test difficulty, an initial exploration was conducted into the lexical properties of the five videos, focusing on lexical density through the Type-Token Ratio, Corrected Type-Token Ratio (CTTR), and the number of new vocabulary items introduced in each video. CTTR was the most useful as it is reasonably stable when text differences are moderate, it is easy to reproduce in articles, and it is particularly common in research on lexical richness and L2 writing (Baayen, 2001; Malvern et al., 2004; McCarthy & Jarvis, 2010). However, these analyses did not reveal significant patterns and were therefore not included in the main analyses. Nevertheless, this remains a potentially valuable direction for future research, as more fine-grained measures of input difficulty may contribute to a better understanding of the decline in scores observed across testing moments. Regarding group composition, an initial exploration was also conducted into how participation changed across the different testing moments. A more detailed analysis, however, could further examine whether students with stronger academic performance were systematically more likely to be absent during later tests. Such an analysis could be carried out by linking students’ test performances to their partial and final exam grades in order to determine whether a participation bias emerged over the course of data collection. Since the study was conducted in an authentic

classroom setting, these factors could not be fully controlled and may therefore have influenced the results.

Third, the sociodemographics used to assess prior language knowledge and cross-linguistic influence were based on self-report measures, including open-ended questions. Self-assessment may be imprecise, as students can overestimate or underestimate their linguistic background and language proficiency. Future studies could therefore complement self-reported data with short placement measures or standardized language background questionnaires in order to obtain more objective learner profiles. It should also be noted that the group division for the CLI analysis was based on students' responses in the survey. However, all student in the 'Dutch' group also know French to a certain degree, that they acquired through educational context only.

Fourth, the study focused specifically on A1–A2 learners of Spanish for specific purposes within one academic context. While this strengthens the internal consistency of the study, it may limit the generalizability of the findings to other learner populations, proficiency levels, educational settings, or target languages. Future research could investigate whether similar effects emerge in different instructional contexts or with learners from other linguistic and educational backgrounds.

A fifth limitation concerns the linguistic variety of Spanish used in the audiovisual materials. The videos predominantly contained Peninsular (Iberian) Spanish, while Spanish is a pluricentric language with several major variants, including Mexican, Caribbean, Andean, and Rioplatense Spanish. Exposure to different accents and pronunciation patterns may influence learners' comprehension, vocabulary acquisition, and overall processing of audiovisual input. Future studies could therefore examine whether subtitle effectiveness differs depending on the Spanish variant used, whether learners show faster adaptation and new vocabulary acquisition for particular variants, and to what extent familiarity or prior exposure to a specific variety influences comprehension and acquisition outcomes. Such research could contribute to identifying the most pedagogically appropriate audiovisual input depending on the learner population.

Furthermore, the analyses remained within the methodological scope of a master thesis. The data were primarily analysed using a linear approach, without extensively modelling potential non-linear developmental trajectories or variability in individual learning patterns, since subtitle-conditional non-linear curves would make the model extremely complicated. As a result, the available data did not provide sufficient explanatory power to fully account for the lower scores observed at the end of the semester compared to the beginning. Future research using more advanced longitudinal analyses, such as growth curve modelling, or dummy variables could provide a more nuanced understanding of fluctuations in vocabulary acquisition and listening comprehension over time.

A sixth limitation is that listening comprehension was only used as an additional outcome measure in this study, which mainly focused on vocabulary learning. As a result, more general concerns about how and when comprehension develops through audiovisual input assisted by subtitles are still not fully investigated. Future studies could look more closely at the connection between vocabulary growth and understanding development, especially from a long-term standpoint. Furthermore, the current study solely considered cognitive overload as a theoretical mediator. Future studies could use techniques like eye-tracking technology to experimentally investigate cognitive load, enabling a more thorough examination of attention distribution and subtitle processing.

Lastly, an important avenue for future research concerns learners with limited literacy skills. Some beginning A1 learners may experience difficulties with reading and writing, such as illiterates, children, or people using other writing systems than the latin alphabet, raising the question of whether intralingual subtitles genuinely facilitate language acquisition for these populations or whether the written modality may instead increase processing demands. Further research is needed to determine whether intralingual subtitles function as a support mechanism or as an additional cognitive burden for learners with limited literacy skills.

Additionally, future research could extend the current design by including a condition with interlingual subtitles in the learner's L1, which would allow for a more direct comparison of the relative effectiveness of intralingual versus interlingual subtitles for incidental vocabulary acquisition in beginner learners.

Appendix

Appendix 1: Codebook

Characteristics	Description	Examples	Code
Error-free/ correct	Unique solution, complete orthotypographical match (see the solutions key)	Precio Empresa familiar Gimnasio	When correct: 1 When incorrect: 0
Almost correct (or phonetically similar)	The word isn't a complete orthotypographical match because of missing accents or letters. The spelling errors in the word are reasonable and the intended correct word is still legible in the answer. Key: the answer is recognizable as the intended word despite minor errors. But also: The word has been written in such a way that the word can be phonetically recognized. Key:	Virtually correct examples: 'Precio': 'presio', 'preso', 'pricio', 'prescio' 'Prácticas': 'Practicás', 'Practica', 'Practico', 'Practicar' 'Esperar': 'Espirar', 'Espera', 'Espere' 'Luminosa': 'Luminoso' 'Fundador': 'Fundator', 'Fondador' 'Gimnasio': 'Gymnasio', 'Gymnasío', 'Gymasio' Phonetically similar examples: 'Gimnasio': 'Ymnasio', 'imnasia' 'Prácticas': 'praticar' 'Jefa': 'efe/efa' 'Leche': 'Laiché', 'lêche'	When virtually correct: 1 When virtually not correct: 0

