

# Transforming Foreign Language Teaching Practices through Communicative Classroom-based Assessment Programs

Charline Rouffet





# **Transforming Foreign Language Teaching Practices through Communicative Classroom-based Assessment Programs**

Charline Rouffet

Copyright 2024 © C. Rouffet

The Netherlands. All rights reserved. No parts of this thesis may be reproduced, stored in a retrieval system or transmitted in any form or by any means without permission of the author.

ISBN/EAN: 978-90-393-7662-1

# **Transforming foreign language teaching practices through communicative classroom-based assessment programs**

**De praktijk van het vreemdetalenonderwijs veranderen door middel van communicatieve toetsprogramma's**

(met een samenvatting in het Nederlands)

## **Proefschrift**

ter verkrijging van de graad van doctor aan de  
Universiteit Utrecht  
op gezag van de  
rector magnificus, prof. dr. H.R.B.M. Kummeling,  
ingevolge het besluit van het college voor promoties  
in het openbaar te verdedigen op  
vrijdag 31 mei 2024 des ochtends te 10.15 uur

door

**Charline Maguy Lydie Rouffet**

geboren op 28 oktober 1983  
te Nemours, Frankrijk

**Promotor:**

Prof. dr. H.C.J. de Graaff

**Copromotor:**

Dr. C.G. van Beuningen

**Beoordelingscommissie:**

Dr. L. Baartman

Prof. dr. H.H. van den Bergh

Prof. dr. C.A.M. van Boxtel

Prof. dr. M.C. Michel

Prof. dr. K. Van den Branden

Dit proefschrift werd (mede) mogelijk gemaakt met financiële steun van Hogeschool Utrecht.

**Pour Chloé et Oscar,**

gardez toujours votre curiosité et  
votre envie d'apprendre et de découvrir d'autres langues et cultures.

# Table of Contents

<b>Chapter 1 General Introduction</b>	11
1.1 Research Context	14
1.2 Research Background	15
1.3 Research Focus	17
1.4 Structure Overview	19
<b>Chapter 2 Constructive Alignment in Foreign Language Curricula: An Exploration of Teaching and Assessment Practices in Dutch Secondary Education</b>	23
2.1 Introduction	25
2.1.1 FL Teaching in the Dutch Educational System	25
2.1.2 Implementation of CLT: Influencing Factors	27
2.1.3 Constructive Alignment: The Crucial Role of Assessments	28
2.1.4 Current Study	29
2.2 Method	30
2.2.1 Participants	30
2.2.2 Instruments, Data Collection and Data Analysis	31
2.3 Results	34
2.3.1 Degree of Alignment (RQ2.1)	34
2.3.2 Conceptual and External Factors Influencing FL Teachers' Pedagogical Choices (RQ2.2)	38
2.4 Discussion	40
2.4.1 Limitations	42
2.4.2 Implication for Practice	42
<b>Chapter 3 Designing a Communicative Classroom-Based Foreign Language Assessment Program for Dutch Secondary Schools: A Design based Research Project within a Professional Learning Community</b>	45
3.1 Introduction	47
3.1.1 Classroom-based Assessment	48
3.1.2 The Role of Language Assessment Literacy in CBA	49
3.1.3 Co-design of a Communicative CBA program	50
3.1.4 The Current Study	50
3.2 Method	51
3.2.1 Participants	51
3.2.2 Co-design Process	51
3.2.3 Data Collection and Analysis	56
3.3 Results	57
3.3.1 Challenges and Design Decisions regarding Validity	57
3.3.2 Challenges and Design Decisions regarding Feasibility	58
3.3.3 Final Design	63
3.4 Discussion and Conclusion	70
3.5 Limitations and Implications	72

<b>Chapter 4 The Implementation of a Foreign Language Classroom-based Assessment Program in Dutch Lower Form Secondary Education</b>	<b>75</b>
4.1 Introduction	77
4.1.1 The Current Study	79
4.2 Method	80
4.2.1 Educational Context	80
4.2.2 Design of an Intervention with “Built-in Adaptations”	81
4.2.3 Participants	82
4.2.4 Implementation’s Procedure	82
4.3 Data Collection and Analysis	83
4.3.1 RQ4.1: Actual Implementation (Operational)	84
4.3.2 RQ4.2: Perceived Implementation (Perceived and Experiential)	85
4.4 Results	86
4.4.1 RQ4.1: Actual Implementation (Operational)	86
4.4.2 RQ4.2: Perceived Implementation (Perceived and Experienced)	88
4.5 Discussion and Conclusion	96
4.5.1 Limitations of this Study	98
4.5.2 Recommendations for Teaching Practice	98
<b>Chapter 5 The Effects of the Implementation of a Communicative Classroom-based Assessment Program on Foreign Language Teaching Practices</b>	<b>101</b>
5.1 Introduction	103
5.1.1 CLT in Practice	103
5.1.2 Constructive Alignment and Washback Effect	105
5.1.3 Classroom-based Assessment	105
5.1.4 Current Study	106
5.2 Method	107
5.2.1 Participants	108
5.2.2 Intervention	108
5.2.3 Data Collection and Analysis	110
5.3 Results	113
5.3.1 Observed Effects	113
5.3.2 Perceived Effects	115
5.4 Discussion	122
<b>Chapter 6 General Conclusion and Discussion</b>	<b>127</b>
6.1 Aims and Setup of the Project	129
6.2 Summary of the Main Findings	132
6.3 Limitations	136

6.4 Suggestions for Educational Practice	137
6.5 General Conclusion	138
<b>Bibliography</b>	140
<b>Appendix A</b>	151
Observation Instrument for Communicative FL Teaching	150
<b>Appendix B</b>	151
Handleiding Communicatief toetsen en formatief handelen	151
<b>Nederlandse samenvatting</b>	152
Inleiding	152
Opzet en belangrijkste bevindingen	154
Conclusie	160
Aanbeveling voor de onderwijspraktijk	160
<b>Dankwoord/Remerciements</b>	162
<b>About the Author</b>	165
<b>List of Publications</b>	166



**CHAPTER 1**



# General Introduction



As strange as it may sound, I came to the Netherlands with the goal of learning English. At the age of 20, after having taken English classes in France at secondary school and university for almost eight years, I was still unable to communicate in this language, despite my efforts, interest and motivation. My cousin, who was doing a post-doc in Utrecht, convinced me that everyone in the Netherlands spoke English and that I would easily improve my English language skills by spending my summer there. In about two months, working in a restaurant and hanging out with international students, I indeed managed to interact in English. I was not very proficient, but I was able to participate in a conversation and get to know new people. And that is how I met my Dutch husband. I had clearly achieved my goal beyond expectations.

Once living in the Netherlands, I soon realized that even though many Dutch people speak English quite fluently, it is clearly not the language used in everyday life. That is why I decided to take a Dutch language course. As with English, I learned Dutch mainly by talking to people and practicing the language in real-life situations. However, the classes I took also really contributed to my learning process. They were quite different from the language classes I took at school in France, in which we almost never heard our teacher speak the target language and mainly had to learn and apply grammar rules in isolated, irrelevant sentences. During the Dutch lessons, we basically talked a lot of Dutch with the teacher and with each other and developed our language skills by communicating about contemporary topics directly related to our interests. I do not remember having to learn grammar rules, lists of irregular verbs or a bunch of words by heart. I just looked these formal aspects up when I needed them to express myself or to better understand a text, an audio document, or my interlocutor, whom I often asked directly what a word or expression meant. This was actually my first experience with communicative language teaching (CLT). The primary focus of the lessons was on the development of our Dutch language skills, with the aim to be able to understand information and interact with others, or, in other words: to be able to communicate in the new language. Attention to formal aspects (e.g., correct grammar, sentence structure and spelling) was only provided when it was needed to help me and my fellow students achieve this ultimate communicative goal.

At the time, I was studying French modern literature, which I did partly in Amsterdam through the Erasmus program. I initially aspired to teach French literature, but inspired by the Dutch language classes, I decided to become a teacher of French as a foreign language (FL) instead and enrolled in the teacher training program at the Vrije Universiteit in Amsterdam. Even though this is more than fifteen years ago, the methodology courses at that time were comparable to those offered to prospective FL teachers today. The emphasis was on communicative language teaching and the use of the target language, based on the Common European Framework of Reference (CEFR). I was immedi-

ately very enthusiastic about this way of FL teaching. I thought that this communicative approach was the reason why most Dutch people could speak English so well, but I soon learned that this was not the way FLs were usually taught, at least not everywhere.

My first internship in the lower form of a secondary school in the Netherlands was a bit of a shock. I quickly realized how difficult it was to put the CLT theory I learned in my methodology classes into practice. My pupils did not understand a word of French and were not at all used to a French teacher speaking to them in French. In my vain attempts to do so, I faced strong protests, not only from them but also from their parents. They found that my communicative approach was not efficient and that my authentic tasks were a waste of time because, obviously, the focus was not on the grammar rules, verb conjunctions and the large number of words pupils had to learn for the assessment. I actually had to agree with them when I understood what kind of assessments my students had to pass. For these assessments students had to apply relatively complex grammar rules and they had to be able to translate large number of words and sentences correctly. Teaching French in a communicative way to achieve this goal was indeed not the most efficient way. The way FLs were taught in this secondary school was in fact very similar to how I had learned English for years in France, which for me at least had proven not to be effective at all.

## 1.1 Research Context

Unlike the French educational system and unlike many others in the world, the Dutch educational system is decentralized. It means that although the government determines the subjects to be taught, the learning objectives and 50% of the final examinations, no specific pedagogical approach is prescribed. Therefore, schools, departments and sometimes even individual teachers are free to decide how they wish to translate the learning objectives into their teaching and assessment practices, and into the remaining 50% of the final examinations. The advantage of this autonomy is that teachers can choose the pedagogical approach that suits their classroom practice and beliefs best, but the challenge is that both external and conceptual factors can influence their pedagogical choices (Scheerens, 2016).

As a young teacher, I found it very difficult to question the way lessons were taught in my school, let alone the way students were assessed. This was not only due to my lack of experience and influence but also to school requirements and agreements within the language department, especially regarding assessments. The pedagogical choices made when it came to assessments were often based on practical considerations, such as the use of ready-made assessments from the textbook, but also on some of my colleagues'

more traditional views on language teaching. As an individual teacher I could introduce more authentic materials and communicative activities in my lessons, but I also had to keep in mind the less communicative nature of the assessments and spend time on assessment preparation to avoid frustration from both students and their parents.

When I became a teacher trainer, I saw my students struggling to introduce CLT in their practice in the same way I did. As I read about the issue, I realized that this problem was well known and faced by many other language teachers around the world (Dos Santos, 2020; Kissau et al., 2015; Little, 2007). A large number of studies has been conducted to understand why introducing CLT in practice proves to be so difficult (Ahmed, 2016; Chang, 2011; Coskun, 2011; Nishino, 2011; Sarab et al., 2016; Zhu & Shu, 2017). Studies often identified external factors contributing to this difficulty, such as the large number of students per class, the limited number of teaching hours, the lack of suitable materials available, and traditional examinations. To deal with some of these external parameters, almost all language teachers in the Netherlands use a textbook to help them prepare their lessons. In fact, these textbooks have been shown to determine Dutch FL teachers' curricula to a large extent (Fasoglio et al., 2015). Whereas recent textbooks offer a lot of authentic and interactive materials, in line with the communicative approach, they also still include many traditional activities such as grammar-focused gap-fill exercises. Moreover, productive language skills, and speaking skills in particular, receive relatively little attention in textbook materials (Van Batenburg, 2019). Additionally, assessments included in the textbooks do not often assess students' ability to use the FL in a communicative way. Instead, most of the textbooks that are widely used in the Dutch context include assessments that tend to mainly assess knowledge of grammar rules and translation of words and sentences out of context. As the more authentic and meaningful tasks that are nowadays included in textbooks are not or minimally included in assessments, these more communicatively oriented activities are often considered optional by teachers and students alike.

## 1.2 Research Background

Today, the general consensus is that the goal of FL education is to learn to communicate effectively. The ability to communicate in another language enables a better understanding of other cultures and is essential in the multicultural society we live in (Savignon, 2017). In order to help students to acquire this communicative competence, language teaching should be communicatively focused, providing students with enough opportunities to practice FL skills in authentic situations. This observation was made more than 50 years ago by Hymes (1972), and many studies have since then shown the effectiveness of CLT (Savignon, 2017). For this reason, in most educational

systems today, including the Dutch one, the learning objectives of FL curricula are communicative in nature, and formulated based on the CEFR (Council of Europe, 2020).

In line with these objectives, assessments should be communicative as well. As we already mentioned, however, external factors – such as the textbooks used – can play an important role in teachers' pedagogical choices. Such factors can push teachers to use assessments that are not communicative (enough). Moreover, more conceptual factors, such as teachers' lack of knowledge and expertise on how to teach and assess in a communicative way, have also been identified as affecting teachers' pedagogical choices (Graus & Coppens, 2016; Tang et al., 2012; Wang & Cheng, 2009). This lack of knowledge and expertise can be due to a lack of consensus on how to interpret CLT and on how it should be introduced into a specific educational context to teach and assess students, as multiple interpretations exist in the literature (Spada, 2014).

Howatt (1984) originally identified two versions of CLT. The first one, called the *strong version*, is based on Krashen's (1984) natural approach of language teaching. According to this version, FL teaching should reflect the natural way of learning a (first) language, by providing learners a lot of input and interaction opportunities in real-life situations, without affording particular attention to grammar rules and other formal aspects of the language to be learned. In the other interpretation of CLT, called the *weak version* by Howatt, communicative language teaching practices should be supported by instruction of formal aspects of the language. However, this should always be done to support performance in a communicative situation or task. To clarify the different ways language form can be taken into account in FL teaching, Long (1991) made the distinction between *focus-on-forms*, which corresponds to the systematic treatment of formal aspects of the language prior to the development of language skills, and *focus-on-form*, which corresponds to the more integrated approach as intended in the weak version of CLT. To be effective, focus-on-form can be planned as well as unplanned and implicit as well as explicit but should always be provided in the context of a communicative situation or task (Ellis, 2015).

As shown in several studies (Genesee, 2004; Lightbown et al., 2002), what version of CLT is effective depends on the teaching context. For the strong approach to be successful a lot of exposure to the target language is required, which can best be achieved in situations of immersion or when many teaching hours are available. However, the majority of secondary schools in the Netherlands teach FLs as separate subjects. The teaching of two FLs in addition to English is compulsory in lower form (i.e., first three years) and becomes optional in upper form of Dutch secondary education. As in most secondary school systems, the number of hours allocated to FL subjects is limited, and for FLs other than English, there is little opportunity to hear and practice these languages outside of

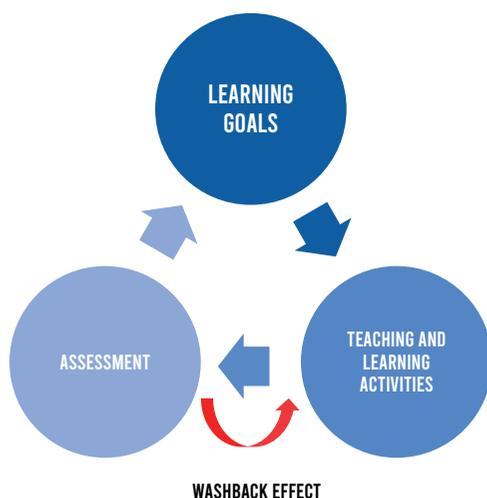
the classroom. For these reasons, the weak interpretation of CLT has been most often adopted in Dutch secondary school FL programs. Moreover, this interpretation was more similar to the traditional way of teaching languages and therefore more recognizable for FL teachers, as it is closer to the way they learned FLs themselves at school (Graus & Coppen, 2016). Yet, the weak interpretation has often been misunderstood and is not always translated well into teaching practices, still resulting in a focus on forms instead of on form. Indeed, in FL education in the Netherlands, and particularly in the lower form, much emphasis is often placed on the knowledge of formal language aspects, such as lists of grammar rules to be learned, prior to the development of language skills (Moonen et al., 2013; Westhoff, 2007). However, many studies have shown that it is essential to integrate instruction of formal aspects with the development of the language skills within authentic and realistic situations (Ellis, 2015; Lightbown & Spada, 2013; Spada & Tomita, 2010).

### 1.3 Research Focus

Several reasons can explain why the actual teaching practice is not always sufficiently communicative in relation to the communicative objectives. However, assessment in particular seems to play a crucial role, exerting an important influence on teachers' pedagogical choices and teaching practices, as many studies conducted in different educational contexts have shown (Chan, 2020; Rea-Dickins & Scott, 2007; Spratt, 2005). This influence - also known as *washback effect* (Alderson & Wall, 1993; Green, 2007) - can be negative when the assessments provided do not reflect the learning objectives. However, when assessments are well aligned with the objectives, this washback effect can also be positive. Whereas earlier studies on such washback effects have often focused on centralized educational systems in which large-scale assessments are determined at the national level and developed by testing experts, fewer studies have investigated the influence of classroom-based assessments (CBA), developed by language teachers themselves, on their own teaching practices (Muñoz & Alvarez, 2010).

In the decentralized Dutch educational system, secondary school teachers, particularly in the lower form, are asked to select, modify, develop, and administer CBA as part of their daily classroom practice. In potential, introducing communicative CBA could have a positive washback effect and steer teachers' pedagogical choices towards a more communicative pedagogical approach (learning and teaching activities) - as such ensuring better consistency in curriculum design (i.e., *constructive alignment*; Biggs, 1996; see Figure 1.1). However, the development and introduction of a communicative CBA program is a complex and challenging process which requires thorough knowledge of assessment methods. Assessing students' communicative competence asks for a wider

range of assessment procedures than the development and administration of more knowledge-based assessments (Stoynoff, 2012). To assess their students' communicative competence, FL teachers need to, for example, facilitate interaction in authentic situations and take into account the many factors that can affect performance in communicative tasks, such as students' socio-cultural background, thematic knowledge or emotional state (Bachman, 2002). Previous research has shown that FL teachers often do not feel sufficiently trained to successfully complete this complex task (Fulcher, 2012; Jin, 2010; Vogt & Tsagari, 2014).



**Figure 1.1** Constructive alignment, with washback effect (based on Biggs, 1996)

What makes the task even more challenging today is that the way assessment is conceived has changed in recent years. The emphasis is no longer exclusively on the summative function of assessment, aimed at establishing the level of students' knowledge and skills at the end of the learning process, but also increasingly on the formative function. Not being part of a formal decision, formative assessment activities offer students the opportunity to practice without formal consequences for their school or future academic career and provide both students and teachers insight into the learning process (Black & Wiliam, 1998; Dixson & Worrell, 2016). This paradigm shift has led to reflection on how FLs should be assessed in secondary classroom practice, whether summatively, formatively or through a combination of both. As Black and Wiliam (2018) point out, every assessment, even assessments that primarily serve a summative function, should also help students learn (i.e., have a formative purpose). As such, an effective communicative CBA program should include not only summative assessments, but also formative assessment activities, with the ultimate aim of both types of assessment being to enhance students' FL learning. In the Netherlands, secondary

school teachers are becoming increasingly aware of the importance of the formative function of assessments. However, the actual introduction of formative assessment activities into practice is still limited and emphasis remains on the summative function of assessments (Krijgsman et al., 2023; National Institute for Curriculum Development [SLO], 2015; Commissie Kwaliteit Schoolexaminering, 2019).

The present design-based research project investigates which tools FL teachers need to be able to implement a communicative CBA program with an adequate balance between informal and formal assessments. The ultimate aim is to provide insights into if and how the implementation of such a CBA program can lead to concrete changes in FL teaching practices and thus create more alignment in FL curricula and help improve students' FL language skills. To achieve these goals, an exploratory study was first conducted to map the actual degree of constructive alignment in current FL curricula in Dutch lower form secondary education and to gather insights into the factors influencing FL teachers' pedagogical choices (Chapter 2). A design-based study (Chapter 3) was then conducted in collaboration with FL teachers to develop tools for the development and introduction of a communicative CBA program, taking into account the factors identified in the previous study. Finally, an implementation study (Chapter 4) and an effect study (Chapter 5) were conducted to determine 1) how FL teachers used the tools to develop and introduce a CBA program in their own practice, and 2) which effects the implementation of a communicative CBA program had on their teaching practices.

## 1.4 Structure Overview

This dissertation contains six chapters. **Chapter 2** explores the current degree of alignment in Dutch FL curricula and the factors influencing teachers' pedagogical choices. For this study, the following research questions were raised:

*RQ2.1: What is the degree of alignment between communicative learning goals, pedagogical approaches, and assessment practices in FL curricula of lower form secondary education in the Netherlands?*

*RQ2.2: Which external and conceptual factors do teachers report to be of influence on the observed degree of alignment?*

In order to map out the current situation in Dutch secondary education regarding the degree of alignment between communicative learning objectives on the one hand and the teaching activities and assessments on the other, different methods have been used: lesson observations ( $N = 31$ ), analysis of formal assessment materials (used for sum-

mative purposes), teacher interviews ( $N = 21$ ) and a student survey ( $N = 473$ ). Teacher interviews ( $N = 21$ ) were also used to determine the underlying factors explaining the observed degree of alignment.

Results of Chapter 2 served as input for the study reported in **Chapter 3** that corresponds to the design phase of the research project. This chapter reports on the challenges faced by FL teachers in the design and implementation of a communicative CBA program and on the suggestions made to overcome them. This inventory of challenges and suggestions then served as input for the formulation of final design principles, which were subsequently translated into a set of tools aimed at helping FL teachers to develop communicative CBA programs suitable for their own practice. The research questions guiding this study are:

*RQ3.1: What are the challenges regarding feasibility and validity that FL teachers face when developing and implementing a communicative classroom-based assessment program and which suggestions can be made to overcome them?*

*RQ3.2: What are the characteristics of a feasible communicative classroom-based assessment program for foreign languages in lower form secondary education in the Netherlands, and how can these be translated into a set of tools for the implementation of communicative CBA programs?*

A group of 21 FL teachers of the most commonly taught FLs in Dutch secondary schools (i.e., English, French, German, and Spanish) from 15 different schools took part in a Professional Learning Community (PLC) with the aim of formulating design principles and developing tools to help other FL teachers in different school settings with the implementation of a valid and feasible CBA program. The PLC participants took part in eight interactive working sessions, organized and supervised by the researcher. Notes on the proceedings of each PLC session as well as the exchanges, remarks, and questioning of the participants were reported by the researcher in a logbook. Feedback from the participants on the try-outs were used to improve the design.

**Chapter 4** reports on how the tools designed in Chapter 3 were used by 32 other FL teachers of 14 language departments from 10 different schools to implement their own communicative CBA program, and on how teachers and students experienced this implementation. This study addressed the following research questions:

*RQ4.1: How is a communicative classroom-based assessment program implemented by language departments in Dutch lower form secondary education based on guidelines and tools made available to them?*

*RQ4.2: How is the implementation experienced by the teachers and students involved?*

Observations were used to determine whether the CBA programs had been developed and introduced as intended (i.e., met the design principles). Teachers' and students' perspectives on the implementation were then investigated through logbooks ( $N = 14$ ), interviews ( $N = 8$ ), and a student survey ( $N = 423$ ).

**Chapter 5** investigates which effects the implementation of communicative CBA programs had on teaching practices in the 10 schools participating in this last phase of the project. The research question guiding this study is:

*RQ5: What are the observed and perceived effects of the structural implementation of a communicative CBA program on FL teaching practices in terms of pedagogical choices?*

This study compared the teaching practices in regular classes (i.e., classes with a non-communicative CBA program) with the teaching practices in intervention classes (i.e., classes in which a communicative CBA program was implemented). Data were collected through lesson observations ( $N = 24$ ), logbooks from the team of teachers of each language department participating in the study ( $N = 14$ ), and student surveys in regular classes ( $N = 330$ ) and in intervention classes ( $N = 423$ ).

**Chapter 6** summarizes and discusses the main project findings and provides suggestions for educational practice.

Chapters 2, 3, 4, and 5 were written as stand-alone articles for publication in academic journals and can thus be read independently. For this reason, there may be some overlap when reading the entire dissertation.

**CHAPTER 2**



# Constructive Alignment in Foreign Language Curricula: An Exploration of Teaching and Assessment Practices in Dutch Secondary Education

This chapter is based on an article that has been published as: Rouffet, C., Van Beuningen, C., & De Graaff, R. (2023). Constructive alignment in foreign language curricula: an exploration of teaching and assessment practices in Dutch secondary education. *The Language Learning Journal*, 51(3), 344-358.

## Abstract

While Communicative Language Teaching (CLT) has gained worldwide recognition as an effective approach to language teaching, its implementation in foreign language (FL) classrooms remains difficult. In the Netherlands, communicative learning goals have been formulated at the national level but are not always reflected in daily FL teaching and assessment practices. As constructive alignment between learning goals and teaching and assessment approaches is a precondition for effective teaching, it is important to gain a deeper understanding of the degree of alignment in Dutch FL curricula and the factors influencing it. The current study therefore aims to take a systematic inventory of teaching and assessment practices in lower form FL classrooms regarding the translation of national communicative learning goals into learning activities and assessments. Findings revealed that teaching activities and classroom assessments predominantly focused on knowledge of grammar rules and vocabulary out of context and, to a lesser extent, on reading skills. External factors, such as teaching and assessment materials available, and conceptual factors, such as teachers' own conceptions of language learning, were identified to contribute to a lack of alignment in Dutch FL curricula. Assessments in particular seem to exert a negative washback effect, thereby impeding effective implementation of CLT.