	The answer is a phonetic approximation of the word.		
Synonym	The answer can be found in the official Spanish dictionary (DRAE) as a synonym for the correct word/answer.	'Escritorio': 'mesa' 'Revista': 'periódico' 'Alimentación': 'comida', 'nutrición' 'Seguir': 'continuar' 'Consejo': 'aviso'	When synonym: 1 When not a synonym: 0
Incorrect	1) The answer doesn't have a common root or doesn't show any morphological resemblance with the solution. 2) The answer is not a synonym 3) The answer is given in another language 4) Is a synonym for the correct word/answer, but it was written incorrectly: The word isn't a complete orthotypographical match	'Cita': correo, comienza, calidad, cena, comida, concordancia, cadena 'Barrio': busco, arbre, sale, avenida, hijo 'desde': después, vente, delante, antes 'llamada': 'appel' 'fecha': 'datum' 'Leche': 'latte'	When incorrect: 1 When correct: 0

Nothing entered	The student did not enter anything but was present on the day of the test/exam. (refer to list of presence)	Empty cell	Cell empty: 1 Cell not empty: 0
Not present at the test/exam	The cell is empty because the student was not present at the test/ exam. (refer to list of presence)	Empty cell	999

Appendix 2: Informed Consent

Informed Consent NON-ANONYMOUS RESEARCH

Dear participant,

You wish to participate in a scientific study on **Incidental Vocabulary Acquisition and Listening Comprehension in A1 Learners of Spanish as a Foreign Language for Specific Purposes** carried out by Almudena Basanta and Lieve Vangehuchten from the University of Antwerp, and Pilar Pérez from the University of Vienna.

In a moment, we will present you a questionnaire and some questions on the video you've just seen. This will take about 15 minutes.

When participating in this study, you are recognizable to our researchers in the data we collect. According to **GDPR-legislation**, we are **required to register your first- and surname, as well as student ID number, when you agree to participate**. We store this information in which you are recognizable separately and digitally for 10 years in password-protected files in a separate location on secure servers of the AUHA (Association of University and Colleges Antwerp). This information is only accessible to Almudena Basanta and Lieve Vangehuchten.

Immediately after the data collection and before the start of the analyses, **we make our datasets pseudonymous**. This means that in our datasets we replace data with which you are recognizable with a pseudonym (another name or a code). In another so-called 'key file' we link this pseudonym to your identity. We store key files separately, in password-protected files in a separate location on secure servers of the AUHA. Those who have access to the key files will treat them confidentially and the key files will not be shared further. Only Lieve Vangehuchten has access to both the datasets and the key files and can still make you recognizable in the datasets/results. For all other people, you are no longer immediately recognizable in the datasets and results of this study. The pseudonymized datasets are stored digitally for up to 10 years after the end of the study.

You have the right to view files in which you are recognizable as long as these files are stored. You can only view data about yourself in these files. Changing or deleting data may be exceptionally allowed if this does not jeopardize the purpose of the research.

You have the right to refuse to participate and to stop your participation at any time. You do not have to explain this and this will not entail any disadvantage or loss of benefits. The data already collected can be deleted at your request if this does not jeopardize the purpose of the research.

The researchers of this study have the right to share anonymous databases (in which participants are not recognizable to anyone) **with (inter)national colleagues** in the context of (this and/or later) scientific research. These data files are not accessible to other parties,

including yourself or companies/organizations that do not fit within scientific research institutions.

If you have any further questions or comments, please contact **the contact person of this project** lieve.vangehuchten@uantwerpen.be. If you have any questions or comments about the collection of data with which you are recognizable, you can also contact the **Data Protection Officer or the Privacy Service of the University of Antwerp**: privacy@uantwerpen.be. If you believe that one of the aforementioned persons is not processing your personal data lawfully and in accordance with the legal requirements, you have the right to lodge a complaint with the Data Protection Authority (contact details available at: <https://www.gegevensbeschermingsautoriteit.be/>). We do recommend, in case of complaints, to first contact our data protection officer. Often any problems or misunderstandings can be solved quickly and easily in this way.

Please answer the questions below and indicate if you want to participate in the study. Many thanks in advance!

- I have read and understood the information in the information document. YES/NO
- I voluntarily agree to participate. YES/NO
- In case you participate, fill in your NAME & FIRST NAME:
- In case you participate, fill in your Student number:

In Qualtrics the form will be presented as follows:



Informed

Consent

NON-ANONYMOUS

RESEARCH

Dear

participant,

You wish to participate in a scientific study on **Incidental Vocabulary Acquisition and Listening Comprehension in A1 Learners of Spanish as a Foreign Language for Specific**

Purposes carried out by Almudena Basanta and Lieve Vangehuchten from the University of Antwerp, and Pilar Pérez from the University of Vienna.

In a moment, we will present you a questionnaire and some questions on the video you've just seen. This will take about 15 minutes.