## 2.1 Introduction

Communicative language teaching (CLT) is nowadays widely accepted and recognized as a productive pedagogical approach in foreign language (FL) education (Richards, 2006). CLT aims to promote the development of communicative language skills in realistic situations (Littlewood & Yu, 2011). The approach emphasizes the learning of language through meaningful spoken and written interactions, with an appropriate degree of spontaneous production, rather than through memorization and the learning of grammatical rules out of context (i.e., *focus-on-forms*, Long, 1991). These formal aspects are certainly present in the communicative approach, but they are to be addressed within a meaningful context and always in combination with a focus on the development of communicative skills (*focus-on-form*; Long, 1991). Table 2.1 below summarizes the main characteristics of the communicative approach.

**Table 2.1** Main characteristics of CLT (based on Larsen-Freeman & Anderson, 2011)

CLT principles
Use of the target language in meaningful spoken and written interactions: the context in which a communicative activity takes place should be connected to students' experiences, interests, and/or imagination.
Use of a functional grammar approach. Attention to form aspects of language (grammar, vocabulary, pronunciation), should be embedded in meaningful, communicative activities.
Use of functional language activities integrating different language skills, with an appropriate degree of free production, which prepare learners for communicating in real-life situations.
Focus on the development of strategies to help students perform language tasks and/or to solve problems in communication.
Use of authentic materials instead of isolated representative sentences out of context.
Use of appropriate scaffolding that enables students to progress in their language development (regarding form and/or content).

Although language teachers today recognize the relevance of teaching languages in a communicative way, they do not always succeed in introducing this approach into their daily teaching practice (Little, 2007; Kissau et al., 2015). This situation seems also to occur in the Netherlands and particularly in lower form (Moonen et al., 2013; Westhoff, 2007), which is the context in which the current study took place.

### 2.1.1 FL Teaching in the Dutch Educational System

The Dutch educational system is partly decentralized. The role of the government is limited to the determination of the subjects to be taught, the setting of attainment targets and the conception of final examinations. FL teaching in the Netherlands is CLT-oriented. Official curriculum documentation (College voor Toetsen en Examens, 2019; Meijer & Fasoglio, 2007) as well as standards for language teacher training (10voordel-

eraar, 2017), mention the communicative approach indirectly by referring to the CEFR and CEFR-levels: students should be able to use different language skills in authentic situations in order to communicate effectively in a foreign language at a specific CEFR level depending on the class they are in and the language they are learning. In a recent proposal for updating the FL curriculum, elaborated by teachers, school leaders and experts, CLT principles are also mentioned only indirectly, by referring to the CEFR. The proposal does stress the importance, however, of paying more attention to the development of productive skills and social interaction (curriculum.nu, 2019).

The present study focuses on contexts where FLs are taught as separate subjects<sup>1</sup>. Foreign languages are compulsory subjects within the first three years of Dutch secondary education (for pre-vocational and pre-academic streams). In upper form, students can choose to (dis)continue studying other foreign languages aside from English, which remains a compulsory subject. The final examination consists of two equally weighted parts, a national reading examination developed by the Dutch National Testing Institute (CITO), and school examinations developed and assessed by the schools themselves, covering all other skills. In lower form secondary education, all assessment formats and content are chosen by (language) departments of individual schools. The government inspects schools on the achievement of the attainment targets but does not prescribe pedagogical approaches or textbooks and only provides suggestions for the number of teaching hours per subject. Therefore, schools, many of which are semi-private, can choose their own teaching approach. Furthermore, different FL departments within these schools make their own plan on how to integrate attainment targets. Finally, at the classroom level individual teachers also make their own choices on how to translate the national learning goals into daily teaching and assessment classroom practices. As such, interpretations of the general targets can differ from one level to another, and from one school to another (Scheerens, 2016).

Although learning objectives at the national level are formulated from a communicative perspective, FL teachers, particularly in lower form, tend to focus more on knowledge of grammar rules, vocabulary and chunks taught out of context in teaching and assessment practices than on the development of language skills (Fasoglio et al., 2015; West & Verspoor, 2016). In upper form education teachers focus heavily on reading comprehension, which is the only skill assessed in nationwide exams and which constitutes 50% of the final grade (Fasoglio et al., 2015). As a result, students who perform well during their school career and up to their final exams may still have difficulties or even be unable to

---

<sup>1</sup> Over 100 schools in the Netherlands offer bilingual programs in which English is used as the medium of instruction for half of the subjects, which are taught according to the principles of content and language integrated learning or CLIL (Michel et al., 2021). Although CLIL provides an interesting example of communicative content-based teaching for foreign languages, this is not the focus of this study, as the majority of schools in the Netherlands teach FLs as separate subjects.

communicate in the foreign languages they have learned (Schnabel et al., 2016). It is important to understand why the introduction of the communicative approach appears to be difficult in FL teaching in the Netherlands.

### 2.1.2 Implementation of CLT: Influencing Factors

In other educational contexts, a number of studies (Ahmed, 2016; Chang, 2011; Coskun, 2011; Kissau et al., 2015; Nishino, 2011; Sarab et al., 2016; Zhu & Shu, 2017) sought to identify factors explaining the difficulty in integrating a communicative approach into daily teaching practice. Although such factors are always context dependent to a certain extent, two broad categories of similar influential factors emerged from the different studies: *external factors* related to organizational issues, policies and (lack of) school support, and *conceptual factors*, related to teachers' training and experiences shaping their conception of CLT. Coskun's (2011) qualitative study, for example, conducted in the Turkish context and Sarab and colleagues' study (2016) in the Iranian context both report that, from a teachers' perspective, challenges in the implementation of CLT mostly relate to external factors, such as large class size, lack of time to develop communicative teaching resources and traditional grammar-based examinations. Chang (2011) furthermore interviewed eight teachers from two universities in southern Taiwan on their perceptions and experiences regarding the integration of CLT into their curriculum in order to determine which factors promote or hinder this integration. She identified four groups of factors explaining the difficulty of introducing CLT, two types of external factors (i.e., educational system factors and CLT support), and two types of conceptual factors (i.e., teacher factors as well as student factors). The study shows, for example, that schools and national educational systems do not facilitate CLT integration as they place students in overcrowded classes with insufficient contact hours and as they impose exams that are not communicative (external factors). Her study also shows that teachers are not sufficiently trained to apply a communicative approach in practice and that students need to grow accustomed to this way of teaching (conceptual factors). In another study, Wang and Cheng (2009) explore the problems faced by language teachers in China in the transition to CLT. The necessary changes that Wang and Cheng identify are also related to both external (materials available, introduction of other types of activities and assessment) as well as conceptual factors (traditional view on language learning). According to Wang and Cheng, teachers need to change the way they think about FL learning and teaching. They need to move from a conception based on the acquisition of knowledge to a conception based on the development of language skills. Wang and Cheng also stress the importance of teachers having a clear understanding of what CLT is and knowing how to apply it in practice.

Indeed, a large number of studies point out the gap between teachers' beliefs about CLT on the one hand and the implementation of this approach in their teaching practice on

the other. They indicate that many teachers who think they are using a communicative approach do not have a clear definition of it (Burke, 2011; Richards, 2006). This may be related to the fact that - despite the attention given to the CLT approach in the literature - language teachers may not have received sufficient and/or appropriate training in their teacher education programs on the effective application of such an approach (Graus & Coppen, 2016; Savignon, 2017). Besides, student teachers' own conceptions of language teaching also play a role in the way they interpret and later introduce CLT principles in their classrooms. This conception is shaped both by the way in which the pre-service teachers themselves have been taught foreign languages in secondary education, but also by their tutors in the schools during their internships, their colleagues and the textbooks they use (Graus & Coppen, 2017).

All these external and conceptual factors affect the way CLT objectives are translated into lesson goals, learning activities and assessments. They can lead to a lack of constructive alignment which may in turn explain the difficulty of introducing CLT in practice.

### **2.1.3 Constructive Alignment: The Crucial Role of Assessments**

Constructive alignment, an educational concept developed by Biggs (1996), highlights the importance of aligning learning objectives with learning activities and assessment practices in a curriculum to enhance students' learning and achievement. Biggs and Tang (2007) defined four characteristics that a curriculum should meet to be constructively aligned:

- Learning objectives must be clearly defined and specified in learning activities.
- The learning environment should provide activities that enable students to achieve the learning objectives.
- The assessment should match and cover the learning objectives.
- Student performance should be assessed using rating scales that bring together criteria related to the learning objectives.

In the development of an aligned curriculum, Biggs emphasizes the importance of identifying the objectives that students need to achieve in assessments before organizing teaching and learning activities. Assessment practices indeed play a crucial role in constructive alignment, as they are known to influence both teaching and learning, known as *washback* (Alderson & Wall, 1993; Green, 2007). This washback can be positive when assessments are aligned with learning objectives but can be negative when alignment is lacking.

If the learning objective is foreign language communication, assessments and learning activities are aligned when they are communicative in nature. In contrast to traditional

assessment formats that assess FL skills indirectly (e.g., multiple choice questions in which knowledge of specific vocabulary or grammar rules are assessed), communicative assessments aim to provide a more direct evaluation of students' general FL ability. As such, communicative assessment tasks should be sufficiently authentic and diverse to match the variety and complexity of real-life communicative situations (Morrow, 2018).

In (partly) decentralized educational systems such as the Dutch one, schools and teachers have a lot of control over assessments which would, in principle, grant them ample opportunity to generate positive washback (Hakim, 2018). However, as Fasoglio and colleagues (2015) and West and Verspoor (2016) have indicated, FL assessments developed by individual schools are not always communicative in nature, and there is considerable variation in formats and content among schools and even language departments within the same school. This situation could explain the difficulty of implementing communicative teaching; if assessments are not sufficiently communicative, learning activities will not be sufficiently communicative either, because teachers want to prepare their students for the assessments (i.e., negative washback).

#### **2.1.4 Current Study**

There currently seem to be no studies available in the Dutch context that systematically investigate the coherence between the different elements of FL curricula, nor studies that seek to provide insight into the reasons and arguments underlying teachers' decisions in using non- or less communicative activities and assessments, particularly in lower form of secondary education. As discussed above, both external and conceptual factors affecting teachers' pedagogical choices can create a lack of alignment between CLT learning goals on the one hand, and non-CLT teaching and assessment approaches on the other hand. It is important to identify which factors affect pedagogical choices at classroom level in a specific context in order to seek solutions to enable teachers in that context to teach and assess in a communicative way, and as such allow them to align their curriculum more effectively. The current study does so for the Dutch context, by exploring the degree of constructive alignment in Dutch FL curricula and the factors affecting teachers' pedagogical choices regarding the translation of communicative learning goals into learning activities and assessments.

As such, the research questions guiding the current study are:

*RQ2.1: What is the degree of alignment between communicative learning goals, pedagogical approaches, and assessment practices in FL curricula of lower form secondary education in the Netherlands?*

*RQ2.2: Which external and conceptual factors do teachers report to be of influence on the observed degree of alignment?*

A variety of measurement tools were used to gather both quantitative and qualitative data on the various components of the curriculum in specific schools and for different languages taught at A2 level. The A2 level was chosen because it is the most common level to be attained for the different languages taught in lower form in the Netherlands.

## **2.2 Method**

First, classroom observations ( $N = 31$ ) were carried out to determine the extent to which national communicative objectives were reflected in lesson goals and learning activities. Assessment programs and assessment materials were then analyzed to establish their degree of alignment with the other two components of the curriculum (RQ2.1). A survey among students ( $N = 473$ ) was conducted to complete the inventory and provide a student perspective on classroom practices regarding learning objectives, learning activities and assessments. Interviews ( $N = 21$ ) were then carried out to gain more in-depth insight into the factors influencing FL teachers' pedagogical choices regarding the implementation of CLT goals in their teaching and assessment practices (RQ2.2).

### **2.2.1 Participants**

#### *Teachers*

A group of FL teachers from different schools in the Netherlands were recruited for classroom observations, supply of assessment materials and interviews. Thirty-two teachers from 14 different schools and 18 language departments voluntarily participated. The participating teachers were all certified, in-service, and taught at least one class at A2 level of the CEFR. They had varied years of experience ranging from one to 25 years. The four most widely taught languages in Dutch secondary schools were represented in the participant group: two English teachers, 17 French teachers, nine German teachers and three Spanish teachers. This apportionment does not reflect the distribution of different FLs in the Dutch educational system (English being the most widely taught FL in the Netherlands) but can be explained by the fact that this is a convenience sample from the network of the researcher, a French teacher educator. Teachers from 10 participating language departments (three German departments, three French, two Spanish and two English) supplied assessment materials.

#### *Students*

Students from A2-level classes of the participating teachers anonymously filled in a digital survey halfway through the school year. They were all lower form students aged

12 to 15. They filled in the survey about the language lessons taught by one of the participating teachers: English ( $n = 17$  students), French ( $n = 308$ ), German ( $n = 76$ ) and Spanish ( $n = 72$ ).

## 2.2.2 Instruments, Data Collection and Data Analysis

### *Classroom observations*

All but one of the selected teachers ( $N = 31$ ) were observed giving a lesson at level A2. Video recordings were made of the lessons observed. The teachers did not receive any prior instructions, except to teach as they normally do. They were observed using an observation tool developed by the Foreign Language Education Research Group at Utrecht University of Applied sciences (see Appendix A). This observation tool maps the extent to which the pedagogical approach of FL teachers adheres to CLT principles. The instrument focuses on observable teacher behavior and communicative teaching activities for each of the language skills. Each category, focusing on a specific language skill, includes similar items for CLT activities based on the criteria that communicative activities should meet as described in the theoretical framework: meaningfulness, creativity (appropriate degree of free production or relevant comprehension activities), authenticity, use of scaffolding, and development of strategies. The format of the instrument is based on ICALT (International Comparative Analysis of Learning and teaching) (Maulana et al., 2017; RUG, 2019). As in ICALT, each category can be scored on a four-point Likert scale representing the extent to which FL teachers implement the principles of communicative language teaching, ranging from 1 (*not seen, although the situation called for it*) to 4 (*seen to a large extent*). Additional field notes were taken during the observations. The instrument also includes an open question on lesson objectives. Teachers were asked to confirm, specify, or indicate after the classroom observation what the lesson goals were.

All observations were performed by the researcher. To guarantee inter-rater reliability five observations were conducted on-site by a second observer, namely by members of the research group, who co-developed and validated the observation tool. To check intra-rater reliability five observations were conducted a second time, a year later by the first author based on the video recordings. Scores were then compared to the observations made on-site. For both inter- and intra-rater reliability, a Cohen's  $\kappa$  analysis showed substantial agreement (inter-rater reliability:  $\kappa = .617$  (95% CI, .499 to .734),  $p < .001$ ; intra-rater reliability:  $\kappa = .616$  (95% CI, .498 to .698),  $p < .001$ ). The scores from the Likert scale were used to generate quantitative data that were analyzed descriptively for each observation category (i.e., meaningfulness, degree of free production, etc.). The additional field notes were used to interpret and illustrate the scores. Open-ended responses regarding the lesson objectives were coded into three categories: lesson goals related to the development of FL skills, lesson goals related to the development of

knowledge of grammar rules outside authentic situations (focus-on-forms) and lesson goals related to the development of content knowledge (e.g., culture, literature).

### ***Analysis of Assessment Programs and Assessment Materials***

In secondary schools in the Netherlands, classroom-based assessment programs (i.e., set of formal assessments, with a summative function, fixed over a school year in a specific grade) are developed per grade based on nationally formulated core objectives. The assessment programs for lower secondary education (A2 level) from a subsample of 10 of the 18 language departments participating in the study were compiled and analyzed in terms of the number of formal assessments assigned per year, assessment formats, assessment content in relation to language skills and/or to language knowledge, and weighting assigned to each formal assessment. In each assessment program the skills or knowledge assessed and the weighting of each summative assessment in the overall yearly average were indicated. We first inventoried which content was planned to be assessed within each formal assessment. We identified four types of content: 1) assessments addressing one specific language skill, 2) assessments addressing two or more language skills, 3) assessments combining the assessment of one or more language skills with the application of grammar rules or the translation of words out of context, and 4) assessments addressing only the application of grammar rules or the translation of words out of context. We then inventoried the frequency and weighting of the different summative assessments administered per school year, per language department.

In addition to the assessment programs, the participating teachers submitted 43 representative formal language assessments used in their schools at level A2. We recorded 88 tasks spread over the various assessments, and for eight of these tasks we also received their accompanying rating scales. We classified the different assessment tasks according to the knowledge or skills they aimed to assess, and the learning goals targeted. Assessments of language skills and the accompanying rating scales were analyzed using a checklist of criteria that communicative assessments should meet. This checklist was adapted from the list of criteria developed by the National Institute for Curriculum Development in the Netherlands (National Institute for Curriculum Development [SLO], 2013) and supplemented by the criteria defined in theory on communicative language teaching and assessments (Bachman, 2010; Brown, 2005; Green, 2013; Morrow, 2018). The checklist was divided into three categories: construct validity, content validity and evaluation criteria (rating scales). Criteria related to CLT (i.e., meaningfulness, authenticity, unpredictability, appropriate degree of free production and skills integration) could be marked as present or absent.

### ***Student Survey***

To add a student perspective to the inventory of classroom practices, participating teachers had their students complete an online survey with multiple-choice items. The questions focused on students' perceptions of learning goals, the proportion of type of knowledge and skills covered during the lessons, and the degree to which this knowledge and these skills were assessed. We first conducted a descriptive analysis of the results and examined the frequency of student responses to the various questions. We then compared the results with those obtained from the other instruments.

### ***Teacher Interviews***

Twenty-one teacher interviews were conducted. The interviews were semi-structured and included 11 open-ended questions. The interviews averaged 20 minutes in length and were recorded digitally. The audio recordings were transcribed verbatim as Word documents and pseudonymized (T1 to T21) before being deleted in accordance with the approval obtained by the Faculty Ethics Assessment Committee Humanities of Utrecht University. Teachers were interviewed individually, following their observed lessons. The main objective was to better understand the basis on which teachers' pedagogical choices regarding lesson goals, learning activities and assessments were made (RQ2.2). The first two interview questions concerned the choice of lesson goals in relation to CLT. The next three questions focused on the choice of learning activities. Finally, the last six questions concerned the choice of assessments.

Once collected and transcribed, the interview data were first analyzed interpretatively and thematically according to the steps specific to qualitative research analysis (Corbin & Strauss, 2008): categorizing the data into themes, identifying noticeable patterns, and synthesizing the categorized data into general findings. We categorized the data into the following two themes, which were identified a priori based on our literature review: conceptual factors and external factors influencing teachers' pedagogical choices. Under these themes, we then coded the data in an inductive way based on the respondents' answers (Corbin & Strauss, 1990). In this case, the codes were short sequences of words that described what kind of conceptual and/or external factors participants mentioned in their answers to the interview questions. Subsequently, codes were clustered under overarching categories (e.g., lack of time; school requirements). All themes, categories, and examples of codes within each category were then brought together in a coding scheme. The resulting coding scheme was used by the researcher to analyze each of the participants' transcripts ( $N = 21$ ). To ensure reliability, a member of the research group - English teacher, teacher trainer and PhD student - selected at random two interview transcripts and used the coding scheme to code them. Cohen's  $\kappa$  was then run to determine the agreement between the two coders. We found that there was substantial agreement ( $\kappa = .632$  (95% CI, .415 to .815),  $p < .0001$ ). The frequency with

which each of the different codes appeared in the responses given by all the teachers interviewed was analyzed to identify noticeable patterns. In order to further illustrate, clarify and explain these quantitative data, example response excerpts were selected and added to the results section.

## 2.3 Results

### 2.3.1 Degree of Alignment (RQ2.1)

The results relating to the first research question concerning the degree of alignment of Dutch FL curricula are reported below under the following curriculum components: lesson goals, learning activities and assessments in relation to CLT learning objectives.

#### *Lesson Goals and Learning Activities*

Table 2.2 shows the type of lesson goals identified in the observed lessons. Almost half of the lesson goals formulated relate to the development of language skills, while the other half focus on forms (e.g., grammar rules). Only a small proportion of the observed lessons included content related objectives, such as culture and literature.

**Table 2.2** Lesson goals in observed lessons (N = 31)

Lesson goals	Frequency	Percentage
Developing FL skills	14	45%
Developing knowledge of grammar rules and vocabulary outside authentic situations (focus on forms)	13	42%
Developing content knowledge (e.g., culture, literature)	4	13%

In everyday teaching practice, however, the focus on grammar rules and words seems to be even more accentuated. When asked directly in the interview following the classroom observation, 35% of the 21 teachers honestly admitted that the communicative focus of the lesson observed was not representative of the way they usually teach. One teacher said for instance: "This lesson is not representative of the way I teach. I have already finished the regular program and the assessments. My activities are usually mostly aimed at preparing the students for the formal assessments." T13.

Table 2.3 shows the extent to which the learning activities used during the observed lessons adhered to the criteria of CLT and, as such, could be expected to contribute to the development of communicative language skills.

**Table 2.3** Adherence to CLT criteria of learning activities in observed lessons focusing on the development of language skills ( $N = 31$ )

Language skills		Meaningful activities	Appropriate degree of free production/Use of functional comprehension activities	Authenticity	Appropriate scaffolding	Development of communicative strategies
	N	Range 1-4 $M (SD)$	$M (SD)$	$M (SD)$	$M (SD)$	$M (SD)$
Speaking	14	2.93 (.73)	1.93 (.92)	1.43 (.65)	1.79 (.90)	1.64 (.74)
Writing	9	3.11 (1.06)	2.56 (1.24)	2.11 (1.06)	2.11 (1.17)	2.33 (1.12)
Reading	13	2.69 (.75)	2.23 (1.02)	1.69 (.75)	2.00 (1.71)	2.31 (1.03)
Listening	10	3.00 (.82)	2.60 (.97)	1.60 (.70)	1.30 (.49)	1.30 (.49)
All language skills	46	2.94 (.84)	2.33 (1.04)	1.70 (1.03)	1.80 (1.19)	1.90 (.84)

1 = not seen, although the situation called for it; 2 = barely seen; 3 = sufficiently seen; 4 = seen to a large extent

The results show that although most activities were meaningful (*sufficiently seen*), they did not ask for an appropriate degree of free production or functional comprehension in accordance with the CEFR level and were often not authentic enough (*barely seen*). Students were frequently asked, for instance, to answer multiple choice questions about a text or an audio document or they had to complete strongly guided writing or speaking tasks that were not authentic, such as the translation of isolated sentences out of context. Results furthermore show that most activities, especially those focusing on listening and speaking skills, were not accompanied by sufficient scaffolding and did not focus enough on the development of communicative strategies (*barely seen*). For instance, students had to perform a speaking or listening task without any prior guidance and/or strategies. As a result (and as reported in the observer's field notes), students either failed to complete the tasks or used non-communicative strategies, such as writing and translating a dialogue, and then tried to read it with comprehensible pronunciation.

An analysis of the students' views on the knowledge and skills addressed in their FL lessons (Table 2.4) shows that their perceptions reflect our own observation that almost half of the lessons focus on grammar rules and vocabulary (see Table 2.2). Moreover, while we saw a relatively equal distribution of attention to the different language skills (i.e., listening, speaking, reading, writing) in observed proficiency promoting learning activities (see Table 2.3), according to students, activities focusing on reading skills are

far more frequent in daily teaching practice than, for example, activities promoting speaking skills (Table 2.4).

**Table 2.4** Knowledge and skills addressed most frequently in lessons according to students ( $N = 471$ )

Knowledge and skills	Frequency	Percentage
Grammar rules and words	209	44.5%
Reading skills	120	25.5%
Speaking skills	64	13%
Writing skills	36	8%
Listening skills	34	7%
Culture	8	2%

### Assessments

An analysis of the content, format, frequency and weighting (i.e., percentage of the assessment result in the final grade) of assessments present in the assessment programs of 10 language departments within the participating schools shows considerable variation between language departments. Almost all language departments summatively assess the knowledge of grammatical rules, vocabulary and chunks out of context alone or in combination with one or more language skills, but not in an integrated way.

Table 2.5 presents the average frequency and weighting per assessment format for the 10 language departments. For the majority of the language departments, assessments of the knowledge of grammatical rules and/or vocabulary and chunks out of context constitute more than 50% of the final grade. The remaining 50% includes assessments of language skills. Speaking skills are the least frequently assessed and have the lowest weighting in the determination of the final grade.

**Table 2.5** Frequency and weighting of assessments ( $N = 101$ ) within assessment programs of FL departments ( $N = 10$ )

Content	Frequency	Average weighting
Knowledge of grammatical rules, vocabulary and/or chunks out of context	36	36%
Knowledge of grammatical rules, vocabulary and/or chunks out of context combined with one or more language skills	19	40%
Development of two or more language skills	8	18%
Development of reading skills	13	26%
Development of listening skills	9	13%
Development of writing skills	8	15%
Development of speaking skills	8	11.5%

Table 2.6 presents the analysis of tasks in representative FL assessments. Among the 43 assessments made available to us by the participating teachers, we counted 88 tasks which we divided into seven categories. Frequencies in Table 2.6 indicate that tasks focusing on formal aspects of the language outside a communicative context were used the most (63.7%). Within tasks assessing language skills, multiple choice reading tasks were the most frequently administrated (13.6%). Tasks assessing language skills in an open (creative) way were scarce (5.7%).

**Table 2.6** Focus of tasks ( $N = 88$ ) in representative FL formal assessments ( $N = 43$ )

Tasks	Frequency	Percentage
Application of grammar rules or vocabulary outside a communicative context	21	23.9%
Translation of words, sentences, or chunks	35	39.8%
Multiple choice reading assessment	12	13.6%
Strongly guided writing task	4	4.5%
(Recorded) pre-written and prepared dialogue or presentation	7	8%
Open (creative) task	5	5.7%
Multiple choice listening assessment	4	4.5%

Table 2.7 presents the frequency with which communicative aspects or criteria were encountered in tasks assessing language skills ( $n = 32$ ) within the FL formal assessments submitted ( $N = 43$ ). Table 2.7 shows that although most of the tasks were meaningful for the students, they were not often embedded in authentic situations, did not ask for a sufficient degree of free production in accordance with the intended CEFR level, and did not measure students' ability to communicate spontaneously by adapting language to new situations. Furthermore, only one assessed different language skills in an integrated way.

**Table 2.7** Communicative aspects present in representative FL assessment tasks assessing language skills ( $N = 32$ )

Communicative aspects	Frequency	Percentage
Meaningful communication	27	84.4%
Authentic situations	10	31.2%
Unpredictability	1	3.2%
Appropriate degree of free production	3	9.4%
Skills integration	1	3.2%

Eight of the 32 communicative tasks were evaluated with a rating scale. As reported by the teachers, the submitted rating scales were all developed intuitively. Seven of the scales were numeric, meaning teachers could give a score per category using numbers or percentages. One scale was more descriptive, using three levels corresponding to *insufficient*, *sufficient* and *good* task performance. Criteria described per level for each category were mostly quantitative (e.g., *less than five verbs are used correctly*) or appreciative (e.g., *the presentation is creative*). All scales were analytic with various categories. They all included formal aspects of the language (such as grammar, vocabulary, spelling, or pronunciation) and task completion. Only one scale mentioned other aspects, such as the use of communicative strategies. None of the scales included CEFR can-do statements.

Results from the assessment analysis coincide with students' experiences regarding formal assessments. According to students, knowledge of grammar rules and words is also the most often assessed and weighs the most heavily (Table 2.8).

**Table 2.8** Most frequently assessed knowledge or skills according to students ( $N = 470$ )

Knowledge and skills	Frequency	Percentage
Grammar rules and words	346	74%
Reading skills	75	16%
Writing skills	34	7%
speaking skills	9	2%
Listening skills	6	1%
Culture	0	0%

The results presented above show that, in general, neither the classroom activities nor the assessment tasks used in participating schools sufficiently cover the CLT learning goals as established at the national level and/or meet theoretical criteria of effective CLT implementation.

### 2.3.2 Conceptual and External Factors Influencing FL Teachers' Pedagogical Choices (RQ2.2)

Below we present the factors the teachers in this study reported to be of influence on the determination of lesson goals, learning activities and assessments in relation to CLT learning goals.

To the question on how learning goals were set for each lesson and how classroom activities were chosen, a large majority of the teachers interviewed answered that the selection of lesson goals and learning activities was based on their textbooks, either exclusively (57%) or partially (33%). Some of the teachers in this study (24%) indicate that they often determine the lesson goals themselves and that most of the learning

activities are based on the CEFR. Fewer teachers did not explicitly formulate any lesson goals (10%). A relatively substantial proportion of the teachers (43%) indicate that lesson goals were not directly linked to CLT. Table 2.9 presents the reasons reported by teachers for not formulating lesson goals and not selecting learning activities more in line with CLT.

**Table 2.9** Factors of influence on the implementation of CLT learning goals and activities reported by FL teachers in interviews

Factors	N (= 21)	Examples of explanations given by teachers
Agreements within FL departments	17	<p>"In the lower form we use the textbook, as agreed in the department. I do not like it because the communicative character disappears a bit." T10</p> <p>"I try to do something outside the textbook on a regular basis, but I have difficulty doing so because within our department it is important that we work in a consistent way. I would like to make more use of authentic texts." T8</p>
Assessment formats and content	15	<p>"Assessment format does have an influence on my choices. When the assessment is very traditional, like with a separate grammar section, I tend to explain grammar explicitly in class." T9</p> <p>"The assessments included in the textbook largely determine the way I teach. If I do something other than grammar exercises in class, then I feel I'm not preparing my students properly for the assessment." T7</p>
Time and expertise	10	<p>"I find that difficult. I work full-time, so sometimes I just have to do something. Because I don't have much time to prepare properly, and I think that it is a pity. So not all lesson goals fit and I'm not going to formulate lesson goals every lesson. I think that working outside the textbook is better, but it requires a lot of preparation and I miss then a clear line. You develop while you are doing it and then it can get a bit messy." T11</p> <p>"My problem is time. I teach for four days. My timetable is simply full. I don't have the extra time to be able to develop professionally and do things differently. I have to do that in my own time. I also work in a small department, which is also a part of the problem." T5</p>
Perceptions on language teaching	5	<p>"The emphasis, in my view, should be on grammar, because that's where almost everything depends on, so if students do that well, they will then write better. And then they will also recognize certain structures in texts, and that's why we made these choices." T21</p> <p>"I think it's important that students learn words, that they get a mark. Without marks, they won't actually learn them. It doesn't have much to do with the learning goals and that's a pity, but it is a school system." T3</p>

External factors such as agreements within FL departments, assessment formats and content, and lack of time and expertise have been frequently reported by the FL teachers interviewed as having a strong influence on the way they shape their lessons.

Some of the teachers interviewed mentioned conceptual rather than external factors of influence on their pedagogical decisions regarding formal assessments, such as the idea that it is important to teach and assess words or grammar rules outside of context first in order to help students to improve language skills later on. The reasons teachers in

this study provide for awarding marks to motivate students to learn grammar rules and vocabulary out of context indirectly reveal the value these teachers place on the prior acquisition of these skills in learning a language.

## 2.4 Discussion

The aim of this study was to draw up an inventory of FL teaching and formal assessment classroom practices in lower form education in the Netherlands. More specifically, the first objective of the study was to determine their degree of alignment with the intended, CLT oriented Dutch FL curriculum. This was done by observing how the communicative learning objectives formulated at the national level were translated into lesson goals, learning activities, and assessments by individual schools and teachers.

The results of the current study show a lack of alignment, especially regarding the national CLT goals on the one hand and the type of formal assessments used on the other hand. When zooming in on learning activities first, we see that, although an important part of the lessons observed focused on the development of language skills, learning activities provided during the lessons were often very guided and did not ask for an appropriate degree of free production in line with CLT criteria (Larsen-Freeman & Anderson, 2011). Moreover, students reported a disproportionate distribution of attention across different skills in learning activities, with a predominant focus on reading skills. This tendency has previously been reported by the Dutch National Institute for Curriculum Development (Fasoglio et al., 2015). The results of the study support this tendency and identify the CLT criteria which were least discernible in the learning activities provided by FL teachers. Although most activities were meaningful, they did not sufficiently include other CLT criteria, namely: authenticity, an appropriate degree of free production, appropriate scaffolding, and a focus on the development of communicative strategies. When looking at the assessment materials, the results show that language skills were less often assessed than knowledge of grammar rules and vocabulary and chunks out of context and weighed less in the overall average, whereas they are essential in real-life communication (Larsen-Freeman & Anderson, 2011; Morrow, 2018). In the assessment of language skills, we observed a disproportional focus on reading skills. In addition, on the rare occasions that productive language skills were assessed, it was not in a communicative way, but rather through heavily guided tasks. Furthermore, the rating scales used to evaluate these assessment tasks, if any, did not include CEFR can-do statements at A2 level. This can be taken as another indicator of the misalignment between assessment practices on the one hand and CLT principles on the other.

We can conclude that we see a lack of alignment between the communicative learning objectives formulated at the national level and the learning activities and assessments observed and reported by the FL teachers in this study, which do not (sufficiently) cover the communicative learning objectives. Results from the student survey further confirm this lack of alignment between the intended communicative objectives on the one hand and realized learning activities and the administrated assessments on the other hand.

The second objective of this study was to gain more insight into the external and conceptual factors of influence on Dutch FL teachers' pedagogical choices, to further explain the observed lack of alignment. Results show that teachers' pedagogical choices are largely influenced by agreements within their language departments regarding the use of a specific textbook and the assessments it includes. The teachers in this study also report that they lack the time and expertise to develop their own materials and that they need a textbook to guide their pedagogical choices. These external factors were also reported in other contexts (Ahmed, 2016; Chang, 2011; Coskun, 2011; Kissau et al., 2015; Nishino 2011; Sarab et al., 2016; Zhu & Shu, 2017). In addition to these external factors, results show that teachers' pedagogical choices also relate to their own conceptions of language learning, such as the importance of achieving knowledge of grammar rules and vocabulary prior to the development of language skills, and the importance of assessing this knowledge directly in order to stimulate students to practice. Whereas form-focused instruction may contribute to the development of language skills, it only does if this knowledge is developed within a meaningful communicative activity (Ellis, 2015; Long 1991; Spada, 2011). In our data, however, activities focusing on grammar rules and vocabulary and chunks were mostly provided out of context. Teachers' conceptualizations of CLT were also reported to be of influence in the studies mentioned earlier (Chang, 2011; Wang & Cheng, 2009), but they may even be of greater influence in the partly decentralized Dutch educational system where teachers have a large degree of autonomy in the selection, development and implementation of teaching and assessment materials.

Both our inventory of classroom practices as well as teachers' reports on factors influencing the degree of constructive alignment in their curricula point towards a crucial role of assessments. Given the preeminent focus on grammar, vocabulary and strongly guided language use in observed assessment practices, and the washback effect assessments are known to exert (Alderson & Wall, 1993), it is not surprising that CLT implementation remains problematic. Moreover, the observation that reading skills are addressed relatively often in both teaching and assessments in lower form FL classes might well be a washback effect of the exclusive focus on reading in the national final FL examinations in the Netherlands. The teachers themselves also pointed to assessment formats and content as a major influence on the way they shaped their teaching.

### **2.4.1 Limitations**

This study was conducted in the context of authentic classroom practice. This enhances ecological validity, but also implies that many factors may have had an (unintended) impact on the results. All the teachers participating in the project did so on a voluntary basis with the intention of working on their own professional development. This means that these teachers and the schools involved were aware of the purpose of the current study. As such, results from classroom observations and interviews may have been influenced by teachers' commitment to the project. Moreover, teachers were observed only once at a given moment in the school year and in a specific phase of their teaching program. It is a snapshot in time that cannot fully reflect the diversity of teaching and learning activities provided during a school year. The results need to be considered in this light. The student survey was introduced to triangulate or nuance the results from the classroom observations in this regard. However, it consisted of a multiple-choice survey filled in by students aged 12 to 15 during the lessons. Even if the results are likely to reflect an existing tendency due to the high number of respondents coming from different classes and schools, the reliability of the results cannot be fully guaranteed. Additional interviews could have been insightful to corroborate the survey findings.

### **2.4.2 Implication for Practice**

The results of this study indicate that although the large majority of the teachers in this study showed their intention to teach foreign languages by adopting a communicative approach, they may not succeed in doing so in an optimal way. This is largely due to external factors such as a lack of communicative materials available, a lack of time to develop lessons, and agreements made within FL language departments. The textbooks frequently used by FL teachers do often not include enough communicative activities and fail to put enough emphasis on such activities, by including formal assessment tasks that are not sufficiently communicative and mainly assess knowledge of grammar and vocabulary out of context. Ideally, publishers could place more emphasis on communicative activities and formal assessment tasks in FL textbooks. Moreover, teacher training programs should put emphasis on the why and how of aligning an FL curriculum with the national CLT learning goals. They could explicitly train prospective teachers, for example, to use textbooks selectively and focus on the communicative activities they contain, to select CLT teaching and assessment materials from other sources, and/or how to develop these themselves. In addition, in daily teaching practice teachers should be allocated more time to select, adapt or develop communicative teaching and assessment materials. Finally, language departments could be advised to collectively develop and implement communicative classroom-based assessment programs. This last point seems to be crucial since assessment influences the pedagogical choices made by teachers regarding the selection of learning activities and time spent to develop students' language skills. However, this also requires that FL teachers have

a CLT oriented conception of language learning and know how to develop a communicative classroom-based assessment program. Here again we see an important role for teacher training programs. In addition, design-based research in which teachers and researchers co-create, implement, and evaluate communicative classroom-based assessment programs aligned with CLT goals could provide insight into characteristics of effective and realistic CLT assessment, and as such, inform both FL teaching practice and teacher training.

**CHAPTER 3**



# Designing a Communicative Classroom-Based Foreign Language Assessment Program for Dutch Secondary Schools: A Design based Research Project within a Professional Learning Community

A concise version of this chapter has been published in the collated papers for the ALTE 7th International Conference, held in Madrid: Rouffet, C., Van Beuningen, C., & De Graaff, R. (2021). Designing a communicative foreign language assessment program for Dutch secondary schools: A design-based research project within a professional learning community. In *Collated Papers for the ALTE 7th International conference, Madrid* (pp. 179-183).

## Abstract

The implementation of a communicative classroom-based assessment (CBA) program is a complex task and teachers do not always have sufficient expertise to do so. In an attempt to enhance constructive alignment and to support foreign language (FL) teachers, a professional learning community consisting of the researcher and 21 FL teachers from different Dutch secondary schools worked together in co-creation. They shared the practical and conceptual challenges they faced when implementing a communicative CBA program in their own practice and sought solutions to overcome these challenges. Design principles were then formulated, and tools (formats and examples) were developed based on these principles. This paper reports on the challenges faced during the design process and present the guidelines and tools designed. Results show that practical challenges were often related to the limited time allocated to teachers to prepare lessons, and to develop or select communicative assessment activities. In addition to this practical challenge, teachers also faced conceptual challenges regarding the operationalization of communicative principles, such as unpredictability, focus-on-form, and the integration of different language skills within one assessment. Finally, teachers were also concerned about students' extrinsic motivation, with fewer assessments focusing only on language skills.

### 3.1 Introduction

The previous chapter highlighted the importance of constructive alignment, defined as aligning classroom activities and assessments with learning objectives to enable a congruent pedagogical approach. Although communicative learning objectives have been formulated at the national level for FL teaching in the Netherlands, they are not always reflected in everyday teaching and assessment practices. Assessments in particular seem to have a strong effect on the effective implementation of communicative language teaching (CLT). This *washback effect* (i.e., the effect assessments assert on teaching; Alderson & Wall, 1993; Green, 2007) can be negative when assessments are not aligned with the learning objectives but can in turn be positive when assessments and learning objectives are well aligned. Since learning objectives focus on the ability to use the foreign language (FL) in real-life communication, assessments should assess this ability through tasks that mimic authentic and realistic communicative situations (Wigglesworth, 2008). Assessments should therefore be based on communicative tasks, as recommended by the Council of Europe (2020).

The washback effect of assessments on teaching and learning activities has been extensively researched in the context of large-scale summative language assessment, partly because of the high-stakes issues involved in this type of assessment: access to education, graduation, access to citizenship, etc. (Al Hinai & Al Jardani, 2021; Rea-Dickins & Scott, 2007; Spratt, 2005). However, less attention has been paid to lower stakes FL teachers' classroom-based assessment (CBA) and its effects on teaching (Fulcher, 2012). Moreover, the way teachers plan CBA and make professional judgements about students' learning progression and achievements remains relatively unexplored (Yan & Zhang, 2018). Yet teachers are required to select, modify, develop, and administer assessments on frequent occasions, for example to decide if students can pass on to the next grade at the end of a school year. To provide concrete evidence to justify these decisions, classroom-based assessments are most often used in a summative way (Black & Wiliam, 2018). Interestingly, over the last twenty years we have seen a significant shift in the way assessment is conceived in assessment literature. The notion of formative assessment, with the aim of helping learners to progress through the various stages of their learning, has been given increased prominence over the notion of summative assessment, more oriented towards the final outcome and used for selection or judgement at the end of the learning process (Black & Wiliam, 1998).

While there are clear theoretical distinctions between summative and formative assessment, the aim in both cases should be to enhance students' learning (Black & Wiliam, 2018). As assessments with a summative function, especially when developed by teachers themselves, can also be part of instruction and be used in a formative way

to help students learn, we chose not to distinguish between summative and formative assessments as such, but between formal assessments with both summative and formative functions and informal assessments with only a formative function. An effective classroom-based assessment program designed by teachers should not only include formal assessments but also informal assessment activities (i.e., not being part of any official measurement or judgment) with the ultimate aim of both assessment types being to enhance students' learning (Dixson & Worrell, 2016; Rea-Dickins, 2001).

In light of the above, the goal of the current design-based study was to develop tools for the design of a communicative and feasible classroom-based assessment program, incorporating both formal and informal assessment activities that are aligned with CLT learning goals, thereby having the potential to generate a positive washback effect on CLT implementation. In this chapter we present the results of a collaboration project between the researcher and FL teachers.

### **3.1.1 Classroom-based Assessment**

According to Davison (2019), the term CBA refers to any assessment (formal or informal), conducted by those responsible for teaching and learning, on an ongoing basis. This process includes the collection of evidence of students' learning, the formulation of judgements about it and its use in making instructional decisions (McMillan, 2013). As mentioned earlier, all assessments, whether formal or informal, should be used to facilitate students' learning (Black & William, 2018). This formative assessment process is often described in three stages: identifying where learners need to go in their learning, where they are now and how teachers can help them to achieve the learning goals (Hattie & Timperley, 2007; Wiliam & Thompson, 2007). In the first stage, teachers need to clarify learning objectives using student-friendly language, and/or by engaging students in a discussion about learning objectives. In the second stage, teachers should gather evidence of students' learning that can take various forms, from spontaneous and informal assessment opportunities through student-teacher interactions, to more planned and formal assessment forms aimed at providing students and teachers with evidence of learning progress (Hill & McNamara, 2017). In the third stage, teachers should use assessment outcomes, whether formal or informal, to adapt their teaching to meet the needs of their students (Turner & Purpura, 2016). Apart from incorporating these three formative stages, CBA should also aim to increase students' engagement and motivation, by making use of (informal and formal) assessment tasks that students perceive to be relevant and informative to them (Shepard et al., 2018). In FL education, assessment tasks that are communicative in nature and linked to the CLT learning goals are likely to fulfill these requirements (Wigglesworth, 2008).

### 3.1.2 The Role of Language Assessment Literacy in CBA

The organization of CBA by teachers in their different classes and grades is a complex and challenging process requiring a thorough knowledge of assessment methods that can be applied in practice. CBA might even be more challenging for language teachers assessing communicative competence. Being skill-oriented, communicative CBA includes a broader range of assessment procedures than more knowledge-oriented assessments and relies less on objectively scorable item types such as multiple-choice assessments (Stoynoff, 2012).

Earlier research has shown that many FL teachers are not sufficiently trained to take on this complex CBA task (Fulcher, 2012; Jin, 2010; Vogt & Tsagari, 2014). In his study conducted among 278 language teachers, Fulcher (2012) attempted to determine the assessment training needs of FL teachers through a survey that was disseminated in different countries and contexts. This study, like Vogt and Tsagari's study (2014), conducted in seven European countries, reveals a lack of knowledge about assessment in the classroom and a need for more concrete assessment examples and models appropriate to different classroom practices. According to both studies, teacher professional development on language assessment is too theoretical and insufficiently concerned with assessment in the classroom, even though teachers are dealing with it on a daily basis. According to Fulcher (2012), the methods used in FL teacher training programs still place too much emphasis on traditional large-scale summative language assessment based on psychometric analysis. Far less emphasis is put on the assessment of communicative tasks incorporating different language skills and on the formative aspect of CBA. Although the CEFR and the manuals developed by the Council of Europe on assessments (North et al., 2009; Council of Europe, 2011) stress the importance of using real-life tasks based on an action-oriented approach, they do not provide clear guidelines on how to achieve this in the conception of CBA (Fischer, 2020). Harding (2014) indicates that individual FL teachers do indeed face challenges when designing communicative language assessments, due to practical concerns and different interpretations of the communicative approach. Vogt and Tsagari (2014) furthermore report that teachers would like to learn more about designing assessments and developing and using rating scales adapted to their own classrooms. However, in view of the relatively vague responses of Vogt and Tsagari's participants (such as learning how "marks come about" p. 390) and the fact that participants were from various contexts, the study points to the need for more qualitative, contextualized research to better understand teachers' assessment literacy and training needs. As recommended by Tsagari and Vogt (2017), it is important to identify and address the reality and constraints that influence teachers' assessment practices by combining theory and classroom practice, involving teachers in collaborative assessment development projects.

### 3.1.3 Co-design of a Communicative CBA program

There is also a growing demand from FL teachers for guidelines on how to design a CBA program for lower form secondary education providing a right balance between formal and informal assessment activities (Krijgsman et al., 2023). To meet this demand and to ground the design not only in theory, but also on the concrete experiences of FL teachers in their practice, we decided to use a co-creative approach to the design of assessment tools that can serve as a base for developing a communicative CBA program - which will be implemented in the next study (Chapter 4). Besides, as Binkhorst and colleagues (2017) point out, various studies have shown that collaboration projects within professional learning communities (PLC's), in which teachers share experiences and engage in critical reflection can ensure sustainable curriculum change, especially when different schools are involved (Chapman & Muijs, 2014; Plomp & Nieveen, 2009; Reeves, 2006; Spada & Lightbown, 2022; Stoll et al., 2006; Van den Akker, 1999). Moreover, materials developed by teachers are more likely to be successfully implemented in practice (Harn et al., 2013; Levin, 2004). PLCs with the aim of (re)designing and/or implementing (part of) a common curriculum, such as assessments, can furthermore improve teachers' pedagogical approach by allowing them the opportunity to put new knowledge and skills directly into practice while, at the same time, contributing to teachers' professional development (Voogt et al., 2011).

### 3.1.4 The Current Study

In lower form secondary education in the Netherlands, assessment programs (a set of formal assessments fixed over a school year in a specific grade) are entirely classroom-based and, as such, developed by teachers at individual schools per grade. In the inventory study conducted prior to this one (Chapter 2), FL teachers reported a lack of time and available materials (i.e., practical factors), but also a lack of expertise in developing a communicative CBA program adapted to their practice (i.e., conceptual factors). As a result, assessment activities used by these teachers, most of which had a summative function, were not always sufficiently valid (i.e., aligned with the communicative learning objectives), generating a negative washback on the implementation of communicative language teaching.

It is important to gain a deeper understanding of how the practical and conceptual challenges teachers reported (see Chapter 2) affect their choices in the selection and design of CBA, to be able to find solutions to overcome them. The first aim of the present study therefore is to further identify and specify the practical challenges (regarding feasibility) and conceptual challenges (regarding validity) that teachers face when developing and implementing a communicative CBA program. The second aim is to develop tools (i.e., formats and concrete examples), together with FL teachers, to support

the implementation of a valid and feasible communicative CBA program. The following research questions are therefore guiding the current study:

*RQ3.1: What are the challenges regarding feasibility and validity that FL teachers face when developing and implementing a communicative classroom-based assessment program and which suggestions can be made to overcome them?*

*RQ3.2: What are the characteristics of a feasible communicative classroom-based assessment program for foreign languages in lower form secondary education in the Netherlands, and how can these be translated into a set of tools for the implementation of a communicative CBA program?*

## 3.2 Method

### 3.2.1 Participants

A group of 21 FL teachers of the most commonly taught foreign languages in Dutch secondary schools (English, French, German, and Spanish) from 15 different schools took part in a Professional Learning Community (PLC) with the aim of identifying challenges they faced during the design and introduction of a communicative CBA program and finding solutions to overcome them. The participants were all qualified teachers with varied years of experience ranging from 1 to 25 years. They all taught at least one class in lower form.

### 3.2.2 Co-design Process

Participants in the PLC took part in eight interactive working sessions of three hours each. The sessions were organized and supervised by the researcher who also took part in the design process. During the eight sessions, participants shared their vision on communicative teaching and assessment practices, compared their practices with theory on communicative assessment, determined the components of a valid and feasible communicative CBA program, formulated design principles, developed tools based on it and tried them out in their practice. In the preliminary design, they determined what should be assessed, how it should be assessed, why it should be assessed, and finally when it should be assessed, following Hill's framework (2017). PLC participants answered the questions from the framework, taking into account conceptual considerations informed by theory and practical considerations informed by their own practices. The PLC participants, together with the researcher then translated these considerations into core components of a CBA program, namely communicative assessment tasks, formal assessments, informal assessment activities, rating scales, and an assessment planning (see Table 3.1).

**Table 3.1** Design process of a classroom-based assessment program

	<b>Conceptual considerations</b>	<b>Practical considerations</b>	<b>Components of a CBA program</b>
What should be assessed?	The ability to communicate in a foreign language	Language skills at A2 level of the CEFR Official attainment targets for FL (lower form)	Communicative tasks at A2 level according to CLT principles
How should it be assessed?	In line with CLT learning goals (according to the principles of constructive alignment)	Through oral and written assessments adapted to large classes	- Formal assessments based on communicative tasks - Informal assessment activities based on (parts of) communicative tasks
Why should it be assessed?	To enhance learning To inform teaching To make judgements To incite positive washback	To document growth in learning To report results to stakeholders (students, parents, school, external authorities)	Rating scales
When should it be assessed?	During the learning process (informal) At the end of the learning process (formal)	On a regular basis During the lessons Outside of the lessons	Assessment planning

All components of the CBA program should be developed in accordance with CLT principles and must be aligned with communicative learning objectives. Communicative assessment tasks can be selected or adapted to compose formal assessments suitable to the teacher's own practice. Informal assessment activities following the three stages of the formative process as described by Wiliam and Thompson (2007), should be introduced in daily teaching practice. Rating scales with success criteria should be used during the formative process and serve as a rating tool for the evaluation of formal assessments. Finally, an assessment planning for the distribution of formal and informal assessments within a school year should be included.