When participating in this study, you are recognizable to our researchers in the data we collect. According to **GDPR-legislation**, we are **required to register your first- and surname, as well as student ID number, when you agree to participate.** We store this information in which you are recognizable separately and digitally for 10 years in password-protected files in a separate location on secure servers of the AUHA (Association of University and Colleges Antwerp). This information is only accessible to Almudena Basanta and Lieve Vangehuchten.

Immediately after the data collection and before the start of the analyses, **we make our datasets pseudonymous.** This means that in our datasets we replace data with which you are recognizable with a pseudonym (another name or a code). In another so-called 'key file' we link this pseudonym to your identity. We store key files separately, in password-protected files in a separate location on secure servers of the AUHA. Those who have access to the key files will treat them confidentially and the key files will not be shared further. Only Lieve Vangehuchten has access to both the datasets and the key files and can still make you recognizable in the datasets/results. For all other people, you are no longer immediately recognizable in the datasets and results of this study. The pseudonymized datasets are stored digitally for up to 10 years after the end of the study.

You have the right to view files in which you are recognizable as long as these files are stored. You can only view data about yourself in these files. Changing or deleting data may be exceptionally allowed if this does not jeopardize the purpose of the research.

You have the right to refuse to participate and to stop your participation at any time. You do not have to explain this and this will not entail any disadvantage or loss of benefits. The data already collected can be deleted at your request if this does not jeopardize the purpose of the research.

The researchers of this study have the right to share anonymous databases (in which participants are not recognizable to anyone) **with (inter)national colleagues** in the context of (this and/or later) scientific research. These data files are not accessible to other parties, including yourself or companies/organizations that do not fit within scientific research institutions.

If you have any further questions or comments, please contact **the contact person of this project** lieve.vangehuchten@uantwerpen.be. If you have any questions or comments about the collection of data with which you are recognizable, you can also contact the **Data Protection Officer** or the **Privacy Service of the University of Antwerp**: privacy@uantwerpen.be. If you believe that one of the aforementioned persons is not processing your personal data lawfully

and in accordance with the legal requirements, you have the right to lodge a complaint with the Data Protection Authority (contact details available at: <https://www.gegevensbeschermingsautoriteit.be/>). We do recommend, in case of complaints, to first contact our data protection officer. Often any problems or misunderstandings can be solved quickly and easily in this way.

Please answer the questions below and indicate if you want to participate in the study. Many thanks in advance!

I have read and understood the information in the information document.

☐ Yes

☐ No

I voluntarily agree to participate

☐ Yes

☐ No

In case you participate, fill in your NAME & FIRST NAME

In case you participate, fill in your Student number

START

Appendix 3: Assistance with third-party research

Table 16: Assistance with third part research

	Date	Theme of Research	Type	Time
1	13/11/2023 25/03/2024	ENG: The development of receptive and productive vocabulary knowledge and the factors that influence it ND: Ontwikkeling van receptieve en productieve woordenschatkennis en factoren die hierin een rol spelen (vak: Zakelijk en Economisch Duits 2)	RESP	1h (30min per semester)
2	29/03/2024	ENG: Comprehension of newspaper articles among German language learners ND: Het begrip van krantenartikels bij taalleerders van het Duits	RESP	50min
3	17/05/2025	ENG: Interpretation of sentences in Spanish ESP: Interpretación de oraciones en español	RESP	20min
4	6/11/2025	ENG: UHasselt: Use of English in guest reviews ND: UHasselt: Gebruik van Engels in gastenreviews	RESP	5min
5	3/12/2025	ENG: Research on writing behaviour ND: Onderzoek over schrijfgedrag	ANDER	1u30
6	29/03/2026	ENG: A study on how young adults use their smartphones while driving	RESP	15min

		ND: Onderzoek naar hoe jongvolwassenen hun smartphone gebruiken in het verkeer		
7	14/04/2026	ENG: Politainment in Spain: A Critical Analysis of Far-Right Discourse on Social Media ESP: Politainment en España: Un análisis crítico del discurso de la extrema derecha en las redes sociales	CODEUR2	2u

Bibliography

Almeida, P. A., & Costa, P. D. (2014). Foreign language acquisition: The role of subtitling. *Procedia – Social and Behavioral Sciences*, 141, 1234–1238.

Baarda, B., & Van Vianen, R. (2025). *Dit is onderzoek!* (4e ed.). Noordhoff., p.82-85

Bairstow, D., & Lavour, J. (2012). Audiovisual Information Processing by Monolinguals and Bilinguals: Effects of Intralingual and Interlingual Subtitles. In *BRILL eBooks* (pp. 273–293). https://doi.org/10.1163/9789401207812_016

Bairstow, D., & Lavour, J. (2015). Sous-titrage, compréhension de films et acquisition de vocabulaire. *Psychologie Française*, 62(3), 249–261. <https://doi.org/10.1016/j.psfr.2015.05.001>

Bartolotti, J., & Marian, V. (2016). Bilinguals' existing languages benefit vocabulary learning in a third language. *Language Learning*, 67(1), 110–140. <https://doi.org/10.1111/lang.12200>

Baumeister, F., Hanushek, E. A., & Woessmann, L. (2025). Out-of-school learning: Subtitling vs. dubbing and the acquisition of foreign-language skills [Preprint]. SSRN.

Bensalem, E. (2018). The efficacy of captions on students' incidental vocabulary acquisition. *Journal of Teaching and Teacher Education*, 6(1), 1–11.