From the conceptual and practical considerations identified, preliminary design principles for each of the components described above were formulated first. Following these design principles, the PLC participants developed a communicative CBA program for their own classes at A2 level for a whole school year. Challenges faced by the participants during the development and introduction of their CBA programs were identified. The solutions found to overcome these challenges were taken into account to formulate

the final design principles (see Table 3.2). Based on these final principles, guidelines and tools for each of the core components of a CBA program (i.e., formats and examples) were designed and gathered in a toolbox. This toolbox aims to help FL teachers who teach at A2 level of the CEFR in lower form secondary education in the Netherlands to put together a communicative CBA program adapted to their own practice (see Appendix B). We chose to develop tools, instead of a ready-made CBA program, to meet teachers' need for flexibility and to respond to the diversity of classroom practices in the Dutch context.

The following tools were designed and gathered in the toolbox: examples of communicative assessment tasks, assessment templates (i.e., templates showing how communicative assessment tasks could be distributed within a formal assessment in relation to the learning objectives at a certain level), examples of formal assessments, a checklist to check the quality of communicative assessments, examples of informal assessment activities for each phase of the formative process, formats and examples of rating scales, and an example of an assessment planning.

**Table 3.2** Final design principles of the components of a communicative classroom-based assessment program

<b>Components</b>	<b>Content-related (validity)</b>	<b>Organizational (feasibility)</b>
<b>Communicative assessment tasks</b>	<p>Based on Brown (2005), Morrow (2018).</p> <ol style="list-style-type: none"> <li>1. Should be meaningful: connected to students' experiences and perceptions.</li> <li>2. Should be authentic: linked to real-life situations.</li> <li>3. Should be unpredictable: encourage students to respond to unprepared questions or comments.</li> <li>4. Should offer an appropriate degree of free production be open enough to enable students to show that they can give their own content to a conversation.</li> <li>5. Should Integrate language skills: encourage students to use different language skills in an integrated way, as it is often the case in real-life communication.</li> </ol>	<ol style="list-style-type: none"> <li>6. Should be easy to organize.</li> <li>7. Should correspond to common themes in FL textbooks.</li> </ol>
<b>Formal assessments</b>	<p>Based on Morrow (2018), Wigglesworth (2008), North et al. (2009).</p> <ol style="list-style-type: none"> <li>8. Should include explicit reference to CEFR can-do statements.</li> <li>9. Should include a variety of communicative tasks with an increase in difficulty.</li> </ol>	<ol style="list-style-type: none"> <li>10. Should be easy to perform during the lessons.</li> </ol>
<b>Informal assessment activities</b>	<p>Based on Hattie &amp; Timperley (2007), Wiliam and Thompson (2017).</p> <ol style="list-style-type: none"> <li>11. Should not be part of formal evaluation.</li> <li>12. Should aim to assess (part of) communicative tasks.</li> <li>13. Should include a formative goal according to one of the steps of the formative cycle (feed-up, feedback or feedforward).</li> </ol>	<ol style="list-style-type: none"> <li>14. Should be easy to organize during lessons.</li> </ol>

**Table 3.2** Final design principles of the components of a communicative classroom-based assessment program continue

<b>Components</b>	<b>Content-related (validity)</b>	<b>Organizational (feasibility)</b>
<b>Rating scales</b>	Based on Bachman and Palmer (1996), Celce-Murcia (2007), Davies (1999), Grabe & Zhang (2016), Lee (2006), Plakans (2021).	
	15. Should bring together criteria related to the learning objectives, corresponding to the can-do statements of the CEFR levels.	19. Should be easy to use to correct and to score assessments.
	16. Should value the use of communicative strategies.	
	17. Should enable the assessment of several language skills used in an integrated way.	
	18. Should be able to be used as a summative and formative tool.	
<b>Assessment planning</b>	Based on Dixon and Worrell (2016), Rea-Dickins (2001).	
	20. Should include the assessment of all language skills in a formal and informal way.	21. Should offer a balanced distribution of the formal and informal assessments of the language skills over the school year.

### 3.2.3 Data Collection and Analysis

#### *Logbook*

Notes on the proceedings of each PLC session as well as the exchanges, remarks and questioning of the participants were reported by the researcher in a logbook. The report of each session was submitted to each of the participants individually for a member check to ensure transparency. They could complete, modify and/or confirm the report.

The qualitative logbook data were analyzed according to the steps recommended by Corbin and Strauss (2008): categorization of the data into themes, formulation of codes and identification of noticeable patterns. Two themes were identified a priori based on the results of the previous study (see Chapter 2): conceptual challenges and decisions (concerning validity) and practical challenges and decisions (concerning feasibility). Under these themes, we then inductively coded the data. Each code corresponds to short sequences of words describing the challenges participants faced during the design process and the decisions they made to overcome them. Examples of codes and logbook entries are presented in Table 3.3.

**Table 3.3** Examples of codes and logbook entries

<b>Theme</b>	<b>Code</b>	<b>Example from logbook</b>
Challenges regarding practical considerations (Feasibility)	Lack of available materials	"How can this approach be applied to textbook use?"
	Lack of time	"In practice you always have too little time to assess speaking skills in a communicative way."
Challenges regarding conceptual considerations (Validity)	How to integrate (all) language skills	"How do we ensure that the receptive skills are sufficiently assessed in an integrated way with a rating scale?"
	How to introduce unpredictability	"How do we prevent the use of a memorized text?"

#### *Stakeholder feedback and try-outs*

The newly designed formal assessments and rating scales were submitted to different stakeholders (colleagues of the PLC participants and their students) for try-outs and feedback. Five teachers (T1 to T5) from five different language departments from the participating schools took part in these try-outs organized with a selected group of students outside of school time. After the try-outs, the teachers in each of the language departments involved asked their students questions about how they experienced the newly designed assessment materials regarding validity and feasibility. They then filled

in a short online questionnaire in which they reported on their students' as well as their own experiences regarding the validity and feasibility of the assessment materials.

Additionally, four pupils from one school took part in a two-hour session supervised by the researcher in which they gave verbal feedback on the rating scales. Notes on the feedback session were taken by the researcher. The feedback of the different stakeholders was gathered and analyzed, first to determine how valid and feasible the new formal assessments and rating scales were perceived and then to identify other practical and conceptual challenges faced by teachers in the incidental use of the assessment materials. Teachers' answers to the questionnaire and the notes about students' perspectives taken by the researcher were first categorized into two themes: one related to validity and another to feasibility. The data were then coded inductively.

## 3.3 Results

### 3.3.1 Challenges and Design Decisions regarding Validity

#### *Results from the logbook*

Three challenges regarding the validity of communicative assessments were identified:

1. How to introduce unpredictability
2. How to integrate grammar and vocabulary
3. How to integrate different language skills within an assessment

To ensure a higher degree of communicative validity in the design of assessment materials (in order to stimulate positive washback), we made decisions within the PLC to overcome these challenges.

#### *Introduction of unpredictability*

Teachers were concerned about the idea of introducing unpredictability in their assessments to enhance validity, because they did not know how to prepare their students to react spontaneously and adequately to unprepared questions and situations. To overcome this challenge, we decided to design assessment tasks that focus on students' ability to perform speech acts which are present in a variety of realistic situations (reporting, corresponding, sharing experience and information, expressing opinions, etc.). These speech acts as well as the themes covered in their textbooks (sports, holidays, school, etc.) are known to students in advance, but in the assessments, students have to perform one of the acts within a new situation. This increases unpredictability while still providing sufficient guidance to prepare students for assessments.

### ***Integration of grammar and vocabulary***

A large number of the participating teachers were still assessing knowledge of grammar and vocabulary and chunks out of context. It was challenging for them to discard this type of assessment and replace it with the assessment of language skills alone. They were afraid that students would no longer learn enough vocabulary, chunks, and grammar rules if these were not assessed directly. To do these concerns justice, we decided to create two categories in the rating scales aiming to assess the use of vocabulary and grammar within the context of a communicative act. In addition, the PLC insisted on the importance of incorporating informal assessment activities, addressing the development of knowledge of vocabulary and grammar rules in the context of communicative learning activities.

### ***Integration of language skills***

The last challenge faced by the teachers was the integration of different language skills within an assessment to enhance authenticity and validity. Teachers were concerned by the fact that one of the skills could be underrepresented or that students could compensate one skill with another, which would not give a clear idea of the level of each skill independently. In order to enhance validity on that point, the assessment template that we developed indicates how assessment tasks should be distributed between the different skills within an integrative assessment.

### ***Results from feedback on newly designed assessment materials***

All of the teachers in the five language departments agreed that the formal assessments developed within the PLC were covering the A2 can-do statements and were therefore at the right level and that the rating scales measured the learning goals present in the assessments. One of the teachers (T2) specified that the descriptions of the different criteria were clear and that they matched the assessment tasks. Another one (T1) mentioned that students appreciated that the tasks were realistic and meaningful to them. However, two teachers (T3 and T5) indicated that for some of their students the formal assessments were too difficult or not challenging enough and therefore demotivating. To overcome this challenge and in order to be able to evaluate a student in a positive way with the CEFR criteria, the PLC suggested including a progression in the tasks to be accomplished within an assessment. For a formal assessment at level A2, in addition to the tasks corresponding to that level, a task at level A1 and a task at level A2+ should be included in order to indicate in the scoring what a student is able or not yet able to do.

## **3.3.2 Challenges and Design Decisions regarding Feasibility**

### ***Results from the logbook***

Teachers identified three challenges regarding the implementation of communicative assessment in their teaching practice.

1. Lack of communicative assessment tasks in available materials
2. Lack of time allocated for assessment administration and scoring
3. Students' lack of experience with communicative assessments

In response to each of these concerns, decisions were made within the PLC to enhance feasibility of the assessment tools to be designed. The different suggestions to overcome these challenges are reported below.

### ***Lack of materials available***

Results from the logbook showed that the majority of the PLC participants used a textbook and did not have much time to develop their own assessment tasks in addition to it. Their first concern in the design of communicative assessments was to be able to keep using their textbooks while preparing their lessons. Most of the learning activities and assessments from the textbooks used were organized thematically (e.g., sports, school, holidays) but focused mainly on knowledge of vocabulary, chunks and grammar rules out of context. Communicative tasks were included but did not constitute the main focus.

Below are a few examples of participants' questions and remarks concerning the use of a textbook in combination with communicative assessment:

"Working even more with a realistic context is a nice aim, but how are we going to achieve it with our textbook?"

"How can CLT be applied with the use of a textbook?"

"I think it is important that the formal assessments match the textbook that supports the preparation during the lessons."

To overcome this problem, we decided to develop examples of communicative assessment tasks that could be adapted to the different themes covered in the textbooks that teachers use. To that end, assessment tasks that were designed within the PLC focused on speech acts, such as sharing experiences, corresponding online or reporting information that can be performed in the context of any theme, both in formal and informal settings in accordance with the corresponding CEFR level. In this way, all activities about a specific theme in the textbook could still be used in preparation for the assessment, and teachers could put more emphasis on the communicative tasks in a specific chapter.

***Lack of time allocated for formal assessment administration and scoring***

The second concern teachers expressed in the PLC meetings is the time allocated to teachers to administer and grade assessments. Communicative assessment requires more time than, for example, the administration of a grammar assessment. Writing, reporting, and speaking are complex operations, and the assessment of these skills requires considerable time, attention, and expertise from FL teachers. One teacher expressed this concern as follows: “There is too little time to develop, to teach, to learn and to assess in a communicative way.”

One solution to compensate for the time-consuming nature of communicative assessment was simply to reduce the number of formal assessments and introduce shorter and more focused informal, formative assessment activities in between less frequent and more comprehensive formal, summative assessments.

The time spent on administering a communicative assessment in an integrative way in classes of about thirty students was another concern, particularly regarding the assessment of listening, speaking, and conversational skills. Participants mentioned their concerns in these terms:

“How to integrate listening skills into a speaking assignment/presentation given the number of students?”

“In practice, you always have too little time to assess speaking skills the way you would like to.”

“I get only 2,5 hours per class for the organization of oral exams. How can I organize communicative assessments in this time?”

One of the suggestions from the PLC to overcome the first concern regarding the organization of the assessment of listening and speaking skills in an integrated way was to assess each skill at different moments with one part completed in class (listening) and one part completed individually or in pairs (presenting) at another moment. In such an asynchronous setting, integration may be less optimal, but can still take place indirectly using the same themes and/or situations in both parts of the assessment (i.e., listening and speaking). To overcome the time issue mentioned, PLC members suggested using teaching time to assess speaking skills by organizing and scoring presentations during the lessons. This can be justified by the fact that, while the presentation is an assessment for some students, it can function simultaneously as a learning activity for others. In addition, presenting in front of an audience as in real life enhances authenticity.

Moreover, to assess conversational skills, we designed assessment tasks that could be performed within 15 minutes by three students at the same time. Students received a description of a realistic situation per group in which each of them had to perform a different but similar task, for example telling someone about past events, for which one of the students was asked to talk about their last weekend, another about their last vacation and the third student about their last day off.

Finally, teachers were also concerned about the time needed to score communicative assessments, as the following excerpt illustrates: "Assessing reading skills with a rating scale takes me too much time." Based on the discussion and exchange of good practices on this issue, the PLC decided to develop rating scales with a holistic part in addition to the analytical part. The holistic part is short and to the point and can be scored very quickly to give a first indication of the global CEFR level at which a student performed a communicative speech act (see Figure 3.1). The analytical part focuses on the quality of student performance. The details of the descriptions allow teachers to score quickly, without having to give extra feedback to justify the score (see Appendix B). In addition to the practical objective, this choice was also motivated by the fact that combining holistic scoring with analytical scoring increases the validity by helping teachers to make better judgements about students' performance (Harsch & Martin, 2013).

Speech acts	(Pre) A1	A1+	A2	A2+
<b>Corresponding online</b> Completing a form Answering a simple mail	Can fill in a form mostly correctly with separate words but is not yet able to write a message.	Can answer correctly questions on a form with short sentences, but still has difficulty in (re) using the sentences correctly to write a message.	Can also write a simple message, but not always in an appropriate way (formal/informal). It is written correctly enough to get the message across despite some errors that disrupt communication.	Can also write a simple message in an appropriate way (formal/informal). It is usually written correctly and there are no errors that disrupt communication.

**Figure 3.1** Format of holistic part of the rating scale

### ***Students' lack of experience with communicative assessments***

The final practical challenge mentioned by the teachers in the PLC was to keep students motivated with fewer grades, as well as with assessments that would not focus on reproduction only. According to the participants, students often perceive assessment of FL skills as being less focused and therefore more difficult to prepare for. One of the questions teachers asked was: "How do we keep students motivated/prepared with fewer

formal assessments?”. To prevent students from “hiding” and not working regularly or effectively between the formal assessment moments, we stressed the importance of the systematic introduction of communicative informal assessment activities based on (part of) communicative tasks during the lessons. Students can thus regularly practice different communicative tasks under the supervision of the teacher and can learn, for example, to identify and integrate the vocabulary and grammatical structures they need to complete the tasks. Moreover, by having to perform authentic and meaningful tasks related to the learning objectives to be achieved on a regular basis, students are likely to be more committed and motivated to learn the language/attain the learning objectives (Shepard et al., 2018).

### ***Results from feedback on newly designed assessment materials***

Even though the teachers and students who were not involved in the design process (i.e., the colleagues of the PLC members and their students) were mostly positive about the feasibility of the formal assessment materials as designed in the PLC, they also put forward several concerns and points in need of improvement.

### ***Assessment preparation***

Teacher 1 specified that the themes present in the assessments were well related to the textbook and that the students were therefore well prepared. Teacher 2 mentioned that the textbook used did not include sufficient communicative learning activities to prepare the students for the assessment and that she had to look for extra materials which was time consuming.

### ***Assessment administration***

Teacher 1 specified that reading and writing assessments were the most feasible, because these could take place in class during a formal assessment, and they were easy to score with the rating scale. Teacher 2 and 4 indicated that speaking assessments were more difficult to organize since the school did not want them to assess students formally during the lessons. They both asked students to film their presentations and to send them in.

### ***Assessment scoring***

Teachers 2 and 4 mentioned their lack of experience with rating scales and indicated that they would need more experience or training to be able to use them correctly and to score more quickly.

### ***Students' experience***

Four of the five teachers reported that their students were satisfied with the new formal assessments and the level of it. Teacher 3 indicated that students had difficulty

understanding the descriptions in the rating scales and needed a lot of explanation to understand their scores.

The four students participating in a feedback session on the rating scale also mentioned that the language used in the rating scales was unfamiliar to them. They noticed that a lot of synonyms (e.g., “chunks”, “idioms”, “fixed expressions”) were used and that it was therefore too complex and confusing to them. They asked for more concrete examples to illustrate the different criteria.

Feedback from the stakeholders resulted in several decisions and recommendations regarding the final design of the different tools. We indicated suggestions in the assessment templates for the administration of formal assessments for listening and speaking skills (e.g., presentations in class, asynchronous setting). Based on the comments made by the students, we have reformulated and clarified the criteria of the rating scale. Fixed terms were chosen to ensure continuity from one level to another and concrete examples were added to illustrate the different criteria. More generally, we recommended teachers to use a textbook and additional materials with sufficient communicative activities and to plan formative activities to work on the criteria of the rating scales with their students (feed-up phase).

### 3.3.3 Final Design

In the final design, consisting of design principles (see Table 3.2) for each component of a communicative CBA program and tools based on it, we took into account the preliminary content-related and organizational design principles supplemented by the insights gathered and directions for improvement determined during the design and implementation process. In what follows, we explain how these principles were translated into concrete tools for the development of a communicative CBA program.

#### **Communicative assessment tasks (principles 1 to 7)**

*Principles 1 and 2: Meaningful communication and authentic situations*

In order to make assessments as meaningful and authentic as possible, all examples of communicative assessment tasks developed in the PLC fall within four realistic contexts, in which students can use the target language in plausible realistic situations:

- Situation 1: In the context of a stay abroad, an exchange or online contacts
- Situation 2: With the language assistant at school or other foreigners in their own country
- Situation 3: At a tourist destination in their own country
- Situation 4: On holiday/school trip in a foreign country

***Principles 3 and 4: Unpredictable language input and creative language output***

The CEFR describes different speech acts for each level, such as reporting, corresponding, sharing experience and information, expressing opinions, or providing suggestions (Council of Europe, 2020). Examples of communicative assessment tasks focus on different speech acts at A2 level. During a communicative formal assessment, students have to perform a few of these acts in new authentic situations (situations that have not been practiced literally beforehand) but around one of the themes at A2 level practiced in class. By doing so, students can show that they are able to spontaneously perform certain speech acts that they might encounter in real life. We have selected relevant speech acts for the target group that students should be able to perform at A2 level. Table 3.4 shows the selection and some examples of communicative assessment tasks based on one or more of these acts.

**Table 3.4** Selection of speech acts at A2 level (Based on Council of Europe, 2020) and examples of communicative assessment tasks

Speech acts	Examples of assessment tasks
<p><b>Understanding an interlocutor</b>                      Describing experiences of events and activities                      Comparing habits                      Theme: School</p>	<p><i>The son of your host family is curious about the Dutch school system. You explain how the school system works in the Netherlands and you compare it with what you know about the school system in his country. You answer his other questions.</i>  <i>Authentic document: timetable for school abroad.</i></p> <hr/> <p><i>You spent a day at school with the daughter of the family. The parents ask you in the evening what you thought of it. You tell them how the day went, what subjects you took and what you did. You compare the day with a school day in Holland. You answer the questions of your host family.</i>  <i>Authentic document: timetable for school abroad.</i></p>
<p><b>Reporting information</b>                      Identifying the main information from documents                      Comparing content                      Theme: Sports and Leisure</p>	<p><i>You would like to go to a summer camp in Germany with friends this summer. You have found an organization that offers different programs. Read the descriptions of the programs on the website and send a message to your friends with a summary: what do the different programs consist of? What are the differences and similarities?</i>  <i>Authentic document: screenshot of website pages</i></p> <hr/> <p><i>You are on holiday with your parents in France. At the reception of the campsite where you are staying there is a board with a lot of information about what can be done and when. You are interested in one of the activities. You get a folder at the reception desk with more information. You read it in detail and make a post on Instagram telling exactly what you are going to do.</i>  <i>Authentic documents: folders about activities</i></p>
<p><b>Corresponding (online)</b>                      Expressing feelings                      Giving an opinion                      Making suggestions                      Theme: Holidays</p>	<p><i>For the study abroad program at your school, you will stay in France with a host family for one week. You communicate in advance with the members of your host family through WhatsApp. You received the following messages. Please respond in French (50 words). Make complete sentences, give an appropriate answer, and ask a question.</i>  <i>Authentic documents: WhatsApp messages</i></p> <hr/> <p><i>You spend the summer holiday at a campsite in Spain with your parents. On the first day you met a nice Spanish guy. You often communicate via WhatsApp. Today you received the following messages. Please respond in Spanish (50 words). Make complete sentences, give an appropriate answer, and ask a question.</i>  <i>Authentic documents: WhatsApp messages</i></p>

***Principle 5: Integration of language skills***

To enhance validity, language skills should be assessed in an integrated way (Davies, 1999; Lee, 2006; Plakans, 2021). Different combinations were possible for this integration. We chose to develop examples of formal assessment integrating reading and writing skills on the one hand and listening and speaking skills on the other, with a distinction between conversational skills (listening and speaking in the context of a conversation) and listening and speaking skills (listening to gather information and then presenting something about it). These combinations appear to be the most common in real life. Examples of tasks integrating different skills can be found in the assessment templates under task specifications (see Figure 3.2).

***Principle 6: Organization***

To ensure that communicative formal assessments are feasible in secondary schools with classes of thirty students, we designed examples of formal assessments with written tasks that can be completed in 60 to 90 minutes and with oral tasks (conversation and presentation) that can be completed in 15 minutes and in groups of three students, during or outside the lessons.

**Formal assessments (principles 8 to 10)**

In order to guide teachers in the development of valid and feasible formal communicative assessments fitting their own teaching practice, we developed assessment templates with a focus on specific speech acts that can be used as a blueprint for the conception of formal assessments. Assessment templates indicate what is being assessed and how (see Figure 3.2). In the upper part of this template, general information can be provided on the assessment content: what skills and themes are assessed and what is the expected CEFR level? In the lower part, this general description can be specified in terms of relevant can-do statements, related speech acts, and appropriate task types. Tasks below and above the expected level can also be included with the corresponding CEFR level. Time and organizational aspects were added in assessment templates. Filling in this assessment template will help teachers to design formal assessments.

<b>Expected CEFR level:</b> <b>Language skills:</b> <b>Themes:</b> <b>Assessment format:</b> <b>Time and organization:</b> <b>Norm:</b>	<b>A2</b> <b>Reading and writing skills</b> <b>Holiday/free time</b> <b>Written assessment</b> <b>45 to 60 minutes. 2 hours grading time per class of 30 students.</b> <b>At least 55% of the points in the reading and writing skills rating scale A2</b>		
<b>Can-do statements</b>	<b>Speech acts</b>	<b>Task specifications</b>	<b>CEFR-levels</b>
Can report isolated information from signs, posters and programs written very simply in the target language and containing illustrations.	Reporting information	Finding and reporting information from posters or flyers	A1
Can write single words to give basic information (e.g., name, address, family).	Corresponding (online) (Making transactions)	Filling in a form/questionnaire to participate to an activity from the poster/flyers that have been read	A1
Can communicate the main points of simple texts clearly stated in the target language.	Reporting information	Finding and reporting information from short texts without illustrations	A2
Can indicate what to do in the evening or at weekends, make suggestions and respond to suggestions, confirm, or change an appointment, offer thanks or an apology.	Corresponding (online) (Expressing basic feelings/ making suggestions/ giving and following directions)	Writing short messages related to the text that has been read	A2
Can express feelings, impressions, and opinions on topics, such as lifestyles and culture, using a combination of standard sentences.	Corresponding (online) (Expressing feelings/ giving an opinion)	Reacting to a long message/text (response to a mail or comments on a post in a blog or in social media)	A2+

**Figure 3.2** Example of an assessment template

## **Informal assessment activities (principles 11 to 14)**

To monitor students' learning and to help students and teachers identify strengths and weaknesses before the formal assessments, we developed examples of informal assessment activities based on formative teaching activities for each stage of the formative cycle: feed-up (what is the student working towards?), feedback (where is the student now?) and feed-forward (what is the student going to work on next?) (Hattie & Timperley, 2007; Wiliam & Thompson, 2017). All these activities include a communicative language goal (content oriented) and a formative goal (process oriented). These informal assessments are based on communicative assessment tasks but are more focused on a specific point or criterion to be trained, such as being able to select essential information or to answer questions spontaneously on a specific topic. In addition, these activities also have an explicit formative objective specific to the formative process, such as getting to know the criteria (feed-up), learning how to conduct self-evaluation or peers-evaluation (feedback), or identifying strengths and weaknesses and establishing a work plan accordingly (feed-forward). Compared to formal assessments, informal assessments are short and easy to administer during the lessons. As with the formal assessments, they can be adapted to any theme. Examples of these informal assessment activities can be found in Appendix B.

## **Rating scales (principles 15 to 19)**

### ***Principle 15: Relation with CEFR***

To assess the level of students' communicative skills, we developed a format for rating scales and examples of three rating scales at level A2 for the evaluation of: conversational skills, reading and writing skills, and listening and speaking skills (see Appendix B). We first developed the analytic part of the rating scales by defining criteria based on the can-do statements from level A1 to level A2+, in order to describe the range between one level below the expected level and one above. We divided the criteria into four categories based on existing models of communicative competence (Bachman & Palmer, 1996; Celce-Murcia, 2007; Littlewood, 2011): sociocultural and interactional competence, strategic competence, and linguistic competence, which is subdivided into grammar, vocabulary and pronunciation or spelling.

### ***Principle 16, 17, 18: Criteria and skills integration***

The criteria are formulated in a qualitative way, emphasize the ability to communicate and value the use of communicative strategies. In order to be able to validly assess the receptive skills in combination with the productive ones and to give balanced emphasis to the assessment of the productive and receptive skills, we added a category *receptive competence*. For this category, we chose to introduce some of the newly developed criteria of the CEFR for cognitive mediation (North, 2016). By reporting in their own language what they have heard or read in the FL, students can show their understand-

ing while completing a task that they are likely to encounter in real life allowing more authenticity and enhancing validity in the assessment of the receptive skills.

### ***Principle 19: Practical use***

In addition to the analytical part of the rating scale we designed a holistic part. The holistic part is short and to the point and describes which speech acts students should perform at the expected level. Teachers can give a quick indication of the general CEFR level of a student while observing students perform the language acts present in the assessment. In addition, the detailed analytical part enables specific feedback and a formative use of the rating scales.

### **Example of an assessment planning (principles 20 and 21)**

The assessment planning we developed as an example (see Figure 3.3 below) consists of six formal communicative assessments (F) in which all the language skills are covered twice a year with an increase in difficulty. In the first half of the school year, the formal assessments are at level A1+ and in the second half at level A2+. Because CEFR levels are not always consistent for all skills this material can be developed and adapted to other levels. During the lessons, informal assessment activities based on (part of) communicative assessment tasks (I) are used in between the formal summative assessment occasions, with an increase in difficulty up to the level that will be assessed for a grade.

	Sept	Oct	Nov	Dec	Jan	Feb	Mar	Apr	May	Jun
Conversational skills A2	I A1	I A1+	<b>F A1+</b>	I A2	I A2	I A2	I A2	<b>F A2</b>		
Listening and speaking skills A2	I A1	I A1	I A1+	<b>F A1+</b>	I A2	I A2	I A2	I A2	<b>F A2</b>	
Reading and writing skills A2	I A1	I A1	I A1+	I A1+	<b>F A1+</b>	I A2	I A2	I A2	I A2	<b>F A2</b>

**Figure 3.3** Example of a communicative assessment planning at A2 level

The final design with explanations and recommendations on how to introduce it into practice has been presented in a toolbox and on a digital platform made available to the teachers participating in the project and to other FL teachers on request (see Appendix B).

### 3.4 Discussion and Conclusion

In this chapter we have presented the results of a collaboration between teachers and a researcher with two main objectives: 1) an empirical objective to identify and specify the challenges faced by teachers when designing and implementing a CBA program and determine the design principles of a feasible and valid CBA program for secondary FL education, adaptable to different school settings (RQ3.1) and 2) a practical objective designing tools (formats and examples) based on the design principles to help other FL teachers in the implementation of their own CBA program aligned with communicative learning objectives and adapted to classroom practice (RQ3.2).

Several studies already pointed out that FL teachers experience difficulties in the development and implementation of a CBA program aligned with communicative learning goals (Fulcher 2012; Jin, 2010; Vogt and Tsagari, 2014). These studies revealed a lack of language assessment literacy, particularly regarding CBA. However, mainly based on survey data, results of these studies did not identify teachers' specific needs as the answers given were often too vague or context specific (and therefore difficult to address in professional development programs). In addition, due to a lack of experience and expertise in the design of CBA, in their responses to the open questions of the survey, participants did not take into account the full range of practical and theoretical challenges they may face while designing a CBA program. Therefore, clear theoretical and practical guidelines on how to design a CBA program in specific contexts in secondary FL education were still needed.

In order to better understand the needs of FL teachers when designing a CBA program and the underlying issues behind their choices and decisions, we decided to involve them in a collaborative project in which they were observed and guided during the design process. The results of the co-design study presented in this chapter gave a complete and more precise picture of the challenges faced by FL teachers in the design and implementation of a CBA program, tailored to secondary education in the Netherlands and adaptable to specific school contexts. In addition, this collaborative project translated suggestions and decisions to overcome these challenges into final design principles and, based on them, concrete tools for the development of a communicative CBA program. The results showed that practical challenges were often related to the limited time allocated to teachers to prepare lessons, and to develop or select communicative learning and assessment activities. This time issue appeared to be even more pressing when the textbook used did not include sufficient communicative tasks or when teachers used their textbooks as the curriculum as such instead of as a tool. Other practical challenges were related to assessment administration and scoring within the allocated time. Teachers were also concerned about students' extrinsic motivation, with fewer

assessments and when focusing only on language skills. In addition to these practical challenges, teachers faced conceptual challenges regarding the operationalization of unpredictability, the integration of grammar and vocabulary within a communicative assessment task, and the integration of different language skills within one assessment.

Decisions made to address practical and conceptual challenges are summarized below in the form of suggestions to be taken into account when designing assessments aligned with communicative learning goals:

1. Language skills should be assessed in an integrative way, combining receptive and productive skills, to reduce the number of formal assessments and to enhance authenticity and validity.
2. Assessment tasks should include new situations to ensure unpredictability but should be based on well-known themes and speech acts to provide enough guidance to prepare students.
3. Rating scales should assess language skills in an integrated way. They should be descriptive, based on can-do statements belonging equally to each language skill. They should include a holistic and an analytical part to save time in scoring, to enable efficient feedback and ensure validity.
4. Informal assessments based on (part of) communicative assessment tasks should be systematically introduced to enhance students' motivation and to guide their learning process.

The final design reveals the importance of combining formal and informal assessments of all language skills in a communicative way (Dixson & Worrell, 2016; Rea-Dickins, 2001). Formal assessments (with both a summative and a formative function) are essential in order to stimulate a positive washback on teaching and learning activities (Alderson & Wall, 1993; Green, 2007) but also to enable teachers to take professional decisions based on students' use of the language to communicate. Informal (formative) assessment activities based on (part of) communicative assessment tasks provided on a regular basis will not only guide teaching and enhance learning but may also increase students' motivation. Indeed, as Shepard and colleagues (2018) point out, students feel more engaged when they have to complete meaningful tasks that are directly related to what they have to learn. Moreover, the regular introduction of informal assessments during the lessons can provide opportunities to reduce the number of formal assessments giving students enough time to practice and experience communicative activities and to receive feedback on their performance before being formally assessed. Furthermore, the design of communicative assessment tasks integrating several language skills has enabled experimentation with a form of assessment not often used in the classroom, in which several language skills, in this case productive and receptive skills, are assessed in

an integrated way (Grabe & Zhang, 2016). Being more authentic and realistic, integrated assessments are more valid and more likely to generate positive washback on teaching activities and on the students' learning process (Weigle, 2004).

In practice, FL teachers participating in the design process found it difficult to develop and administer such assessments, partly due to a lack of experience but also due to the difficulty of assessing the different skills in a balanced way. Indeed, this latter point has been the subject of much research, revealing mixed results. While some studies report slight difference in the proportion to which the different skills are assessed in integrated assessments (Lee, 2006; Gebril, 2010), others indicate that the productive skill tends to predominate over the receptive skill, the latter only being used as input (Cumming et al, 2006). In order to avoid such an imbalance in the integrated assessment of productive and receptive skills, we decided to give more emphasis to the assessment of the receptive skills by creating a category "reception" and by introducing some of the newly developed mediation criteria of the CEFR for cognitive mediation (North, 2016). Finally, the choice to focus on speech acts rather than on ready-made situations in order to increase unpredictability, as recommended in the CEFR manual (Council of Europe, 2020), enhances the potential for a positive washback on teaching and learning activities in which teachers teach their students to react spontaneously as in real life instead of encouraging students to learn a script written in advance by heart.

The meeting of practice and theory during the design sessions highlighted the challenges faced by teachers in developing materials for a communicative CBA program both in terms of feasibility and validity, as evidenced in the researcher's PLC logbook and stakeholder feedback. With the tools that were developed in the current design-based study, we aspire to provide support to other FL teachers to overcome these challenges.

### **3.5 Limitations and Implications**

Most of the teachers involved in the PLC chose to participate on a voluntary basis in the development sessions. They took part in the working sessions in the evening outside of their teaching or professional time. This factor potentially influenced the results in diverse ways. The participating teachers were de facto interested in the topic and were willing to learn and improve their classroom practices regarding assessments. Their interpretations and reactions might therefore have been more positive and constructive than those of other language teachers. On the other hand, the extra workload sometimes led to one or two absences or to a lack of preparation. The try-outs, for example, were not carried out systematically by all participants. However, it is remarkable to note the steady attendance and continuity, as the number of participants remained stable

during the whole design process. Finally, the tools were designed for classroom-based assessment of students in lower form secondary education in the Netherlands. The focus was on validity and feasibility with the aim of generating positive washback on classroom practice. The material has not yet been assessed for reliability.

Our results confirm, as Tzagari and Vogt (2017) have also pointed out, that it is essential to involve teachers in collaborative projects in order to better understand their needs for assessment in the context of their own practice. To support this, schools should facilitate teachers by giving them the opportunity to participate in professional development projects in collaboration with colleagues from other schools and with experts (Chapman & Muijs, 2014; Plomp & Nieveen, 2009; Reeves, 2006; Stoll et al., 2006; Van den Akker, 1999). The results furthermore suggest that methodology courses in FL teacher training should focus more on CBA (instead of on large-scale assessment), as this type of assessment is more common in and relevant to daily teaching practice in lower form education in the Netherlands. Besides, FL teacher training should facilitate on-the-job assessment education (Xu & Brown, 2016) and encourage prospective teachers to put theory into practice by actively taking part in the development of CBA during their traineeships. Finally, while the challenges identified and the characteristics formulated for the design of a valid and feasible assessment program described in this chapter are certainly relevant for FL teachers and departments beyond those participating in this project, research on further implementation in different school settings is crucial to confirm or adapt the design decisions that have been made and to further investigate teachers' and students' experiences.

**CHAPTER 4**



# The Implementation of a Foreign Language Classroom-based Assessment Program in Dutch Lower Form Secondary Education

This chapter has been submitted for publication as: Rouffet, C., Van Beuningen, C., & De Graaff, R. (under revision). The implementation of a foreign language classroom-based assessment program in Dutch lower form secondary education.

## Abstract

According to the principle of *constructive alignment*, assessment should be in line with learning objectives to enable effective teaching. Since the main objective of foreign language (FL) education is to teach students to communicate in an FL, assessment activities should measure this ability. Moreover, an effective assessment program should contain the right balance between formal and informal assessments. Current assessment practices in Dutch secondary FL education do not meet these requirements: FL assessments are often not communicative (enough), and the use of informal assessment activities is limited. FL teachers in our research context have to select, modify, develop and administer classroom-based assessments (CBA) as part of their daily practice. This requires a thorough knowledge of assessment methods and clear guidelines for their application in practice. To guide teachers in this complex task, we designed tools (frameworks and examples) for the development of a communicative CBA program for Dutch lower-form secondary education. The current chapter reports on how these tools were used to implement a communicative CBA program in practice, and on how teachers and students experienced this implementation. Results show a high degree of observed treatment fidelity. Moreover, a large proportion of the teachers perceived more alignment in their curriculum and the majority of the students reported assessment activities being more focused on the development of language skills. This shows that with appropriate tools FL teachers can optimally align the way they assess their students with communicative learning goals.

## 4.1 Introduction

This chapter reports on the implementation phase of a larger design-based research project aimed at investigating the effects of a communicative classroom-based assessment (CBA) program on foreign language (FL) teachers' pedagogical approach, within the context of Dutch lower form secondary education. This study investigates how tools designed in the previous project phase were used by FL teachers to implement a communicative CBA program in their own context.

The large majority of FL teachers worldwide agree that the purpose of learning one or more FLs is to be able to communicate in these languages (Savignon, 2017). However, despite the fact that the learning objectives of many FL curricula are nowadays based on communicative language teaching principles, communicative language teaching (CLT) is not always implemented in practice as intended (Kissau et al., 2015; Little, 2007). Studies on the subject have revealed various factors that may be at the root of the problem, such as a lack of time, materials, or expertise available, national examinations, or teachers' beliefs (Ahmed, 2016; Anani Sarab et al., 2016; Chang, 2011; Coskun, 2011; Nishino, 2011; Zhu & Shu, 2017). These factors often lead to a misalignment between the CLT-oriented learning objectives on the one hand and the learning activities and assessments on the other hand (Barnes, 2017; Taylor, 2005).

As Biggs (1996) has pointed out, in order to achieve educational goals, it is essential to align the three main components of a curriculum (i.e., *constructive alignment*). If for some practical or conceptual reason learning activities and/or assessments are not aligned with the learning objectives, this creates a lack of coherence that hinders the achievement of the educational goals. The assessment component in particular plays a crucial role in teachers' pedagogical choices, and hence affects the way students learn (i.e., *washback effect*; Alderson & Wall, 1993; Green, 2007). For FL education, this means that if assessments are non-communicative (e.g., focus on grammar), the teaching and learning activities may become less communicative as well. To create alignment with the communicative learning objectives set in most FL curricula, it is therefore crucial that FL teachers provide assessments that evaluate students' ability to communicate in the FL. In our earlier study (Chapter 2), in the context of Dutch lower form secondary FL education, we identified a lack of constructive alignment between the national CLT-oriented objectives on the one hand, and the type of assessments on the other. Observed assessment activities focused mainly on knowledge of grammar rules and vocabulary out of context, and activities aimed at assessing the language skills in a communicative way were scarce. The results furthermore showed that these assessments seemed to exert a negative washback effect on teaching activities as the observed classroom activities focused mainly on language forms outside communicative situations. FL teachers in

our study indicated that they lacked the time and expertise to develop and implement communicative assessments.

As previous studies have shown, this misalignment between communicative goals on the one hand and non-communicative assessments on the other can be due to the fact that assessment of communicative skills is complex, and the development, administration and scoring of communicative assessment tasks requires a lot of time, organization and expertise which is even more challenging in secondary education with usually large classes (Jin, 2010; Stoyhoff, 2012). The Council of Europe offers various open access manuals emphasizing the importance of aligning assessments with communicative objectives and using FL assessments that are based on authentic and realistic tasks (ALTE, 2011; North et al., 2009). However, these manuals are more suitable for large-scale assessment developers and language testers than for individual FL teachers who need to develop classroom-based assessments (CBA). The guidelines and advice they offer are not contextualized enough for classroom practice and are therefore difficult to adapt and to implement by FL teachers (Fischer, 2020). Likewise, the guides and manuals on assessment construction, used in FL teaching programs remain largely theoretical and pay little attention to CBA, focusing more on large-scale assessments (Allal, 2016; Fulcher, 2012; Gan & Lam, 2022; Tsagari & Vogt, 2014). However, secondary school teachers are frequently required to select, modify, develop, and administer assessments as part of their daily classroom practice, especially when decisions have to be made at the end of each school year as to whether students can pass to the next grade (Black & William, 2018). It is therefore important that they know how to develop and implement CBA aligned with communicative learning goals.

Besides, to be effective in optimally enhancing students' learning, CBA designed by teachers should not only include formal assessments with a summative function but also informal formative assessment activities with respect to the communicative learning goals (Dixson & Worrell, 2016; Rea-Dickins, 2001). In the past years, studies have demonstrated the benefits of CBA taking place during the learning process to enhance students' learning (see Lewkowicz & Leung, 2021). These studies have highlighted the value of CBA in stimulating learning and adapting teaching but have also revealed the difficulties FL teachers faced to introduce it in practice, especially more informal types of assessments with an essentially formative function (Chen & Zhang, 2021; Gan & Lam, 2022; Pan, 2020; Tsagari, 2020). Frameworks for the implementation of CBA including the most informal types of assessment, such as spontaneous interaction between teacher and students, have been developed (e.g., Hill, 2017). However, FL teachers still lack tools (e.g., concrete examples) to implement CBA into practice (Firoozi et al., 2019; Tsagari & Vogt, 2017).

Recent studies on the implementation of CBA often focused on one specific assessment form with a primarily formative function (e.g., self and peer assessment, use of portfolio, digital tools or feedback in interaction) and were carried out within centralized educational systems (Burner, 2023; Leung et al., 2018; Naghdipour, 2017; Poehner, 2009). Moreover, most of these studies were conducted in the context of higher education, within courses of English as a second language, and focused on the assessment of a specific language skill, most often writing (Burner, 2023). Fewer studies have investigated how secondary school FL teachers, including teachers of other languages than English, can combine formal and informal forms of CBA in a constructively aligned assessment program, aimed at the assessment of all language skills (Burner, 2023; Sultana, 2018). It is therefore important to gather insight into how FL teachers can be supported in the development of valid, feasible and well-balanced CBA programs and to investigate how such programs are implemented in practice.

As pointed out by Binkhorst et al. (2017), in order to ensure a sustainable implementation of a curriculum change - such as an assessment program - into practice, it is essential to include teachers in the design of the new program. The benefits of collaborative work between researchers and teachers have been highlighted in several studies on the subject (Allal, 2016; Harn et al., 2013; Levin, 2004; Spada & Lightbown, 2022). These studies not only show that participants are more likely to introduce materials that they have designed themselves (Levin, 2004), but also that materials which are informed by both theory and practice are more likely to be successfully implemented by other teachers (Harn et al., 2013). In order to achieve this goal, it is also important to consider the process of the implementation in different settings. As several studies have pointed out (Durlak, 1998; Lendrum & Humphrey, 2012) studying the process of the implementation of an intervention not only allows for monitoring if the materials have been implemented as intended, but also helps to identify what works and what is problematic in practice (Lee & Butler, 2021). This enables the formulation of recommendations for a successful implementation (Lendrum & Humphrey, 2012).

#### **4.1.1 The Current Study**

The current study investigates how tools developed in co-creation with FL teachers in a previous study (Chapter 3) were used by other teachers in various schools and language departments to implement a communicative CBA program. We also report on how the FL teachers and students involved experienced the implementation of the CBA programs developed with the help of those tools. The research questions guiding this study are:

*RQ4.1: How is a communicative classroom-based assessment program implemented by language departments in Dutch lower form secondary education based on tools made available to them?*

*RQ4.2: How is the implementation perceived by the teachers and experienced by the students involved?*

The implementation was described and evaluated using the steps of Van den Akker's model of curriculum representations (2013), as described in Figure 4.1. First (RQ4.1), we investigated how the intended CBA programs (*Formal/Written*) - that were developed by the different language departments based on the design principles and assessment tools made available to them (*Ideal*) - were implemented in practice (*Operational*). Furthermore, teachers' perceptions regarding the implementation process (*Perceived*) and students' experiences regarding the implementation (*Experiential*) were investigated (RQ4.2).

<b>Intended</b>	<i>Ideal</i>	Vision (rationale or basic philosophy underlying a curriculum)
	<i>Formal/Written</i>	Intentions as specified in curriculum documents and/or materials
<b>Implemented</b>	<i>Perceived</i>	Curriculum as interpreted by its users (especially teachers)
	<i>Operational</i>	Actual process of teaching and learning (also: curriculum-in-action)
<b>Attained</b>	<i>Experiential</i>	Learning experiences as perceived by learners
	<i>Learned</i>	Resulting learning outcomes of learners

**Figure 4.1** Typology of curriculum representations (Van den Akker, 2013, p.56)

## 4.2 Method

### 4.2.1 Educational Context

In the decentralized Dutch educational system, only school subjects, learning objectives and part of the final examinations are determined at a national level. Departments within schools can choose their own pedagogical approach towards the achievement and assessment of the learning objectives (Scheerens, 2016). The teaching of two FLs in addition to English is compulsory in lower form (i.e., first three years) and becomes optional in upper form of secondary education. The learning objectives for FL teaching are based on communicative principles and linked to the CEFR levels (College voor Toetsen en Examens, 2019; Meijer & Fasoglio, 2007). The national final exam at the end of secondary education is a reading comprehension exam that covers 50% of the final grade, for each of the FL subjects. The other part is based on CBA developed and admin-

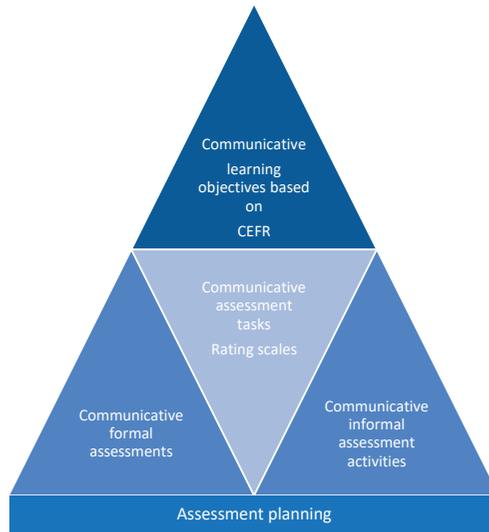
istered in each school by individual departments and teachers. In lower form secondary education, all assessment formats and contents are classroom-based. This implies that teachers must have significant expertise on how to assess their students, which is currently not the case (Chapter 2).

#### **4.2.2 Design of an Intervention with “Built-in Adaptations”**

In a previous study, a group of 21 FL teachers of the most commonly taught FLs in Dutch secondary education (English, French, German, and Spanish) from 15 different schools took part in a Professional Learning Community with the aim of designing an intervention for the development of a valid and feasible CBA program (Chapter 3). Given the variety of school settings in which teachers operate in the Netherlands, we chose to work with FL teachers from different schools and to develop tools (guidelines, formats, and examples) rather than a fixed program. Harn et al. (2013), following Webster-Stratton et al. (2011), call this type of intervention in which participants can adapt materials to match their own practice an intervention with “built-in adaptations” (p. 188). Core components and essential features of the intervention are fixed but participants can introduce them in different ways in their own practice. The flexibility offered by this type of intervention may increase the chance of a successful and sustainable implementation.

The tools of the intervention with “built-in adaptations” were designed with the aim to support FL teachers who teach at A2 level in lower form secondary education in the Netherlands to compile a communicative CBA program adapted to their own practice. Core components of a CBA program were first determined in accordance with CLT principles and aligned with communicative learning objectives: assessment tasks, formal assessments, informal assessment activities, rating scales and an assessment planning (Figure 4.2). For each of the components, design principles (see Chapter 3, Table 3.2) were formulated taking into account conceptual and practical considerations. Tools were then designed based on these principles: examples of communicative assessment tasks, assessment templates to compose formal assessments, examples of informal assessment activities, formats and example of rating scales and an example of an assessment planning. Examples of communicative assessment tasks could be selected or adapted and combined according to an assessment template to compose a formal communicative assessment suitable to the own classroom practice. Assessment templates show how assessment tasks could be distributed within a formal assessment in relation to the learning objectives at a certain level. Examples of informal assessment activities appropriate to communicative learning tasks and for each stage of the formative process (*feed-up*, *feedback* and *feed-forward*; Hattie & Timperley, 2007) can be selected or adapted for introduction in daily teaching practice. Rating scales with success criteria can be used during the formative process in the context of informal assessment activities and serve as a rating tool for the evaluation of formal assessments. Finally, an

example of an assessment planning for the distribution of formal and informal assessment activities within a school year was provided. The design principles and tools were finally compiled in a toolbox (see Appendix B).



**Figure 4.2** Components of a constructively aligned communicative CBA program

### 4.2.3 Participants

#### *Language Departments*

Fourteen language departments with 30 FL teachers from 10 schools in the Netherlands participated in the current implementation study: one English department, six French departments, six German departments and one Spanish department. All participating teachers within these departments held a teaching certification and taught classes at A2 level of the CEFR. Teachers' experience varied between 2 to 20 years.

#### *Students*

Lower form students aged 12-15 ( $N = 423$ ) from eight of the participating schools and 10 language departments filled in a survey about the implementation of the new CBA program. They were taking language classes at A2 level: English ( $n = 46$ ), French ( $n = 211$ ), German ( $n = 158$ ) and Spanish ( $n = 8$ ).

### 4.2.4 Implementation's Procedure

The implementation consisted of the development of a communicative CBA program and its introduction in practice in each of the language departments involved. The participating teachers developed these CBA programs based on the tools made available to them (i.e., designed in the previous research phase; cf. Chapter 3). The CBA

programs were implemented over the full school year 2020-2021 by the participating FL teachers. In May and June 2020, before the start of the implementation in September 2020, a two-hour kick-off session was organized with all of the participating schools, language departments and teachers, online or at the schools depending on the health regulations related to the Corona pandemic at that time. In preparation for this session, participants received access to the toolbox on a digital platform containing explanatory videos and examples on how to use the tools to develop a communicative CBA program adapted to their own practice. During the kick-off session, the researcher presented the different tools to the participants and instructed them on how to use them. All questions and practical issues regarding the development of a CBA program and its introduction in practice in the specific context of their schools were discussed. During the implementation, participants were free to contact the researcher for further details or information or for feedback on assessment materials they adapted or developed based on the tools made available. To observe the implementation process more closely, the researcher took part in four two hours working sessions in one of the schools involved. We chose this specific school for two reasons. First, the school board had decided to facilitate language teachers in introducing communicative assessments in school year 2020-2021. Secondly, the language departments of French and German, consisting of eight language teachers, four French and four German, wished to work together and introduce the program in all their classes of the lower form.

### 4.3 Data Collection and Analysis

In order to answer the research question on how the toolbox materials were implemented in practice (i.e., RQ4.1), we first collected data through the analysis of the CBA programs that participants developed using the toolbox. This was done according to the recommendations given by Harn et al. (2013) on how treatment fidelity of an intervention with “built-in adaptations” in which participants can adapt materials to match their own practice in schools, such as our toolbox, should be evaluated. The main recommendation was to identify the “active ingredients” defined as the core components and essential features of the intervention. Criteria based on the design principles (*Ideal*) were thus first identified and gathered in a checklist in order to determine the actual fidelity of the CBA programs (*Formal/Written*) to the overall concept of communicative assessment. The checklist was used for the analysis of the different components of the CBA programs participating teachers and language departments developed using the toolbox (*Operational*). Classroom observations were furthermore used to establish to what extent informal assessment activities reflected communicative learning goals (*Operational*).