Bird, S. A., & Williams, J. N. (2002). The effect of bimodal input on implicit and explicit memory. *Applied Psycholinguistics*, 23(4), 509–533.

Black, S. (2021). The potential benefits of subtitles for enhancing language acquisition and literacy in children. *Translation, Cognition & Behavior*, 4(1), 74–97.

Blumenfeld, H. K., Bobb, S. C., & Marian, V. (2016). The role of language proficiency, cognate status and word frequency in the assessment of Spanish–English bilinguals' verbal fluency. *International Journal Of Speech-Language Pathology*, 18(2), 190–201. <https://doi.org/10.3109/17549507.2015.1081288>

Chaudron, C. (1985). Intake: On Models and Methods for Discovering Learners' Processing of Input. *Studies in Second Language Acquisition*, 7(1), 1–14. <https://doi.org/10.1017/s027226310000512x>

Danan, M. (2004). Captioning and subtitling: Undervalued language learning strategies. *Meta*, 49(1), 67–77.

- De Maeyer, S., Coertjens, L., & Ardies, J. (2012). *Bivariate en multivariate statistiek: Een Open Leerpakket in R* (7de editie). Lannoo nv.
- Díaz Gutiérrez, E., Narvajás Colón, E., Pérez Cañizares, P., & Suárez Lasierra, M. (2023). *Meta profesional Plus A1–A2*. Klett.
- European Union (2024). Europeans and Their Languages. Special Eurobarometer 540. <https://europa.eu/eurobarometer/surveys/detail/2979> (accessed May 22nd 2026).
- Flores-Salgado, E., & Gutiérrez-Koyoc, A. F. (2024). Working memory and cross-linguistic influence on vocabulary acquisition. *Brain Sciences*, 14(8), 796.
- Frumuselu, A. D., De Maeyer, S., Donche, V., & Gutiérrez Colon Plana, M. M. (2015). Television series inside the EFL classroom. *Linguistics and Education*, 32, 107–117.
- Harald Baayen (2001). *Word Frequency Distributions*. Kluwer.
- I. S. P. Nation (2001). *Learning Vocabulary in Another Language*. Cambridge University Press.
- Krashen, S. (1985). *The input hypothesis: Issues and implications*. Longman.
- Kurokawa, S., Hein, A. M., & Uchihara, T. (2024). Incidental vocabulary acquisition through captioned viewing: A meta-analysis. *Language Learning*. <https://doi.org/10.1111/lang.12697>
- Lombard, M., Snyder-Duch, J., & Bracken, C. C. (2002). Content Analysis in Mass Communication: Assessment and Reporting of Inter-coder Reliability. *Human Communication Research*, 28(4), 587–604. <https://doi.org/10.1111/j.1468-2958.2002.tb00826.x>
- Malvern, D., Richards, B., Chipere, N., & Durán, P. (2004). *Lexical Diversity and Language Development*. Palgrave Macmillan.
- Mayer, R. E. (2002). Multimedia learning. In *The Psychology of learning and motivation* (Vol. 41, Nummer 0, pp. 27–139). [https://doi.org/10.1016/s0079-7421\(02\)80005-6](https://doi.org/10.1016/s0079-7421(02)80005-6)
- McCarthy, P. M., & Jarvis, S. (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>
- Montero Pérez, M. (2022). Second or foreign language learning through watching audio-visual input. *Language Teaching*, 55(2), 163–192.

- Montero Pérez, M., Peters, E., Clarebout, G., & Desmet, P. (2014). Effects of captioning on video comprehension and incidental vocabulary learning. *Language Learning & Technology*, 18(1), 118–141.
- Montero Pérez, M., Peters, E., & Desmet, P. (2018). Vocabulary learning through viewing video. *Computer Assisted Language Learning*, 31(1–2), 1–26.
- Montero Pérez, M., Van den Noortgate, W., & Desmet, P. (2013). Captioned video for L2 learning: A meta-analysis. *System*, 41(3), 720–739.
- Paivio, A. (1986). *Mental representations: A dual coding approach*. Oxford University Press.
- Peters, E., Heynen, E., & Puimège, E. (2016). Learning vocabulary through audiovisual input. *System*, 63, 134–148.
- Peters, E., & Webb, S. (2018). Incidental vocabulary acquisition through viewing L2 television. *Studies in Second Language Acquisition*, 40(3), 551–577.
- Reynolds, B. L., Cui, Y., Kao, C., & Thomas, N. (2022). Vocabulary acquisition through captioned and subtitled video. *Systems*, 10(5), 133.
- Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, 12(2), 257–285. [https://doi.org/10.1016/0364-0213\(88\)90023-7](https://doi.org/10.1016/0364-0213(88)90023-7)
- Van der Kolk, J., & Feijoo, S. (2024). Incidental vocabulary recognition effects of audiovisual input. *Revista de Lingüística y Lenguas Aplicadas*, 19, 218–230.
- Vanderplank, R. (1993). A very verbal medium: Language learning through closed captions. *TESOL Journal* 3.1, 10–14.
- Vanderplank, R. (2013). Effects of captions on language learners. *Language Teaching*, 49, 235–250.
- Vanderplank, R. (2016). The Value of Closed Captions and Teletext Subtitles for Language Learning. In *Palgrave Macmillan UK eBooks* (pp. 43–73). https://doi.org/10.1057/978-1-137-50045-8_3
- Vandervieren, E. (2025). H3: Data en datamatrix. In *Statistiek A* (pp. 36–39).
- Wei, R., & Fan, L. (2022). On-screen texts in audiovisual input for L2 vocabulary learning. *Frontiers in Psychology*, 13, 904523.