To answer the second research question on how the implementation process was experienced by the different stakeholders, we collected data through the analysis of teachers' logbooks (*Perceived*) and a student survey (*Experiential*). The same data sources were collected in all language departments involved. To get more insights into factors influencing the implementation process and to complement the quantitative data obtained from the student survey we collected additional qualitative data through interviews with teachers and students in the school we selected to observe the implementation process more closely.

### 4.3.1 RQ4.1: Actual Implementation (*Operational*)

#### *Analysis of CBA programs*

Language departments ( $N = 14$ ) submitted their CBA programs consisting of assessment materials developed or selected from the tools made available to them in the toolbox and used by each teacher in that department in one of their classes at level A2. We received the following materials (i.e., CBA components): examples of informal assessment activities, formal assessments, rating scales, and assessment plannings. Submitted assessment materials were analyzed using a checklist including criteria that can be seen as the "active ingredients" that a communicative CBA program should feature (Harn et al., 2013), based on the design principles. Using the checklist, we determined to what extent the materials provided by the different language departments met the main criteria of a communicative CBA program. For each criterion, the degree of treatment fidelity was indicated as *high* when almost all of the materials available (80% or more) met the criterion, *acceptable* when a good proportion of the materials (between 50 and 80%) met the criterion, or as *low* when only a small proportion (less than 50%) met the criterion. Finally, we calculated the percentage of language departments that reached a high, acceptable, or low level of fidelity for each of the criteria.

#### *Classroom observations*

Teachers were observed teaching in one of the A2-level classes. They were observed using an observation tool developed and validated by the Foreign Language Education Research Group at Utrecht University of Applied Sciences (see Appendix A). We used this observation tool to measure the degree to which informal assessment activities adhere to CLT principles and had a formative character in accordance with the design principles. The criteria related to the communicative and formative character of the assessment activities were scored on a four-point Likert scale ranging from 1 (*not seen*) to 4 (*seen to a large extent*). All observations were performed by the researcher. The same instrument had already been used by the researcher and members of the research group who co-developed the instrument in a previous study; inter- and intra-rater reliability had been checked and substantial agreement was found (see Chapter 2). Originally, we planned to observe at least one teacher in each of the language departments involved.

Unfortunately, not all planned observations could be carried out due to the restrictions related to the COVID-19 health crisis. Eventually, 12 (of the initial 14) FL teachers (T1 to T12) from 12 language departments were observed.

### 4.3.2 RQ4.2: Perceived Implementation (*Perceived and Experiential*)

#### *Logbooks*

All participating language departments ( $N = 14$ ) filled in a logbook on a regular basis on how teachers within the department experienced the implementation of the CBA program in one of their classes. Data were first pseudonymized per school S1 to S10. The letters *E* (English), *F* (French), *G* (German) and *S* (Spanish) were added to the codes to indicate the different language departments. To guide the structure of the report, participants received three reminder emails at the end of each trimester with a series of questions to which they could respond, such as how they implemented informal and formal assessments in the past trimester, what they perceived as positive, and which issues they encountered, following the recommendations of Lendrum and Humphrey (2012) on how to get insights into the implementation process of an intervention. Responses from all language departments were compiled. The compiled logbooks were then analyzed qualitatively to determine how the implementation of the CBA programs had been perceived by the teachers of the language departments involved.

#### *Student survey*

An online survey with multiple-choice questions was filled in by students in intervention classes ( $N = 423$ ) to provide a student's perspective on the implementation of the new assessment materials. The questions focused on students' perceptions of formal and informal assessments. Results were analyzed descriptively. We examined the frequency of responses to the different questions to determine students' experiences of the implementation (*Experiential*).

#### *Student and teacher interviews*

Two teachers, one French teacher and one German teacher, and six students taking French and German classes from one of the schools participated in semi-structured interviews. The interviews included open-ended questions about the implementation of the assessment program. The interview duration averaged around 20 minutes. All interviews were recorded, transcribed verbatim, and pseudonymized (T1 and T2 for the teachers and S1 to S6 for the students). After transcription, recordings were deleted in accordance with the approval of the Faculty Ethics Assessment Committee Humanities of Utrecht University. Both teachers were interviewed individually online due to the restrictions related to the health crisis at that time. Students were interviewed in groups of three at their school. Topics of the semi-structured interviews were similar to the questions structuring the logbook (for the teachers) and to the questions from

the students' survey (for the students). The interviews were aimed at providing more details on how the implementation was perceived by the teachers and experienced by the students involved.

The qualitative data gathered to answer the second research question (i.e., logbook and interviews) were analyzed following the phases defined by Corbin and Strauss (2008): categorizing data into themes and identifying general patterns. We used the same coding scheme to analyse all data sources related to teachers' perceptions of the implementation. To establish how the implementation process was perceived by teachers, we sorted the data into positive experiences and issues regarding both the development of a CBA program with "built-in adaptations" (Harn et al., 2013), and its introduction in practice. Under these themes, codes were determined based on the participants' answers (Corbin & Strauss, 2008). We based the analysis of student interviews on the last step of Van den Akker's typology of curriculum representations, and investigated how the implementation goals were attained according to learners' experiences (*Experiential*). We conducted a thematic analysis and coded the data based on respondents' input when describing how they experienced the intervention. We specifically looked at students' experiences regarding the learning goals, the informal assessment activities during the lessons and the formal assessments. The codes were then clustered into themes representing the general experiences of the respondents on the implementation. The researcher coded all data; to ensure reliability, 20% of the logbooks and interviews were coded by another member of the research group. We used Cohen's  $\kappa$  to check intra-coder reliability. Substantial agreement was found ( $\kappa = .764$  (95% CI, .566 to .962),  $p < .001$ ).

## 4.4 Results

The results reported below first show how the implementation has been operationalized in the different language departments involved (RQ4.1). We then show how the implementation has been experienced by the different stakeholders (RQ4.2).

### 4.4.1 RQ4.1: Actual Implementation (*Operational*)

#### *Analysis of Assessment Materials*

Treatment fidelity Table 4.1 shows what percentage of language departments developed the CBA programs with a high, acceptable or low degree of treatment fidelity, for each "ingredient" of the different program components.

**Table 4.1** Treatment fidelity of the CBA programs developed by the language departments  
(*N* = 14)

<b>Components of a CBA program and “main ingredients”</b>	<b>High treatment fidelity (&gt;80%)</b>	<b>Acceptable treatment fidelity (50-80%)</b>	<b>Low treatment fidelity (&lt;50%)</b>
<b>Formal assessment activities</b>			
Include meaningful tasks	72%	7%	21%
Include authentic tasks	36%	43%	21%
Include open/creative tasks	21%	43%	36%
Encourage students to react spontaneously	36%	21%	43%
Include tasks that integrate different skills	36%	21%	43%
Are related to CEFR can-do statements	57%	22%	21%
Include a variety of tasks with an increase in difficulty	14%	36%	50%
Include tasks at the appropriate level	64%	7%	29%
<b>Total</b>	<b>42%</b>	<b>22%</b>	<b>33%</b>
<b>Informal assessment activities</b>			
Are not part of any formal evaluation	50%	29%	21%
Include a communicative learning goal	64%	22%	14%
Include communicative principles	57%	29%	14%
Include a formative goal corresponding to a specific phase of the formative cycle	64%	22%	14%
<b>Total</b>	<b>59%</b>	<b>25%</b>	<b>16%</b>
<b>Rating scales</b>			
Bring together criteria related to the learning objectives, corresponding to the can-do statements of the CEFR levels	57%	7%	36%
Enable the assessment of several language skills used in an integrated way	29%	21%	50%
<b>Total</b>	<b>43%</b>	<b>14%</b>	<b>43%</b>
<b>Assessment planning</b>			
Include the assessment of all language skills in a formal and informal way	29%	43%	29%
Offer an even distribution of the formal and informal assessments of the language skills over the school year	14%	43%	43%
<b>Total</b>	<b>22%</b>	<b>43%</b>	<b>36%</b>

In general, formal assessments have been developed with an acceptable to high treatment fidelity in the large majority of the language departments. For some “ingredients” of formal assessments, however, we observed a lower level of treatment fidelity: the stimulation of spontaneous reactions, the integration of different skills, and the increase in difficulty between the different tasks. Informal assessment activities have been developed with a high level of treatment fidelity in more than half of the language departments. Rating scales were mainly lacking fidelity regarding skills integration. Finally, the assessment plannings teachers developed adhered to the main ingredients for only few language departments (22%).

### ***Classroom observations***

Results in Table 4.2 show that most of the informal assessment activities observed were sufficiently meaningful, authentic, asked for an appropriate degree of free production, and had a focus on form instead of on forms (*sufficiently seen*). These results are consistent with those of the material analysis (Table 4.1), which showed that informal assessment activities were developed with a high level of treatment fidelity in the majority of the language departments. Integration of language skills was often *barely seen*.

**Table 4.2** Adherence to CLT criteria and appropriate use of formative tools in informal assessment activities during observed lessons ( $N = 12$ )

CLT criteria	<i>M</i> Range 1-4	<i>SD</i>
Meaningfulness	3.3	.9
Authenticity	3.2	1.1
Appropriate degree of free Production	3.3	.9
Integration of skills	2.5	1.3
Focus on form	3.2	1.1
Formative character	3.0	.8

1 = *not seen*; 2 = *barely seen*; 3 = *sufficiently seen*; 4 = *seen to a large extent*

## **4.4.2 RQ4.2: Perceived Implementation (*Perceived and Experienced*)**

### ***Logbooks***

Teachers’ perceptions of the implementation process as evident from the logbooks are reported below in Tables 4.3, 4.4 and 4.5. Table 4.3 shows the positive experiences and issues teachers mentioned regarding the development of their own CBA program. During the development process, the use of more authentic materials and the inspiration teachers got from the materials made available to them (i.e., toolbox and accompanying teacher guide) were indicated as a positive experience in 6 of the 14 logbooks. Some teachers also mentioned that the development of the assessment materials stimulated teamwork and contributed to their professional development. While nine language

departments mentioned the lack of time to develop and adapt the materials as an issue, two others saw time spent as an investment that “pays off in the end”. Finally, six of the language departments mentioned lacking expertise as a hurdle to develop the materials, and two of them experienced a lack of appropriate activities in their textbooks to do so.

**Table 4.3** Positive experiences and issues regarding the development of a CBA program with “built-in adaptations” reported in logbooks from the different language departments ( $N = 14$ )

<b>Teachers’ experiences during the development of a CBA program</b>	<b>Frequency</b>	<b>Quotes</b>
<b>Positive experiences</b>		
Stimulate the use of authentic materials and give inspiration	6	“I got a lot of inspiration from the examples in the guide.” (S5-G)
Stimulate teamwork and professional development	2	“I really liked the structure because we really learned from it.” (S3-G)
Investment	2	“It takes a lot of time, but the investment pays off in the end.” (S1-F)
<b>Issues</b>		
Lack of time/time consuming	9	“What is negative is how much preparation time it takes and how much time you actually need to create an assignment properly.” (S1-F)
Lack of expertise in the development of CLT materials	6	“We did miss the contact with an expert. We are still beginners and would have liked more guidance.” (S6-G)
Lack of materials available	2	“I find that the textbook does not include enough authentic material.” (S8-F)

In Table 4.4 we report how language departments perceived the introduction of their CBA program in practice. Results show that most language departments valued the fact that the program provided more alignment between the learning goals and the teaching and assessment activities. Teachers from different language departments described their lessons as being more goal-oriented and noticed a positive washback of the formal assessments on the selection of learning activities. They also reported on their enthusiasm and that of their students, about the introduction of the new CBA program in practice. Other points often mentioned as positive were the fact that they had more insights into the learning progress of their students and that the students themselves seemed to have more insights into their own progress. Further aspects mentioned relatively often were the fact that the program seemed to increase students’ confidence, that students were more pro-active during the lessons, and that they obtained better results compared to previous years.

Table 4.5 shows the main issues reported by teachers regarding the implementation of their CBA program. A large proportion of the teachers report the lack of experience they and their students have with CLT and formative assessment activities and mentioned the need to get used to this way of teaching and learning. Almost all of the teachers identified the fact that the implementation was time consuming as an issue, for both the organization of communicative activities (formal and informal) as well as for feedback they had to provide students with.

**Table 4.4** Positive experiences regarding the implementation of a CBA program with “built-in adaptations” reported in logbooks by teachers from the different language departments

(N = 14)

Teachers’ positive experiences during the implementation of a CBA program	Frequency	Quotes
More alignment	12	“More is done with long-term goals, encouraging more growth, and working on language skills.” (S2-E)
Students’ and teachers’ enthusiasm/satisfaction	10	“Students experience it as more pleasant also because they realize that they can do more with the language than the students with a traditional approach.” (S6-F) “On the positive side, I had more fun, spoke more German and felt much more relaxed to help my students to really practice language skills.” (S5-G)
More insights into learning progress	9	“Students really liked to know what went well and what they needed to improve. I have seen students really improve because of this.” (S8-F)
Active and efficient lessons	9	“Students ask for grammatical explanations at the moment they need them for an assignment. The mindset is then different from regular lessons where the teacher says: Today we are going to deal with the <i>conditionnel</i> . They ask how they can build a sentence to suggest something.” (S6-F)
Increased confidence	8	“The formative activities during the lessons have reduced the barrier for practicing productive skills. Students dare more to speak.” (S1-F)
Better results	7	“Students generally got very good results compared to previous years.” (S2-F)
More use of the target language	6	“I notice that I speak more German in class than last year. In previous years, I started well, but it weakened after the autumn break.” (S9-G)

**Table 4.5** Issues regarding the implementation of a CBA program with “built-in adaptations” reported in logbooks by teachers from the different language departments ( $N = 14$ )

Issues reported during of the implementation of a CBA program	Frequency	Quotes
Students' and teachers' lack of experience with CLT	11	<p>“What I experience as less positive are the unexpected twists in my approach. What I thought was well-defined has not necessarily been clear to the students, which leads to confusion.” (S6-F)</p> <p>“Students find it stressful to have to perform within a theme and not know exactly what questions they will be asked.” (S4-F)</p>
Time consuming	11	<p>“We have just finished the oral examination for conversation skills. It takes a lot of time (we did it during the lessons). In the beginning, I spent a bit too much time per group because I let the students speak for too long.” (S3-G)</p>
Lack of school support (colleagues or school administration)	5	<p>“I find it difficult that students do not make a change if only a single subject is using formative evaluation, and the other subjects are taught and assessed in a more traditional way. Students remain then very focused on grades.” (S2-G)</p>
Lack of explicit basic knowledge regarding grammar and words and concerns about transition to upper form and “regular” classes	5	<p>“Grammar is now taught indirectly in a playful way. It has not been assessed as was done in the other grades. Only in the long run we will see what results this produces. The question is whether there will be a gap in grammar knowledge after all.” (S6-G)</p> <p>“I am a little worried about the transition to upper secondary and regular French classes. My students will have more trouble with verb conjugations on assessments (non-communicative) than others. On the other hand, this is perhaps just a matter of learning by heart.” (S4-F)</p>
Tracking progress/keeping an overview of individual students during formative process	5	<p>“Focusing on formative activities phase 1, 2, 3 is difficult for me even though I have regularly tried to do it.” (S5-G)</p> <p>“What I would have liked to do better was to consistently formally (and visibly) assess students.” (S7-S)</p>

### **Teacher interviews**

The two teachers interviewed mentioned similar points as reported in the logbooks. Furthermore, they indicate their willingness to continue teaching and assessing language skills in a communicative way: "I want to continue to train language skills in every class and assess it in a communicative way. I like it and it motivates me." T2

The reorganization of the working time during the lessons was mentioned as a positive aspect as well, as it gave more time to put knowledge into practice and for practicing speaking skills in particular: "What I like is to have more time to practice speaking skills in class. We are not only busy with grammar because that's not assessed like that anymore and when it's time for the assessment of the speaking skill, it's not a surprise for students because we practiced a lot in class." T2

However, the teachers interviewed also indicated that they were confronted with difficulties, mainly of a practical nature, concerning the lack of time and materials available to develop communicative learning and assessment activities, which made the introduction into practice less elaborate and less faithful than they aimed for: "It was also not developed and well-presented enough due to lack of time." T1

In addition to these practical difficulties, teachers mentioned a lack of confidence in the quality of the materials they developed: "I don't know if my assignments are all communicative enough and meet all the criteria." T2

They also indicated lacking expertise on how to work with the CEFR in the selection of activities at the right level and how to assess students using CEFR criteria:

"I found it difficult to do something communicative at level A1 because they can't do much yet." T2

"I found it difficult to find documents at the right level and therefore did not sufficiently assess the receptive skills in a communicative way." T1

"I found the rating scales based on the can-do statements too complicated. It takes me too much time." T2

We also noticed during the working sessions that participants mainly had concerns regarding the selection of materials at the right level of the CEFR. This was an important point of discussion in the group since the learning objectives based on the CEFR are the operationalization of the concept of CLT. This concern therefore came back in the different stages of the development process but especially when teachers had to adapt

and use the rating scales. Using the can-do statements of the CEFR to assess students was difficult for teachers and thus often seen as time consuming.

### **Student survey**

Results from the survey show how students experienced the implementation of the new CBA program in one of their FL classes. When asked about the main focus in their lessons, the majority of the students (54%) answered that the focus was on learning to communicate in the target language (table 4.6). Table 4.7 shows which skills and knowledge are the most frequently assessed according to students. An almost equal proportion of the respondents report writing skills (28%), reading skills (26%) and speaking skills (20%) as being the most frequently assessed. A smaller percentage indicate that listening skills (9%) were assessed most frequently. Knowledge of grammar rules and words out of context is reported as being the most frequent focus of assessments by 17% of the respondents.

**Table 4.6** Main Focus in FL lessons according to students ( $N = 422$ )

Main focus in FL lessons	Frequency	Percentage
Learning to communicate in the target language	230	54%
Developing knowledge of grammar rules and vocabulary outside authentic situations ( <i>focus-on-forms</i> )	149	35%
Preparation for the final examination	28	7%
Developing content knowledge (e.g., culture, literature)	15	4%

**Table 4.7** Knowledge or skills addressed most frequently in assessments according to students ( $N = 423$ )

Knowledge and skills assessed	Frequency	Percentage
Writing skills	118	28%
Reading skills	108	26%
Speaking skills	83	20%
Listening skills	40	9%
Grammar rules and words	74	17%

### **Student interviews**

Students we interviewed were enthusiastic about the new assessment approach compared to the more traditional approach they experienced in previous years. Answers can be categorized into four themes: the communicative nature of the assessment activities, the fact that they had more time to practice language skills during the lessons, the insights into their progress, and finally their increased self-confidence. Table 4.8 presents a selection of quotes illustrating what the students particularly valued.

**Table 4.8** Student's main experiences regarding the implementation of the new CBA program

Main experiences	Quotes
Communicative focus of the assessment activities	<p>"I thought that learning a language was mainly learning words; now the focus is more on a theme. For example, we are now working on reading and writing, and the focus is more on understanding and sometimes the focus is more on speaking." (S2)</p> <p>"This year it is more improvising, that you have to react well at that moment and that you dare to speak. You cannot prepare it all in advance. You can prepare some things: words and grammar you will need." (S3)</p>
More time to put knowledge into practice	<p>"It is now mostly about speaking and other language skills." (S5)</p> <p>"The style of teaching is really different from last year. In year 2, the teacher would explain something and then the teacher would keep explaining and we would not have time to work on an assignment. It's much better now." (S3)</p> <p>"We also have much more time to practice during the lesson, not only to learn theoretical material." (S2)</p> <p>"I experience the lessons as positive. We work more during the lessons and get less homework!" (S1)</p>
More insights into learning progress	<p>"When you have an assessment, you know that you will be given tasks that you have already practiced." (S6)</p> <p>"You get quite a lot of feedback during the lessons. But that is not a bad thing. It is actually very good." (S2)</p> <p>"The teacher asks a lot of questions so she can see what we have learned." (S6)</p> <p>"Especially when we have done an assignment, I like to check it with the class at the end, so you know if you have done it right." (S5)</p>
Increased confidence	<p>"The teacher does not repeat everything in Dutch, and she really asks someone to answer in German. This way you build up confidence because you know that you can do it. It is less relaxing than learning grammar and words at home, but you learn more." (S1)</p> <p>"The focus is more on being able to speak French together and I think that gives us a more positive attitude which makes it easier to learn French." (S4)</p> <p>"Last year we had to learn a lot of words and if we didn't, the teacher got angry, and the class got more rowdy because she wasn't really positive. Now it's more the skills and it's more positive." (S6)</p>

## 4.5 Discussion and Conclusion

The aim of this study was first to investigate how language teachers implement a communicative CBA program based on tools (guidelines, formats, and examples) that have designed in a previous study (Chapter 3) and were made available to them in a toolbox (RQ4.1). The second aim was to determine how this implementation was experienced by the teachers and students involved (RQ4.2). The first research question was investigated through an analysis of the newly developed CBA programs and through classroom observations. The results show an acceptable to high level of treatment fidelity with regard to most “key ingredients” of the CBA programs for the majority of the language departments. The communicative character of the assessment tasks was visible in practice, both in the materials developed as well as in the informal assessment activities observed. These results contribute to the idea that involving teachers in the design and further development of an intervention leads to a high degree of treatment fidelity which, in turn, benefits sustainable change of classroom practice (Harn et al., 2013; Levin, 2004; Harding & Brunfaut, 2020).

To answer the second research question, aimed at exploring teachers’ and students’ perspectives on the implementation process, we first analyzed teachers’ logbooks on the implementation of the CBA programs they developed with the help of the toolbox. Most of the language departments involved were positive about the toolbox. They reported that the tools were inspiring and helped them develop their own CBA program. This result supports the idea that providing clear guidelines on classroom assessment practices in context can contribute to assessment literacy development (Firoozi et al., 2019; Berry et al., 2019). Furthermore, these results underpin the idea that an intervention with “built-in adaptations” not only offers flexibility but also works as a source of inspiration giving teachers the opportunity to develop high-quality interventions tailored to their own practice (Harn et al., 2013). Besides, teachers also noticed the fact that the project, involving language departments rather than only individual teachers, stimulated teamwork, and the exchange of good practices among colleagues, which contributed to their professional development. This confirms Lam’s (2019) recommendation to stimulate collaboration between colleagues to enable sustainable professional development.

Regarding the implementation of the CBA programs in practice, the results show that a large proportion of the teachers perceived more constructive alignment in their curriculum, experienced lessons as being more goal-oriented with more focus on language skills, and assessment activities as having a positive washback effect on their teaching. These results confirm the idea that a valid CBA program developed by FL teachers can promote a positive washback in their classroom, ultimately enhancing students’ learn-

ing (Hakim, 2018). Furthermore, a large number of the participants mentioned having more insight into students' progress and noticing better results in language proficiency as they were better able to adjust their teaching to the needs of their students, and as such to effectively contribute to their students' learning (Black & Wiliam, 1998). They also indicated having more time during the lessons to practice language skills and described students being more proactive. Finally, they stressed their enthusiasm about the new CBA program and noticed enthusiasm amongst their students as well. This aspect is certainly important, given the fact that pleasure and satisfaction stimulate motivation, which can be seen as the driving force behind FL learning (Ellis, 2019).

Even if teachers perceived the introduction in practice as positive in many respects, they also reported a number of issues, such as a lack of experience among students and teachers regarding CLT and the difficulty to keep track of individual students' progress in large classes. Some of the participants also mentioned a lack of school support in not providing sufficient time to learn and to experiment within the project. Finally, some of the teachers shared their concerns regarding the transition to upper form and "regular" classes and a possible lack of "basic knowledge" of words and grammar rules since these components were not taught and assessed directly. These concerns might reveal a limited expertise regarding CLT and on how to focus on form instead of on forms in teaching activities in order to achieve better and more sustainable learning results (Ellis, 2015). Results from the teacher interviews were very much in line with the results from the logbooks. The teachers interviewed mentioned having difficulty working with the can-do statements to assess their students. This lack of experience with the CEFR seemed to contribute to extra workload, as it took teachers more time than usual to select and develop suitable activities. This result shows the importance of training prospective teachers on how to implement the CEFR in their own practice (Fischer, 2020).

A large number of the students surveyed perceived assessment activities being more focused on the development of language skills more than before the project period, with more emphasis on productive skills. This finding is particularly remarkable since in our earlier study exploring current teaching and assessment practices in Dutch FL classrooms (Chapter 2), it was precisely these skills that students mentioned as being the least addressed in lessons and assessments. This suggests that working with the toolbox stimulated teachers in this project to teach and assess productive skills more often. Finally, students we interviewed stressed the fact that they felt more confident to pass the formal assessments since they had time to train and practice during the lessons in informal assessment activities. This last observation is important because self-confidence contributes to learning success. The higher the level of self-confidence,

the more learners dare to express themselves in the FL, and the more likely they are to reach a higher level of language proficiency (MacIntyre et al., 1998).

#### **4.5.1 Limitations of this Study**

Although the results of this study are quite positive regarding both the extent to which teachers succeeded to implement a communicative CBA program as intended (material development and introduction in practice), as well as with regard to how teachers perceived the implementation process and how students experienced it, some limitations should be reported. Despite the fact that teachers were provided with concrete tools designed by other teachers and adaptable to their own context, some of them still experienced difficulties in implementing their CBA program. The lack of experience with using CEFR can-do statements in their assessments, for instance, shows that more professional development is needed.