Winke, P., Gass, S., & Sydorenko, T. (2013). Factors influencing the use of captions by foreign language learners: An eye-tracking study. *The Modern Language Journal*, 97(1), 254–275.

Use of AI Tools

I, Elisa Mames Antuna have used the following AI tools: ChatGPT, Claude, and Scribbr

Table 1

Activities done by AI

Activities	Yes	No
Have text proofread for typos, spelling mistakes, and grammatical errors	+	
Have text corrected for style and wording	+	
Have sentences rephrased	+	
Generate a reference list in APA style	+	
Check reference list for APA style		+
Improve the wording of the research question		+
Get suggestions for Dutch translations of English concepts		+
Get suggestions for possible English formulations of items from foreign-language measurement instruments		+

Table 2

Overview of AI-use

Date	AI-tool	Short description of the task (with link to the prompt and output)
21/01/2026	ChatGPT	Have the text “subject, concept, and definition” proofread for linguistic errors, sentence structure, and style.

		Link: https://chatgpt.com/share/6970c5af-3940-8005-8e95-d8109514666b
26/01/2026	ChatGPT	Have the text about 'control variables' be checked on syntax and grammatical errors. Link: https://chatgpt.com/share/6970c5af-3940-8005-8e95-d8109514666b
26/01/2026	ChatGPT	Have the text about 'data analysis' be checked on syntax and grammatical errors. Link: Link: https://chatgpt.com/share/6970c5af-3940-8005-8e95-d8109514666b
15/04/2026	Claude	Sparring partner with AI to create coding in R and asking questions on whether to include parts of my book on statistics in the methodology section. (The thread was started in February, but the last use was in May). Link: https://claude.ai/share/2ac1d404-3e96-48c4-b9dd-7bfca1c616b3
20/04/2026	Claude ChatGPT	Have the text about 'Model 1' and 'Model 2' (now raw score and detailed production) in the Methodology section be checked on syntax and grammatical errors. Link to Claude: https://claude.ai/share/e9b4e048-0430-46d1-adda-062556eb6659 Link to ChatGPT:
20/04/2026	ChatGPT	Check for spelling, syntax, and flow of sentences: Link: https://chatgpt.com/share/6a15723c-b5c8-83eb-b338-63c055197c66
5/05/2026	ChatGPT	Sparring partner and adaptation of phrasing:

		Link: https://chatgpt.com/share/6a157594-175c-83eb-ba56-5f7f4b492224
15/05/2026	ChatGPT	Check for spelling, syntax, and grammar: Link: https://chatgpt.com/share/6a15754a-24f8-83eb-bf50-30a468be5b05
18/05/2026	ChatGPT	Make paragraphs more concise: Link: https://chatgpt.com/share/6a1572d2-f194-83eb-8d46-01419f7a4659
25/05/2026	Claude	Check that the references are all correctly there in text and bibliography Link: https://claude.ai/share/ccd870bb-c277-42e0-9dfa-e41583b49e47
26/05/2026	ChatGPT	Check for grammar, spelling, syntax or rephrasing possibilities. Link: https://chatgpt.com/share/6a1570b2-d93c-83eb-9495-e1468491bfa5
/	Scribbr	Create APA reference

List of tables

Table 1: Gender Distribution

Error! Bookmark not defined.

Gender	N	%
Female (F)	75	43.6
Male (M)	83	48.3
Not mentioned (NA)	14	7.0

Table 2: distribution by study programme

Error! Bookmark not defined.

Studies	N	%
Business Management	82	47.7
Economic Policy	8	4.7
Commercial Engineer	36	20.9
Political Sciences	1	0.6
Social and Economic Sciences	41	24.4
Communication Sciences	2	1.2
NA	1	0.6

Table 3: distribution by condition

Condition	N	%
With subtitles	88	51.2
Without subtitles	84	48.8

Table 4: Gender Distribution per condition

Condition \ Gender	Gender		
	Female	Male	NA
With	32 (36.4%)	49 (55.7%)	7 (7.9%)
Without	43 (51.2%)	34 (40.5%)	7 (8.2%)

Table 5: Example of one question

Example of a recall question: Yo _____ mucho a Berlín. (reizen; to travel) apellido (1) ; viajo (2) ;congreso (3)	Example of a production question: Hannah estudia P_____ (reclame; advertising) (publicidad)
---	--

Table 6: Overview of the Descriptives for all outcome measures

Measure	N	Observations	Mean	SD	Min	Max
Students						

Recall	171	591	3.91	1.28	0	5
Production (raw)	171	590	1.59	1.43	0	5
Production (detailed)	171	2940	3.32	1.63	1	5
Comprehension	171	590	3.48	1.42	0	7

Table 7: Fixed Effects Estimates of Recall

Predictor	β	SE	t	p
Intercept	4.31	0.15	28.72	< .001
Test	-0.17	0.05	-3.68	< .001
Condition	0.65	0.21	3.05	.002
Test x Condition	-.021	0.06	-3.37	< .001