Furthermore, teacher and student interviews were held with a very small sample, limited to two language departments of one school; although the data were insightful and in accordance with insights gathered from group data, they do not guarantee that teachers and students from other schools would experience the implementation in the same way. Moreover, the mainly positive experiences reported by students might have been affected by socially desirable behavior due to the interview setting with the researcher. Finally, the implementation took place partly during the COVID-19 pandemic, which impacted the degree of treatment fidelity since teachers had to adapt materials to online teaching and were forced to make pragmatic choices, such as reducing the number of formal assessments planned.

#### **4.5.2 Recommendations for Teaching Practice**

The results of our study highlight the importance for teachers to gain expertise on how to develop a communicative CBA program in line with communicative learning objectives, with a right balance between formal and informal assessment activities. As recommended by Berry and colleagues (2019), FL teachers need a contextualized assessment training in which they can gain experience and confidence regarding the implementation of CBA in their own practice. This implies that in-service teachers need to be allocated time and opportunity to practice working with the CEFR, to learn how to select materials appropriate to the level of their students, and to experiment with developing and using assessment instruments based on it. Such room for experimenting with the implementation of communicative CBA programs is also necessary for both teachers and their students to get used to this way of assessing. Furthermore, this study shows the importance of providing teachers with the opportunity to collaborate as a team, focused on the same curriculum development project. We therefore encourage

collaborative work between FL departments within the same school or within the same educational context.

**CHAPTER 5**



# The Effects of the Implementation of a Communicative Classroom-based Assessment Program on Foreign Language Teaching Practices

This chapter is in preparation for submission as: Rouffet, C., Van Beuningen, C., & De Graaff, R. The effects of the implementation of a communicative classroom-based assessment program on foreign language teaching practices in Dutch lower form secondary education.

## Abstract

Classroom-based assessments (CBA) for Foreign Language (FL) education should be aligned with communicative learning objectives and organized in a program with the right balance between formal and informal assessments to enable effective teaching and learning. As demonstrated in Chapter 2, current assessment practices in Dutch secondary FL education do not meet these requirements and seem to exert a negative washback effect on teaching practices that, as a result, are often not communicative (enough). The current study investigates the observed and perceived effects of the implementation of communicative CBA programs that FL teachers, from 10 schools and 14 language departments, developed based on tools (e.g., guidelines, formats and examples) designed in co-creation with other FL teachers (see Chapter 3). In this evaluative study teaching practices in intervention classes (with a communicative CBA program) were compared with teaching practices in regular classes (with a non-communicative CBA program). Data were collected through lesson observations, teacher logbooks and student surveys. Results indicate that, compared to the observed teaching activities in regular classes which exhibit a traditional focus-on-forms, teaching activities in intervention classes adhered significantly more to communicative principles. Moreover, results show that both teachers and students valued the systematic introduction of communicative informal assessment activities in the program as it gave them insights into the learning progress and helped them to effectively steer FL teaching and learning.

## 5.1 Introduction

Over the past decades, in most educational contexts, the learning objectives of foreign language education have been communicative. This choice is motivated by the principle that the ability to communicate in foreign languages (FL), both in written and spoken interaction, is essential in a multicultural and globalized society as it enables communication across borders and a better understanding of other cultures. The positive effect of a communicative language teaching (CLT) approach on the development of learners' language skills has been theoretically and empirically supported since its introduction (Savignon, 2017). However, there are many different ways to translate CLT into teaching practices, and the effectiveness of the communicative approach in developing students' FL skills depends on the actual teaching practices used in a specific context.

### 5.1.1 CLT in Practice

CLT has gained worldwide popularity over the past decades. However, the application of CLT principles in the classroom remains challenging (Dos Santos, 2020; Kissau et al., 2015; Little, 2007). This may be due to contextual factors such as the (usually small) number of teaching hours per week, the (usually large) number of students in the classroom, or the materials and preparation time available (Ahmed, 2016; Chang, 2011; Coskun, 2011; Sarab et al., 2016; Wang & Cheng, 2009). But more conceptual factors also play a role, such as the way CLT is conceived by teachers (Graus & Coppen, 2016; Tang et al., 2012; Wang & Cheng, 2009).

When the communicative approach became popular and was increasingly adopted in FL teaching programs around the world, there were initially two different interpretations. The first, based on Krashen's (1984) natural learning theory, is called the *strong version* (Howatt, 1984). This version defines CLT as a meaning-oriented approach without any or little attention to formal aspects of the language, based on the assumption that, as for first language learners, the FL will be acquired naturally through reception and interaction. This strong interpretation has led to teaching practices that focused exclusively on meaning, with almost no attention paid to formal linguistic aspects. Various studies, as synthesized by Spada (2014), have shown that with this so-called strong approach, learners develop good comprehension skills (Genesee, 2004) and a fairly high level of vocabulary knowledge, but only with considerable exposure which is usually not available in secondary FL education (Lightbown et al., 2002). It was also observed that learners in exclusively content-oriented communicative programs had difficulties with grammatical accuracy in their oral and written production (Lightbown & Spada, 2013; Lyster, 2007). In the other interpretation of CLT, called the *weak version*, teaching practices should pay attention to both form and meaning (Howatt, 1984). This second version has in turn been interpreted in different ways, with, on the one hand,

a form-based approach with explicit and structural focus on predetermined formal aspects of the language to be learned prior to the development of language skills, and on the other hand, a meaning-based approach with an implicit focus on formal aspects when these emerge naturally in authentic communicative situations. This distinction has been referred to by Long (1991) under the terms *focus-on-forms*, corresponding to a traditional explicit and systematic instruction of aspects of the language system, and *focus-on-form*, corresponding (originally) to the implicit treatment of these aspects within meaningful situations. Long stressed the importance of including a focus on form in CLT in order to facilitate incidental language learning while keeping a primary focus on meaning. The definition of the concept has evolved over the years (Ellis, 2015). In today's interpretation, focus-on-form may no longer be exclusively implicit but may include the explanation of an explicit grammar rule, planned or unplanned, as long as it is provided within a communicative situation (Long, 2009). In the same vein, Spada (2014) defined form-focused instruction as any effort to draw learners' attention to formal aspects of the language in communicative and meaning-based settings.

Studies on the effects of form-focused instruction on FL learning have been conducted in different contexts (Ellis, 2015; Spada, 2014). Overall, results indicate that communicative instruction including attention to form in an integrated way is more effective than instruction that exclusively focuses on either form or meaning, especially when students do not have the opportunity to practice a language in an intensive way (Ellis, 2015; Lightbown, 2000; Lightbown & Spada, 2013; Spada & Tomita, 2010). Therefore, for FL acquisition to be successful in the context of secondary education with few teaching hours and large classes, teaching practices should give learners the opportunity to practice language skills in a realistic communicative context by carrying out authentic and meaningful tasks, supported by implicit or explicit form-focused instruction within this communicative context. Furthermore, by providing scaffolding tools for the development of communicative strategies and feedback on authentic communicative performances of students, teachers enable them to improve their language skills in relation to the learning objectives to be achieved (Perrone, 2011; Muranoi, 2007; Robinson, 2001; Swain, 2005). To compensate for the lack of exposure to the target language, secondary school FL teachers should furthermore use the target language not only frequently as a communicative tool but also adequately and consciously as a learning tool, drawing learners' attention to the language aspects to be learned (Dönszelmann et al., 2019).

Remarkably, the weak version of CLT has often resulted in teaching practices with a structure-based, focus-on-forms approach in which predetermined formal aspects are taught out of context and are directly assessed (Lightbown & Spada, 2013). This observation was also made in our recent study in the context of lower form secondary education in the Netherlands (Chapter 2) in which we found that FL teaching activities

and classroom assessments predominantly focused on the explicit knowledge of grammar rules and vocabulary out of context prior to the development of language skills, creating a lack of constructive alignment between national CLT-oriented learning goals on the one hand (College voor Toetsen en Examens, 2019) and the observed teaching activities and assessments on the other.

### 5.1.2 Constructive Alignment and Washback Effect

As Biggs (1996) states, aligning learning objectives with teaching activities and assessment practices in a curriculum can improve students' learning and achievement. In language teaching, if the learning objectives are communicative, it is therefore crucial that teaching practices and assessments are both related to these objectives. In *constructive alignment*, assessments particularly play an important role as various studies have shown that assessments influence both teaching practices and the way students learn (see review Chan, 2020). This influence, known as *washback* (Alderson & Wall, 1993; Green, 2007), can be positive when assessments are aligned with learning objectives, but can also be negative when alignment is lacking. Most studies on washback have been conducted in centralized educational systems with large-scale national examinations and specifically for English as a foreign language (Muñoz & Alvarez, 2010). Fewer studies have investigated the potential washback effects of classroom-based assessments (CBA) developed by FL teachers themselves, while in many educational systems, such as the Dutch one, FL teachers are frequently required to select, adapt and develop their own assessment materials for use in the classroom (Black & William, 2018; Scheerens, 2016).

### 5.1.3 Classroom-based Assessment

Assessments are classroom-based when they are developed and administered by teachers themselves. CBA can have a summative function, aimed at establishing students' level at the end of the learning process but should always have a formative function as well, aimed at providing insights into students' learning progression (Black & William, 1998; Dixon & Worrell, 2016). Hill and McNamara (2012) defined CBA as teachers' and students' reflection on students' skills and the use of this information for teaching and learning. This information collected at key moments of the learning process should help FL teachers shaping their teaching practices and inform students on how to improve the development of their language skills in relation to the learning objectives to be achieved (Purpura, 2008). The effects of CBA on students' learning have been recognized in several studies (Cheng, 2005; Muñoz & Alvarez, 2010; Perrone, 2011; Wall, 2013). However, CBA can take many forms and the ways in which assessment activities are developed and put into practice are also highly diverse and context specific (Saito & Inoi, 2017). Turner and Purpura (2016) made the distinction between spontaneous assessment activities, planned informal assessment activities, and formal assessment activities. Hill and McNamara (2017) developed a framework describing CBA activities ranging from the most

informal and formative (e.g., feedback in interaction, peer feedback or self-assessment) to the most formal assessment activities with a strong summative function (e.g., oral, or written examination). However, notwithstanding such frameworks, FL teachers still lack tools (e.g., concrete examples) to develop, select, plan, and implement a formal and informal communicative CBA program into practice (Firoozi et al., 2019; Tsagari & Vogt, 2017).

Moreover, most of the studies on the implementation of CBA are conducted within a centralized educational system and often investigate one specific type of CBA. Fewer studies have measured the washback effects of the implementation of a complete CBA program, combining informal and formal assessment activities, on teaching practices from a curriculum alignment perspective (Sultana, 2018). This seems to be a gap, as a CBA program aligned with the communicative curriculum goals could be expected to ensure a better application of CLT practices and, as such, contribute to the efficient development of language skills.

#### **5.1.4 Current Study**

##### ***Educational context***

The Dutch educational system is decentralized, which means that the role of the government is limited to determining the subjects to be taught, the objectives to be achieved, and the development of a part (50%) of the final examination. For FL education learning objectives in official curriculum documentation are formulated from a communicative perspective with reference to the CEFR levels (College voor Toetsen en Examens, 2019; Meijer and Fasoglio, 2007). The government inspects schools with regard to the achievement of these learning objectives but does not prescribe specific teaching practices and assessments. Therefore, schools, language departments, and language teachers within these schools have considerable autonomy in the way they integrate the learning objectives into their teaching and assessment practices (Scheerens, 2016). In the lower form of secondary education in the Netherlands, all assessments are classroom-based with a predominantly summative function (see Chapter 2). They are often presented in a so-called *assessment program* in which all assessments are planned for a school year in a specific grade. This CBA program determines, to a large extent, what teaching practices look like. It is therefore essential that the assessments included in the CBA program are aligned with the communicative learning goals.

For practical reasons such as large classes or limited preparation time and school requirements, the majority of Dutch FL teachers base their teaching practices and classroom-based assessments on the textbook they use (Fasoglio et al., 2015). Most commercial textbooks for FL secondary education in the Netherlands offer a form-based or structure-based approach, more commonly known as the presentation-practice-pro-

duction (PPP) method (Anderson, 2017; Criado, 2013; Van Batenburg et al., 2020). In this method, teaching and practice are aimed at learning predetermined formal aspects of the language. These aspects are first presented explicitly in a specific order. Controlled activities out of context are then provided to practice them. Finally, students are required to apply the instructed forms in strongly guided language tasks, often similar to translation tasks. The primary focus is on the knowledge of the specific linguistic forms instead of on the achievement of clear communicative goals (Ellis, 2010). Similarly, assessment tasks mainly focus on the knowledge of instructed formal aspects instead of on students' ability to communicate in the FL.

### **Objective**

In the present study, corresponding to the last phase of a larger design-based research project, we report on the effects of the implementation of a communicative CBA program on teaching practices within the context of Dutch lower form secondary education. The study focuses on classes at the A2 level of the CEFR, which is the most common level to be attained for most languages taught in lower form in the Netherlands (i.e., English, French, German, Spanish). In previous project phases, tools that aim to support FL teachers in developing a communicative CBA program were designed in co-creation with teachers (Chapter 3) and evaluated in terms of practicality (Chapter 4). The current study investigates how the implementation of communicative CBA programs, which are based on the materials designed and trialed in previous studies, can have a positive washback effect on FL teaching practices. As such, the research question guiding the study is:

*RQ5: What are the observed and perceived effects of the structural implementation of a communicative CBA program on FL teaching practices in terms of teachers' pedagogical choices?*

## **5.2 Method**

We used a mix of quantitative and qualitative data collection methods to determine the effects of the implementation of a communicative CBA program (i.e., intervention) on teaching practices. Within a quasi-experiment in situ, we first used an indicator-based approach to assess the visible effects of the intervention on teaching practices during classroom observations conducted in classes with and without the intervention but within the same schools and with the same teachers. We then used a qualitative analysis to analyze teachers' logbooks to assess the perceived effects of the intervention on their teaching practices. Additional data were collected through student surveys to obtain more insight into everyday classroom practices from a student perspective.

## 5.2.1 Participants

Participants in this study are the same as in the previous study. Whereas in the previous study (Chapter 4) the focus was on the implementation process of a CBA program, and on how this implementation was perceived by teachers and experienced by students, the present study evaluates the effect of the implementation on actual teaching practices.

### *Language departments and FL teachers*

Fourteen language departments with 32 FL teachers from 10 schools in the Netherlands participated in this study: one English department, six French departments, six German departments and one Spanish department<sup>2</sup>. The number of participating teachers ranged from one to four per language department. All participating teachers taught classes at A2 level of the CEFR. In each of the departments at least one of the teachers taught at two classes in the same grade: one class (part of the control group) with a regular CBA program, largely based on assessments included in the textbook and similar to the CBA programs analyzed in our previous inventory study (Chapter 2), and one intervention class in which a communicative CBA program was implemented. All participating teachers hold a teaching certification and had various years of experience ranging from two to twenty-five years.

### *Students*

Two cohorts of students from A2-level classes of all participating schools and language departments were involved in the study. They were all lower form students aged 12 to 15. Students in the first cohort (part of the control group) attended the regular language classes and CBA programs of the teachers involved during the 2019-2020 school year prior to the implementation of the newly developed CBA programs. Students in the second cohort (intervention group) were part of the intervention classes in which the communicative CBA programs were introduced during the 2020-2021 school year. In the first cohort 330 students filled in the survey about the language lessons taught by one of the participating teachers: English ( $n = 67$ ), German ( $n = 131$ ), French ( $n = 110$ ), Spanish ( $n = 22$ ). In the second cohort 423 students filled in the survey: English ( $n = 46$ ), German ( $n = 158$ ), French ( $n = 211$ ), Spanish ( $n = 8$ ).

## 5.2.2 Intervention

The intervention consisted of two parts: 1) supplying teachers with tools for the development of a communicative CBA program, and 2) the development of one's own communicative CBA program based on these tools, and its introduction in teaching practice.

---

<sup>2</sup> This is a convenience sample that does not reflect the distribution of different FLs in the Dutch educational system (English being the most widely taught FL in the Netherlands).

### ***Supplying tools for the development of CBA programs***

Given the Dutch context in which schools, language departments and teachers themselves have considerable autonomy, we decided not to offer an intervention in the form of a predetermined program, but to design guidelines, formats, and examples for the development of a communicative CBA program instead. This type of intervention in which participants can adapt the materials to their own practice is called an intervention with “built-in adaptations” (Harn et al., 2013, p.188). The core components and essential features of such interventions are fixed, but participants can introduce them in different ways, depending on their context.

The materials for this study’s intervention were designed in co-creation with FL teachers in an earlier research phase (Chapter 3) and collected in a toolbox made available to the participants of the current study on a digital platform (see Appendix B). The different tools were designed according to CLT principles and aligned with communicative learning objectives. We first determined which core components (e.g., formal, and informal assessment tasks, rating scales, assessment planning) should be included in the CBA program and we then formulated design principles (Chapter 3, Table 3.2). To encourage a positive washback effect on teaching practices in the context of secondary education with few teaching hours and large classes, we based our design on the interpretation of CLT in which teaching practices should give learners the opportunity to practice language skills in a realistic communicative context by carrying out authentic and meaningful tasks with an appropriate degree of free production supported by implicit or explicit form-focused instruction within this communicative context. As such, a communicative CBA program should include summative assessments assessing the ability to communicate in a FL through authentic and meaningful tasks in which skills are assessed in an integrated way matching real-life communicative situations (Morrow, 2018). Moreover, a communicative CBA program should also include assessment activities with an exclusively formative function on which students can receive feedback and further (form-focused) instruction to help them improve their FL skills without being formally assessed (Black & Wiliam, 2018).

### ***Development and introduction of CBA programs***

Teachers participating in the current study developed their CBA programs based on the materials made available to them in the toolbox. They then introduced the newly developed CBA programs in practice during the whole school year 2020-2021. All CBA programs contained the same components developed in accordance with CLT design principles: an assessment planning, informal assessment activities, formal assessments, and rating scales. The analysis of the newly developed CBA programs conducted in a previous study (see Chapter 4) showed an acceptable to high level of treatment fidelity for most language departments with regard to the different “key ingredients” of the CBA

programs. The communicative character of the assessment activities and the link to the CEFR were clearly included in the materials and visible in the majority of the classroom observations.

**Table 5.1** Core components and “key ingredients” of a communicative CBA program

Core components	“Key ingredients”
Communicative assessment tasks	<ul style="list-style-type: none"> <li>· Meaningful: connected to students’ perceptions</li> <li>· Authentic: linked to real-life situations</li> <li>· Unpredictable and creative: encourage to respond to unprepared questions with an appropriate degree of free production</li> <li>· Integrate language skills</li> </ul>
Formal assessments (summative and formative function)	<ul style="list-style-type: none"> <li>· Include an explicit reference to CEFR can-do statements</li> <li>· Include a variety of communicative tasks with an increase in difficulty</li> </ul>
Informal assessment activities (exclusive formative function)	<ul style="list-style-type: none"> <li>· No part of formal evaluation</li> <li>· Include (part of) communicative tasks</li> <li>· Include a formative goal according to one of the steps of the formative cycle (feed-up, feedback or feedforward)</li> </ul>
Rating scales	<ul style="list-style-type: none"> <li>· Bring together criteria related to the can-do statements of the CEFR</li> <li>· Value the use of communicative strategies</li> <li>· Enable the assessment of several language skills in an integrated way</li> <li>· Can be used as a summative and formative tool</li> </ul>
Assessment planning	<ul style="list-style-type: none"> <li>· Include the formal and informal assessment of all language skills</li> </ul>

### 5.2.3 Data Collection and Analysis

#### *Classroom observations*

Due to restrictions related to the COVID-19 health crisis, not all planned classroom observations could be carried out. Eventually, 12 (of the initial 14) FL teachers (T1 to T12) from 12 language departments who were willing to participate were observed twice, giving a lesson at level A2 in a regular class without intervention (control group) and in an intervention class (experimental group). The teachers were observed using an observation tool developed by the Foreign Language Education Research Group at the Center of Expertise: Learning and Innovation of Utrecht University of Applied Sciences (see Appendix A). This observation tool maps the extent to which teaching practices of secondary FL teachers adhere to CLT principles. The instrument focuses on the quality of communicative teaching activities and on observable teacher behavior in introducing CLT. The instrument measures six quality characteristics of communicative teaching activities: meaningfulness, authenticity, appropriate degree of free production or relevant comprehension activities, integration of language skills; and five aspects related to a teacher’s behavior in introducing these activities in practice: use of scaffolding,

development of strategies, appropriate use of formative tools, form-focused instruction (explicit or implicit), and appropriate use of the target language (as a communicative tool and as a learning tool). Each characteristic can be scored on a four-point Likert scale representing the extent to which FL teachers implement the principles of communicative language teaching, ranging from 1 (*not visible, although the situation called for it*) to 4 (*visible to a large extent*). During the observation, the lesson goals were identified. After the classroom observation the teachers observed were asked to confirm and, if necessary, clarify these goals.

All observations were performed by the same observer (i.e., the researcher). The same instrument had already been used in a previous study (Chapter 2) by the researcher and members of the research group who co-developed the instrument. In this previous study, inter- and intra-rater reliability was checked using Cohen's  $\kappa$  measurement. Substantial agreement between the ratings was found (inter-rater reliability:  $\kappa = .617$  (95% CI, .499 to .734),  $p < .0001$ ; intra-rater reliability:  $\kappa = .616$  (95% CI, .498 to .698),  $p < .0001$ ). The scores from the Likert scale were used to generate quantitative data that were first analyzed descriptively for each item (i.e., meaningfulness, degree of free production, etc.). Paired-samples  $t$ -tests were then conducted to determine if there was a significant difference between the observations of regular classes and intervention classes. The conditions for parametric tests were met.

The lesson goals identified were first coded into three categories: 1) focus on the development of communicative FL skills, 2) focus on the development of knowledge of grammar rules and vocabulary outside authentic situations (focus-on-forms), and 3) focus on the development of communicative FL skills, as well as knowledge of grammar rules and vocabulary outside authentic situations (focus-on-forms). The data obtained were then analyzed descriptively.

### ***Teacher logbooks***

We asked participating teachers ( $N = 32$ ) from each language department involved ( $N = 14$ ) to complete a logbook during the implementation of the communicative CBA program in school year 2020-2021. Teachers represented their language department and completed their logbook as a team. We asked them specifically to report on the extent to which the new way of assessing their students affected their teaching practices and in which way. Data were first pseudonymized per school S1 to S10. The letters E (English), F (French), G (German) and S (Spanish) were added to the codes to indicate the different language departments. The logbooks were first analyzed interpretatively (Corbin & Strauss, 2008). Responses about the perceived effects of the intervention on teaching practices were coded. Codes were short sequences of words that describe general patterns. We then looked at frequencies and percentages of codes in order

to synthesize the categorized data into general findings. In order to further illustrate, clarify and explain these quantitative data, logbook excerpts were selected and added to the results section.

### ***Student survey***

Surveys among students in regular classes of the first cohort (control group), at the end of school year 2019-2020 ( $N = 330$ ) and among students in intervention classes in the second cohort (experimental group), at the end of school year 2020-2021 ( $N = 423$ ) were conducted to compare the students' perspectives on teaching practices with and without the implementation of a communicative CBA program. We used online surveys with multiple-choice questions, for which students had to choose one answer from a predetermined list of options. The questions for this analysis focused on students' perceptions of: 1) the knowledge and skills most frequently addressed in lessons, 2) the main focus of summative assessments, 3) the main focus of formative assessment activities, and 4) the use of the target language by the teacher. Data were first analyzed descriptively. As the data are correlated and clustered, students being nested in schools, language departments and classes with different teachers, we had to take into account the possible interference of these random factors. We therefore fitted a generalized linear mixed model (GLMM) for each of the four dependent variables (questions) to determine if the intervention made a significant difference between the control group and the experimental group when taking random factors into account. The GLMM's were run using IBM SPSS Statistics 28. We first determined which model fitted the data best. Given the fact that there was often an overlap between participating school and language department, as in most cases only one language department participated per school, we expected that teacher as random factor would explain the most variance. To check this assumption, we first calculated the ICC value of each possible random factor. As expected, teacher as random factor explained the most variance ( $ICC = .50$ ). We first ran the models with this intercept only, and then ran them a second time with the intervention as fixed factor to check if the models fitted the data. As the distribution of the data appeared to be too unbalanced to fit the model with a multinomial analysis, we chose to perform a binomial analysis instead. In order to do so, we simplified the data by combining the different predetermined answers for each of the dependent variables into two main categories: 1) focus on formal aspects of the language, versus 2) focus on the development of language skills for the three first questions, and 1) frequent use of the target language versus 2) non-frequent use of the target language for the last dependent variable. In total, four GLMMs were run separately with the following dependent variables: main focus of teaching activities, main focus of assessment activities, main focus of formative assessment activities, use of the target language by the teacher.

## 5.3 Results

The results reported below show the observed and perceived effects of the structural implementation of a communicative CBA program on teaching practices in intervention classes compared to regular classes.

### 5.3.1 Observed Effects

#### *Classroom observations*

Table 5.2 shows the type of lesson goals formulated by FL teachers in observed lessons for regular and intervention classes. Whereas in regular classes only 33% of the lesson goals focused exclusively on the development of communicative FL skills, in the intervention classes all lesson goals focused on that aspect. In regular classes, 25% of the lesson goals were formulated exclusively in terms of traditional focus-on-forms, and 41% focused on forms partially.