Table 8: Fixed Effects Estimates of Production

Predictor	β	SE	t	p
Intercept	2.07	0.17	12.44	< .001
Test	-0.14	0.04	-3.13	.002
Condition	-0.36	0.24	-1.53	.127
Test x Condition	0.02	0.06	0.37	.709

Table 9: Fixed Effects Estimates of Detailed Production

Predictor	β	SE	t	p
Intercept	3.04	0.18	16.80	< .001
Test	0.09	0.04	2.43	.016
Condition	0.23	0.23	1.14	.256
Test x Condition	-0.02	0.05	-0.37	.713

Table 10: Fixed Effects of GLMM on 'Almost Correct' & 'Error-free'

Predictor	Almost Correct				Error-free			
	β	SE	z	p	β	SE	z	p
Intercept	-1.96	0.28	-6.93	< .001	-0.74	0.24	-3.14	.002
Test	-0.14	0.06	-2.11	.035	-0.17	0.05	-3.53	<.001
Condition	0.08	0.26	0.31	.760	-0.42	0.27	-1.56	.119
TestX Condition	-0.02	0.09	-0.26	.792	0.03	0.07	0.39	.696

Table 11: Fixed Effects Estimates of Comprehension Score

Predictor	β	SE	t	p
Intercept	3.18	0.17	18.35	< .001
Test	0.14	0.04	2.61	.009
Condition	-0.06	0.25	-0.24	.813

Test x Condition	-0.05	0.08	-0.66	.508
------------------	-------	------	-------	------

Table 12: Fixed Effects Estimates of Recall & Prior Knowledge as moderator

Predictor	β	SE	t	p
Intercept	4.78	0.16	30.37	< .001
Test	-0.38	0.04	-8.70	< .001
Condition	-0.55	0.21	-2.59	.010
Prior Knowledge	0.51	0.15	3.48	< .001
Test x Condition	0.21	0.06	3.30	.001

Table 13: Fixed Effects Estimates of Production & Prior Knowledge as moderator

Predictor	β	SE	t	p
Intercept	1.51	0.18	8.42	< .001
Test	-0.12	0.04	-2.81	.005
Condition	0.45	0.24	1.91	.057
Prior Knowledge	0.65	0.20	3.22	.002
Test x Condition	-0.02	0.06	-0.39	.695

Table 14: Fixed Effects Estimates of Recall & CLI

Predictor	β	SE	t	p
Intercept	4.95	0.16	30.50	< .001
Test	-0.37	0.04	-8.33	< .001
Condition	-0.69	0.21	-3.23	.001
CLI: Romance	0.37	0.24	1.56	.122
CLI: Multilingual	-0.05	0.15	-0.35	.731
Test x Condition	0.21	0.06	3.27	.001

Table 15: Fixed Effects Estimates of ProductionI & CLI

Predictor	β	SE	t	p
Intercept	1.78	0.19	9.51	< .001
Test	-0.13	0.04	-3.04	.003
Condition	0.16	0.24	0.66	.509
CLI: Romance	0.37	0.32	1.15	.252
CLI: Multilingual	-0.06	0.20	-0.29	.773
Test x Condition	-0.001	0.06	-0.02	.984

List of images

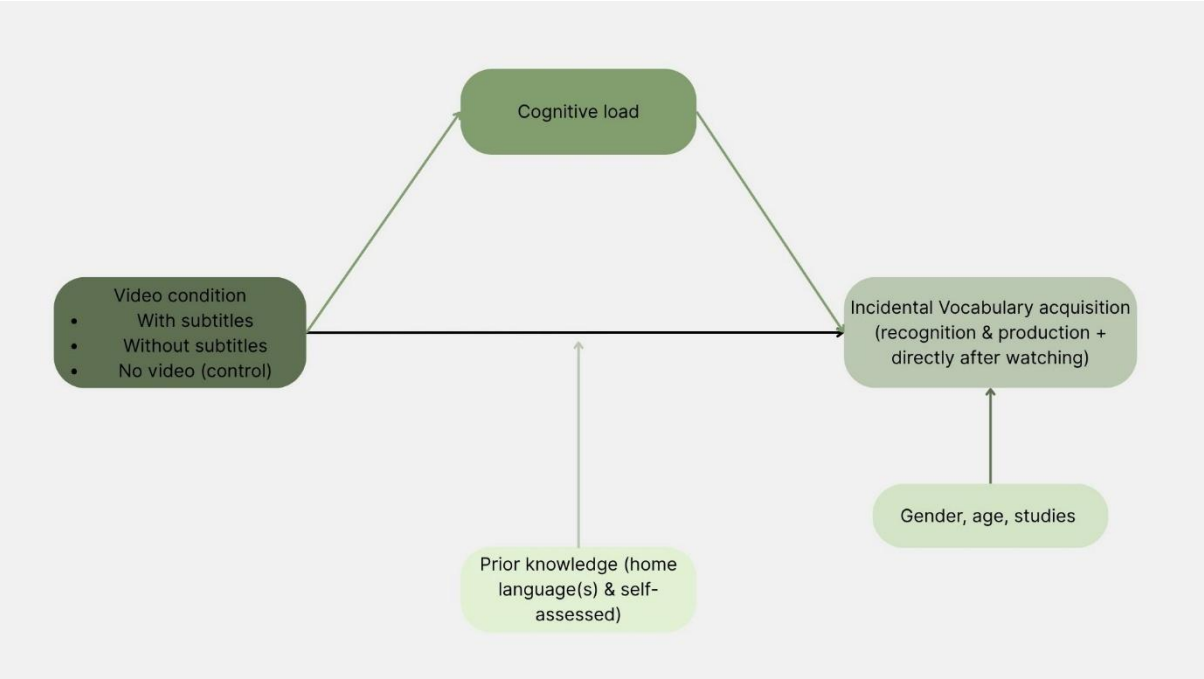


Figure 2: Conceptual Model

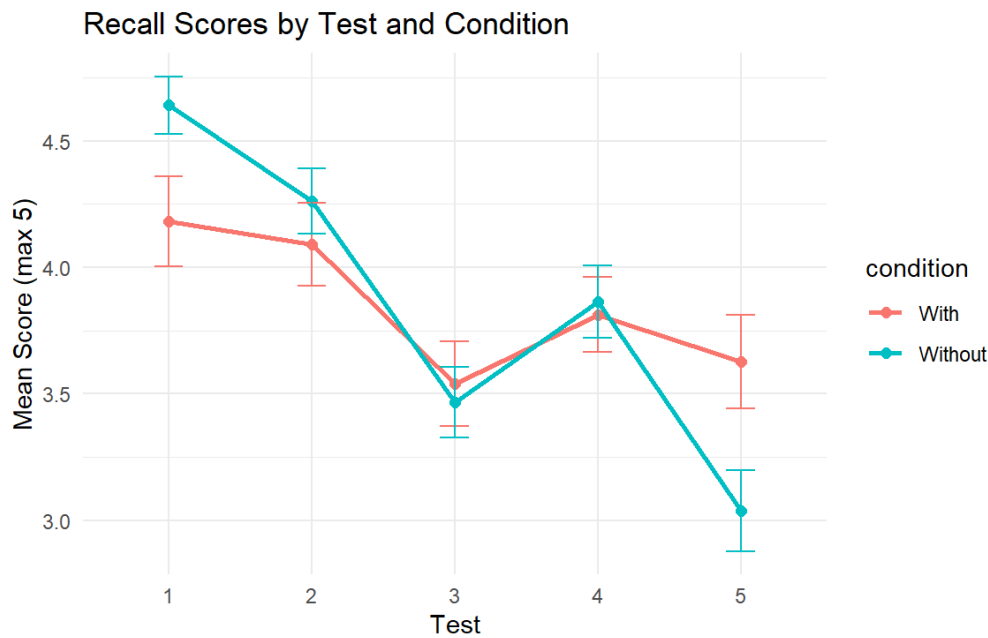


Figure 2: Recall Scores by Test and Condition

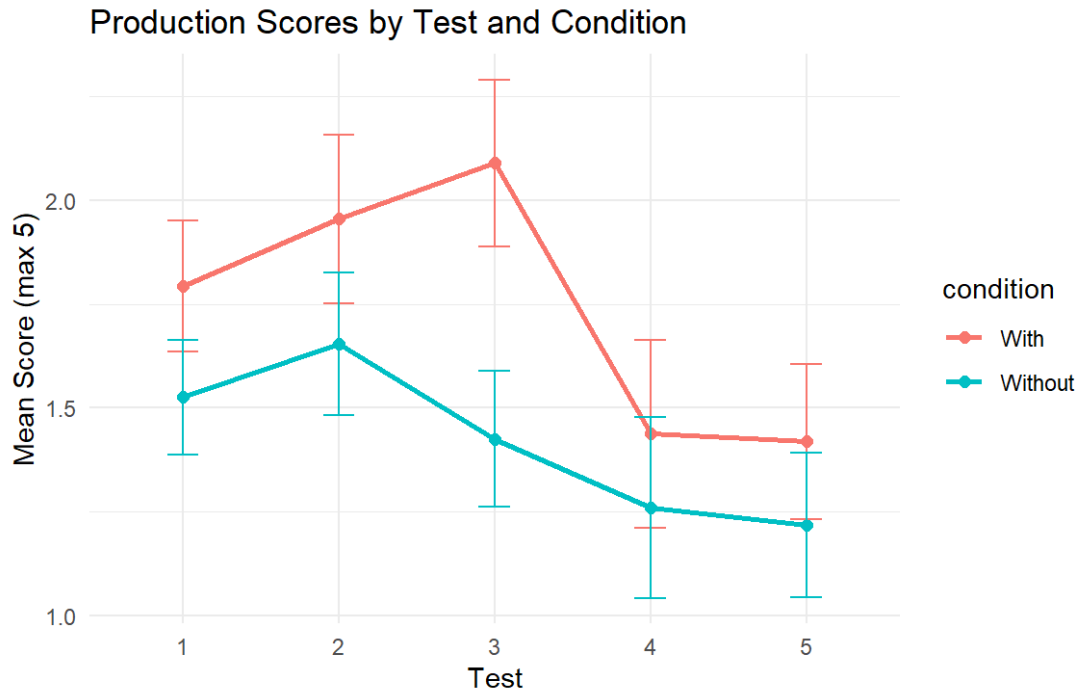


Figure 3: Production Scores by Test and Condition



Figure 4: Production Category Scores by Test and Condition



Figure 5: Comprehension Scores by Test and Condition

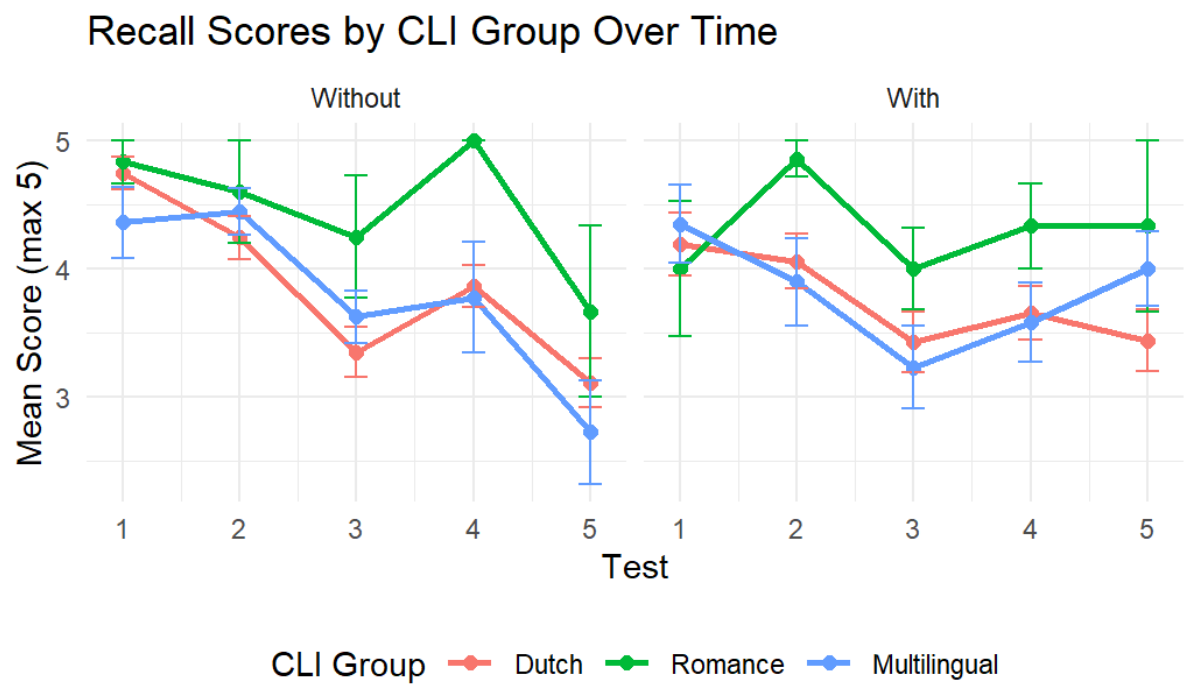


Figure 6: Recall Scores by CLI Group Over Time

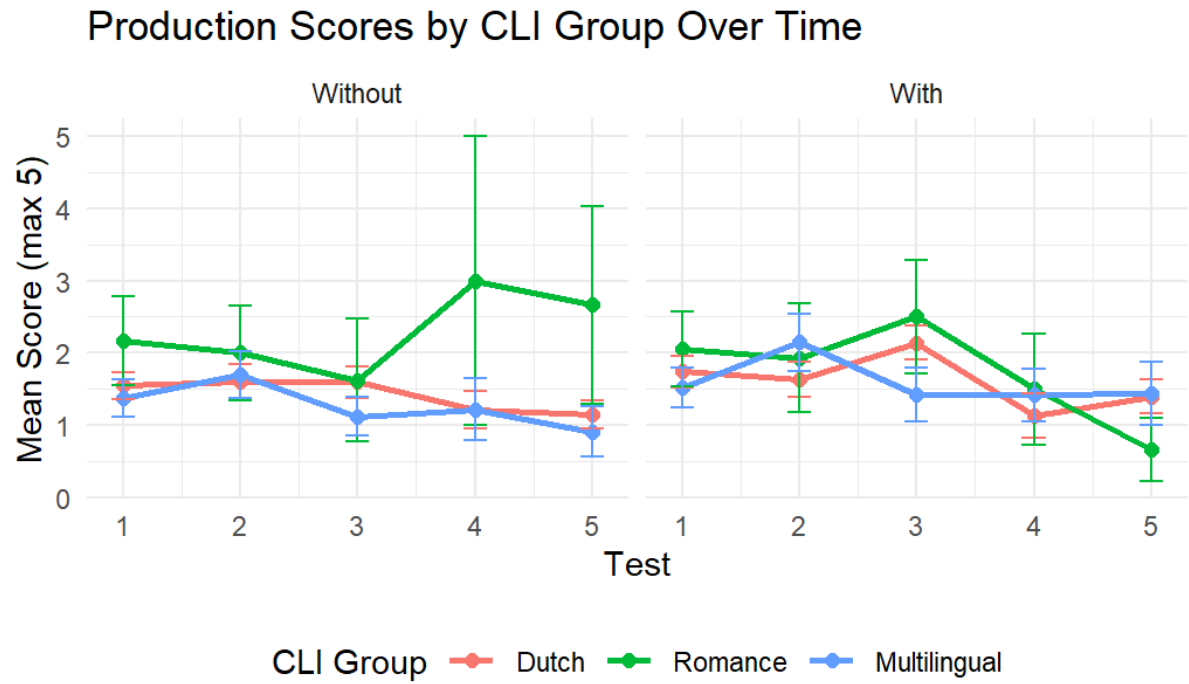


Figure 7: Production Scores by CLI Group Over Time