**Table 5.2** Lesson goals in observed lessons in regular classes ( $N = 12$ ) and in intervention classes ( $N = 12$ )

Focus of the lesson goals	Regular classes		Intervention classes	
	<i>N</i>	%	<i>N</i>	%
Developing communicative FL skills	4	33%	12	100%
Developing knowledge of grammar rules and vocabulary outside authentic situations (focus-on-forms)	3	25%	0	0%
Developing communicative FL skills, as well as knowledge of grammar rules and vocabulary outside authentic situations (focus-on-forms)	5	41%	0	0%

Results in table 5.3 show the difference between the two groups regarding the overall communicative nature of the teaching practices observed and the adherence to the different CLT criteria.

**Table 5.3** Difference between intervention classes ( $N = 12$ ) and regular classes ( $N = 12$ ) regarding the adherence to CLT criteria of teaching practices during observed lessons

Range 1-4	Regular classes		Intervention classes		<i>t</i>	<i>df</i>	<i>p</i>	Cohens' <i>d</i>
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>				
Meaningful activities	2.37	1.28	3.38	.92	-2.2	11	.024	.64
Appropriate degree of free production	2.00	1.22	3.56	.73	-3.5	9	.004	1.2
Use of functional comprehension activities	1.16	1.04	2.67	1.15	-3.0	2	.048	1.7
Authenticity	1.53	1.22	2.93	1.06	-3.1	11	.005	.89
Appropriate scaffolding	1.80	1.29	2.38	1.18	-1.6	11	.063	.48
Development of communicative strategies	1.73	1.27	3.13	1.11	-3.8	11	.001	1.1
Formative activities	1.84	1.30	3.04	.81	-2.8	11	.009	.79
Skills integration	1.2	.39	2.6	1.04	-4.8	11	.001	1.38
Focus on form	1.6	.56	2.6	.93	-4.2	11	.001	1.20
Use of the target language as a learning tool	2.33	1.07	3.00	1.16	-1.4	11	.08	.58
Use of the target language as a communication tool	2.17	.39	2.42	1.11	-5.4	11	.30	.21
Overall communicative value	2.03	1.07	3.04	.66	-3.3	11	.003	.97

1 = not seen, although the situation called for it; 2 = barely seen; 3 = sufficiently seen; 4 = seen to a large extent

Comparing the overall communicative value of teaching practices in the intervention classes ( $M = 3.04$ ,  $SD = 0.66$ ) to those observed in the regular classes ( $M = 2.03$ ,  $SD = 1.07$ ), we observe a significant and large difference between the two groups, [ $t(11) = -3.3$ ,  $p = .003$ ], with teaching activities within intervention classes showing more adherence to CLT criteria than those within regular classes for all criteria except scaffolding.

Regarding the integrated teaching of language skills we see a significant difference between the two groups, [ $t(11) = -4.8$ ,  $p < .001$ ] with more integration in the intervention classes than in the regular classes with a large effect size. Likewise, there was significantly more focus on form in the intervention classes than in the control group (where there was more focus on forms) [ $t(11) = -4.2$ ,  $p = .001$ ], again with a large effect size. In both cases the intervention classes show better results (more visible application of CLT principles) than the regular classes.

Regarding the use of the target language, results of the paired *t*-tests show no significant difference between the two groups. However, with respect to the use of the target language as a learning tool, the medium effect size could be interpreted to signal a trend: teachers seem to use the target language as a learning tool more often in intervention classes than in regular classes.

### 5.3.2 Perceived Effects

#### *Teachers' logbooks*

The perceived effects of the intervention regarding pedagogical choices, as reported by teachers of the FL departments ( $N = 14$ ) involved, are gathered in Table 5.4 below. Teachers' mentioning of effects regarding other aspects, such as students' motivation or practical issues were not part of the current analysis, as such aspects were already in focus in our previous study on the implementation process (see Chapter 4).

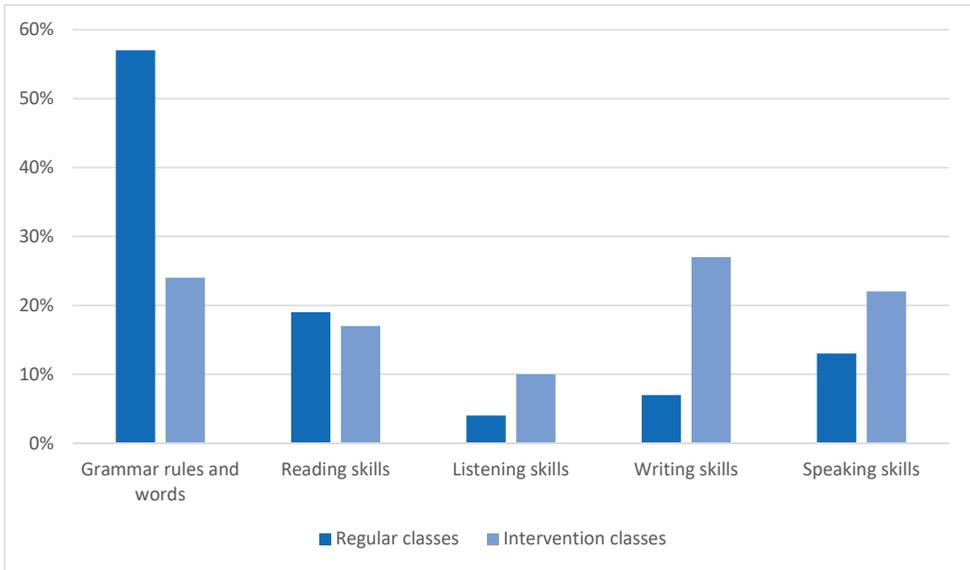
**Table 5.4** Perceived effects of a communicative CBA program on teaching practices in terms of pedagogical choices reported in logbooks by the FL teachers of 14 FL departments

Perceived effects	Frequency of the FL departments (%)	Quotes
More goal-oriented	7 (50%)	"The goals from the rating scales were always central; you could fall back on them and refer to them. More is being done with the long-term goals, practicing more language skills." (S7-E) "I notice that I pay much more attention to communication skills because I want students to be well prepared for the assessment." (S3-G)
More focus on productive skills	6 (43%)	"We have been practicing speaking skills all the time in class." (S5-F) "I do a lot of speaking and writing activities and link reading to writing, as will be done in the assessment." (S3-G)
More use of the target language	5 (36%)	"I notice a difference in my teaching style because I speak much more French, at my own initiative and so do the students." (S1-F) "I personally notice that I speak more German in class compared to last year. In previous years I started well but weakened after the fall break." (S10-G)
More focus on form instead of focus on forms	5 (36%)	"Particular grammar rules come up because of the conversations tasks students have to perform. Rules are explained at that time and are immediately applicable." (S5-F) "We are practicing language skills all the time and vocabulary and grammar only when it is relevant to the tasks to be performed." (S7-E)

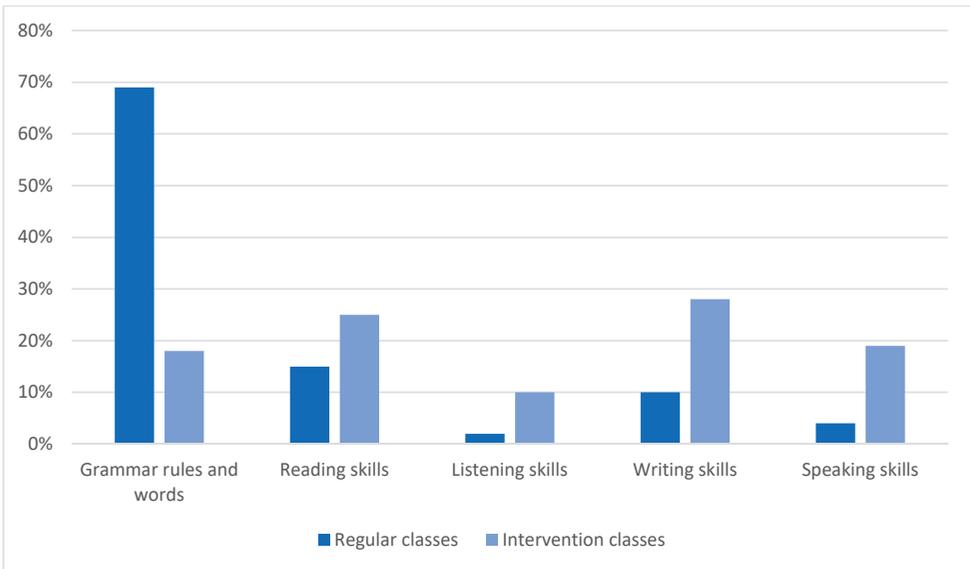
In most of the FL departments involved teachers report lessons being more goal-oriented, more focused on language skills, - and particularly productive skills - with a focus on form instead of on forms. They also report an increase in the use of the target language.

### ***Student surveys***

We first report on the results of the descriptive analysis. When comparing the answers with regard to lesson focus of students from the regular classes (control group) with those of students from the intervention classes (experimental group), the knowledge and skills most frequently addressed in intervention classes seem less focused on knowledge of formal language aspects and more on the development of FL skills, in particular on the development of productive skills (Figure 5.1). Figure 5.2 shows which knowledge and skills are the most frequently assessed according to students. Results are in line with those in Figure 5.1. Students of the intervention classes report less focus on formal aspects of the language in assessments and more on their language skills, particularly their productive skills.

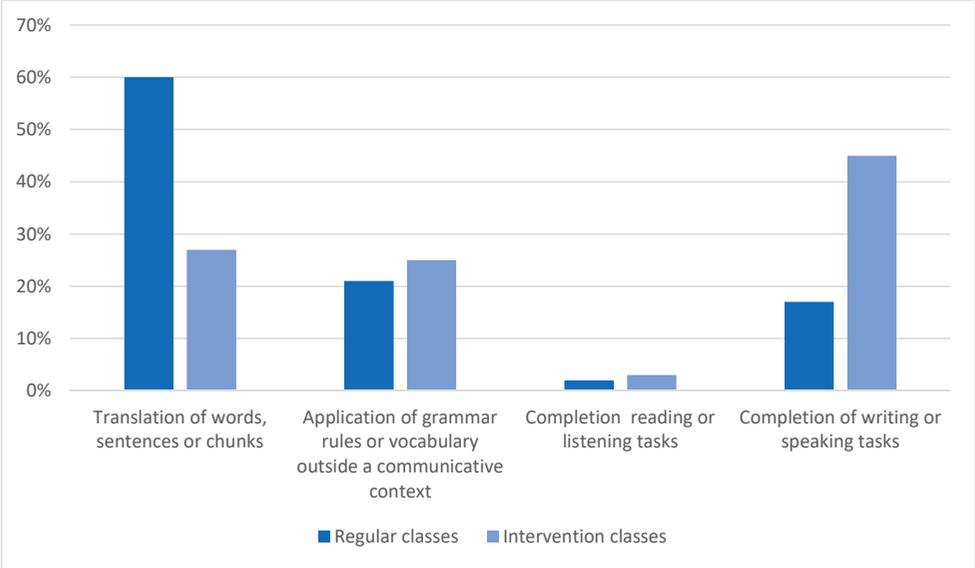


**Figure 5.1** Knowledge and skills addressed most frequently in *teaching activities* according to students from regular classes ( $N = 330$ ) and intervention classes ( $N = 423$ )



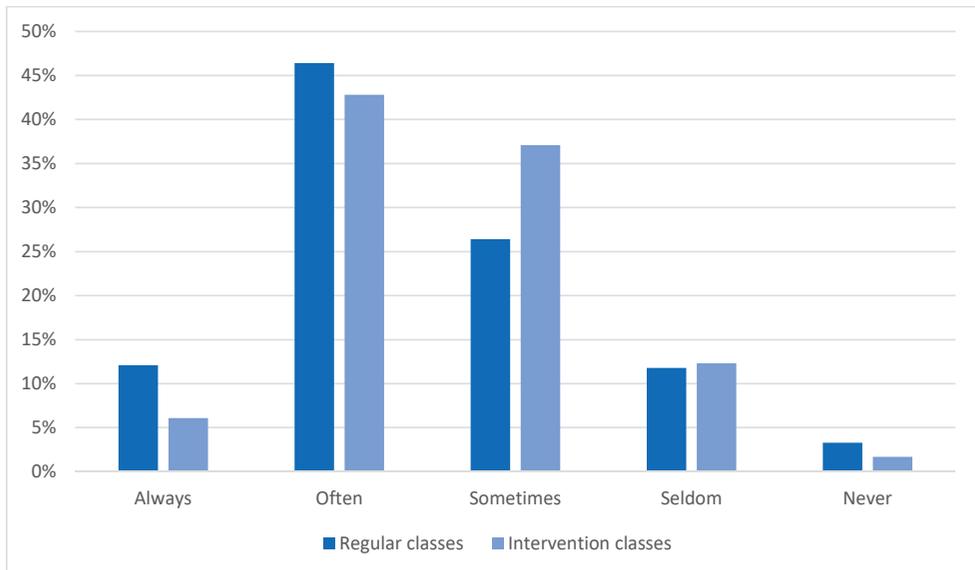
**Figure 5.2** Knowledge and skills addressed most frequently in *assessment activities* according to students from regular classes ( $N = 330$ ) and from intervention classes ( $N = 423$ )

Results reported in Figure 5.3 show which knowledge and skills formative assessment activities most frequently addressed during the lessons. Whereas more than half of the students in regular classes (60%) report the focus to be on the translation of words, sentences or chunks, a large number of the students in intervention classes (45%) report that formative activities involve the completion of writing or speaking tasks.



**Figure 5.3** Formative activities most frequently addressed according to students from regular classes ( $N = 330$ ) and from intervention classes ( $N = 423$ )

Results reported in Figure 5.4 might indicate that according to students, teachers seem to use the target language slightly less frequently in intervention classes than in regular classes.



**Figure 5.4** Use of the target language by the teacher according to students from regular classes ( $N = 330$ ) and from intervention classes ( $N = 423$ )

Although these descriptive results show that students in the intervention classes seemed to perceive less focus on formal aspects of the language and more on the development of communicative language skills in teaching practices and assessments (teaching activities, summative assessments and formative assessment activities), further GLMM analysis were performed to determine whether these differences were significant.

Results of the GLMMs show a significant degree of variability between the intercepts of each teacher for the two first models but not for the other two (model 1:  $VAR. = 0.517$ ,  $p = .041$ ; model 2:  $VAR. = 1.210$ ,  $p = .038$ ), so we included this random factor only in final models 1 and 2. The accuracy rates of the four final models are substantial and indicate that the models fit the data (Table 5.5).

**Table 5.5** Overall accuracy rates of the final generalized linear mixed models (GLMM's)

Final GLMM's	Dependent variables	Overall accuracy rate
Model 1	Main focus of teaching activities	73%
Model 2	Main focus of summative assessments	82.3%
Model 3	Main focus of formative assessment activities	70.1%
Model 4	Use of the target language by the teacher	70.7%

Table 5.6 shows the results of the four final models. In the three first models the category *language skills* as main focus is chosen as reference category. This means that the results presented in the table refer not to *language skills* as main focus but to *formal aspects of the language* as they contrast with the focus on the development of language skills. Regarding the fixed effect, intervention classes is the reference category. The results in the fixed effect row indicate how students from regular classes are more likely to perceive a focus on formal aspects of the language in comparison to students from intervention classes. In the last model the category *frequent use of the target language* is chosen as reference category as well as the intervention classes. The results in the fixed effect row in the last model indicate how students from regular classes perceived the use of the target language by the teacher in comparison to students from intervention classes.

**Table 5.6** Fixed and random effects of the GLMM's

<b>Model 1</b>		<b>Main focus of lessons</b>		
<b>Categories</b>				
Language skills	<i>Reference category</i>			
Formal aspects	<b>p-value</b>	<b>Odds ratio</b>	<b>95% confidence interval</b>	
			<b>Lower</b>	<b>Upper</b>
<b>Fixed effect</b>				
Intervention				
Regular classes N = 330	< .001	3.871	2.726	5.496
intervention classes N = 423	<i>Reference category</i>			
<b>Random factor</b>	.041			
Teachers				
<b>Model 2</b>		<b>Main focus of summative assessments</b>		
<b>Categories</b>				
Language skills	<i>Reference category</i>			
Formal aspects	<b>p-value</b>	<b>Odds ratio</b>	<b>95% confidence interval</b>	
			<b>Lower</b>	<b>Upper</b>
<b>Fixed effect</b>				
Intervention				
Regular classes N = 330	< .001	11.731	7.727	17.808
intervention classes N = 423	<i>Reference category</i>			
<b>Random factor</b>	.038			
Teachers				
<b>Model 3</b>		<b>Main focus of formative assessment activities</b>		
<b>Categories</b>				
Language skills	<i>Reference category</i>			
Formal aspects	<b>p-value</b>	<b>Odds ratio</b>	<b>95% confidence interval</b>	
			<b>Lower</b>	<b>Upper</b>
<b>Fixed effect</b>				
Intervention				
Regular classes N = 330	< .001	3.835	2.682	5.483
intervention classes N = 423	<i>Reference category</i>			
<b>Random factor</b>	<i>Not included</i>			
Teachers				
<b>Model 4</b>		<b>Use of the target language by the teacher</b>		
<b>Categories</b>				
Frequent use	<i>Reference category</i>			
Not frequent use	<b>p-value</b>	<b>Odds ratio</b>	<b>95% confidence interval</b>	

**Table 5.6** Fixed and random effects of the GLMM's continued

			Lower	Upper
<b>Fixed effect</b>				
Intervention				
Regular classes N = 330	.802	1.045	0.741	1.472
intervention classes N = 423	<i>Reference category</i>			
<b>Random factor</b>				
Teachers	<i>Not included</i>			

In model 1, results show that students in the regular classes perceived teaching activities to be focused more on formal aspects of the language and less on the development of language skills than students in the experimental group ( $p < .001$ ). Looking at the *odds ratio*, we can say that the probability of students from the regular classes to indicate lessons being mainly focused on the knowledge of formal aspects of the language is almost four times higher than that of students from the intervention classes ( $OR = 3.871$ , 95% CI [2.726, 5.496]). Results of model 2 show that, according to most students of the regular classes, the main focus of summative assessments is on the formal aspects of the language while in intervention classes it is on the development of language skills. This difference is statistically significant ( $p < .001$ ) with a large effect size ( $OR = 11.731$ , 95% CI [7.727, 17.808]), showing that the odds of students from the regular classes indicating that the emphasis in summative assessment is on formal aspects is almost twelve times higher than that of students from the intervention classes. Similarly, results from model 3 show that students from the regular classes perceived a focus on formal aspects of the language in formative assessment activities significantly more often ( $p < .001$ ) than students from the intervention classes for whom these formative activities focused mainly on the development of language skills. The proportion of students from the regular classes that indicate a main focus on formal aspects of the language in formative activities is nearly four times higher than that of students of the intervention classes ( $OR = 3.835$ , 95% CI [2.682, 5.483]). In model 4, results show no significant differences between the use of the target language by the teacher in regular classes compared to intervention classes ( $p = .802$ ).

## 5.4 Discussion

The objective of this study was to identify which observed and perceived effects the structural implementation of a communicative classroom-based assessment program has on FL teaching practices in the context of lower form secondary education in the Netherlands. Results of the study show that the implementation of a communicative

CBA program focusing on the development of language skills in summative assessments and formative assessment activities has a positive washback effect on teaching practices, as they become more communicative and thus more goal oriented. This can be seen as a positive effect since activities that are aligned with learning goals enable effective teaching and enhance students' learning and achievement (Biggs, 1996). Results also show that lessons observed in the intervention classes were all mainly focused on the development of language skills within a meaningful context, while in regular classes a large part of the lessons observed focused on forms (e.g., grammar rules) out of context. This result can also be seen as a positive washback effect since a focus-on-form instead of a focus-on-forms approach in CLT offers students better opportunities to develop their language skills (Ellis, 2015). Compared to the observed teaching activities in regular classes, teaching activities in intervention classes adhered to communicative principles significantly better. These activities were more meaningful, authentic, functional, or creative, asking for more free production. The teaching of language skills in intervention classes also took a more integrated approach, matching real-life communication. Regarding the use of the target language, we did not see a significant difference, but teachers observed in intervention classes did use the target language as a learning tool. This result can also be seen as a positive effect since the use of the target language as a learning tool tends to be more efficient in helping students develop FL skills (Dönszelmann et al., 2019). Finally, we noticed a more appropriate use of formative activities in intervention classes than in regular classes. These activities were directly linked to CLT goals and assessment criteria to provide students and teachers insight into the learning progress. This is a positive effect since providing formative activities aligned with learning goals helps students to achieve these goals (Black & Wiliam, 2018).

Most of the effects identified through classroom observations were also perceived by teachers and students themselves. In the majority of the language departments involved, teachers report that the implementation of a communicative CBA program - as compared to following a structure-based program - led to more focus on the development of language skills and particularly of productive skills, as well as to a focus on form within communicative activities instead of a focus on forms. Some of them also indicated using the target language more frequently. In the same vein, most students in the intervention classes reported lessons, summative assessments and formative assessment activities being mainly focused on the development of language skills, while according to students from regular classes the focus was mainly on knowledge of formal aspects of the language. There was no significant difference in how students from regular classes and intervention classes reported the use of the target language by their teachers.

The conclusion that we can draw from these combined findings is that the implementation of a CBA program aligned with CLT learning goals can change traditional form-focused teaching practices into a more communicative teaching approach, with a primary focus on the development of language skills (i.e., positive washback effect). However, the findings of this study have to be seen in light of some limitations. This evaluative study was partly conducted during the COVID-19 pandemic, and the restricted measures related to the health crisis had an unavoidable impact on our research plan and methodological choices. Originally, we planned to observe teaching practices before the implementation of the new CBA program in the second part of the 2019-2020 school year. The unexpected first lockdown in the second half of school year 2019-2020 prevented us from doing so. Therefore, we had to adapt the initial research plan by asking the participating teachers to teach at least two classes in the same grade: an intervention class in which they introduced a communicative CBA program for the school year 2020-2021, and a non-intervention class which followed the regular assessment program. Classroom observations in both intervention and regular classes were thus conducted in 2020-2021 during the same school year. As teachers generally do not have much time to prepare their lessons, it is possible that materials and teaching activities were used from one class to another, as these were parallel classes.

As the health crisis partly continued in 2020-2021, some of the restrictions still impacted the normal teaching practice and thus the project in the next school year. Due to a second lockdown, part of the classes between December 2020 and February 2021 were held online and some schools worked with half-size classes, not all classroom observations could be carried out, some participants could not provide all the data requested and some others dropped out of the project during the school year. Furthermore, we were not allowed to observe the lessons with two researchers, and it was too late to obtain the approval of the ethics committee to record the lessons on video. This factor might have negatively impacted the reliability of the observation data obtained. However, the fact that results from the student surveys are in line with those from classroom observations counterbalances the potential lack of reliability.

All these external factors related to the pandemic may certainly have influenced the results. However, we found that despite the particular circumstances, the majority of participating teachers remained in the project until the end and were enthusiastic about it, noticing positive effects on their teaching practices. Yet, it is important to note that the teachers participating in the project did so on a voluntary basis, were interested in the project and were willing to change their more traditional form focused way of assessing their students to a more communicative one.

Notwithstanding these limitations, we can conclude that the implementation of a communicative CBA program showed a positive washback effect on teaching practices as they became visibly and significantly more communicative, enhancing constructive alignment. Studies conducted in other educational systems had already demonstrated the importance of classroom-based assessment as a tool to improve classroom practices and students' learning (Cheng, 2005; Muñoz & Alvarez, 2010; Perrone, 2011; Wall, 2013). However, these studies were all conducted in centralized educational systems in which national high-stake examinations also have a strong influence on teaching practices. As many other studies on washback effects have shown (Kuang, 2020), teachers are strongly inclined to devote their teaching activities to the skills and knowledge being assessed in high stake examination more than to the curriculum goals to be attained (Stecher et al., 2000). For this reason, some researchers have stressed the importance of studying the washback effect of a whole CBA program from a curriculum alignment perspective by investigating its effect on teaching practices in relation to the learning goals to be attained beyond a final high-stake examination (e.g., Sultana, 2018). The results of our study show that a CBA program aligned with learning goals and providing a good balance between formal and informal assessment activities can be used as a steering tool to exert a strong positive washback effect on teaching practices and on the way students learn.

A large, light blue number '6' is centered on the page. A dark blue horizontal bar is positioned to the left of the number, overlapping its left side.

**CHAPTER 6**

# General Conclusion and Discussion



## 6.1 Aims and Setup of the Project

Foreign language (FL) teaching in secondary education in the Netherlands is mainly based on the principles of Communicative Language Teaching (CLT). In the decentralized Dutch educational system, in which only the learning objectives and part of the final examinations are prescribed at a national level, secondary school teachers have substantial autonomy in how to translate learning objectives into practice. For FL education, the learning objectives are formulated from a communicative perspective. However, although FL teachers value the communicative approach, they do not always succeed in translating the communicative learning objectives into their classroom and assessment practices. FL education, particularly in lower form, tends to focus more on knowledge of grammar rules, vocabulary and chunks taught out of context than on the development of language skills (Fasoglio et al. 2015; West & Verspoor 2016). Moreover, the communication levels students reach in subjects such as French and German have dropped and the motivation to learn these languages has severely decreased (Michel et al., 2021). Students who have attended FL classes throughout their entire school career until their final examination still experience difficulties or are not even able to communicate in these FLs. Of course, FL professionals such as language teachers and teacher educators seriously deplore these findings. It is not without reason that the Council of Europe stresses the importance of providing multilingual education that promotes linguistic diversity and the learning of multiple languages. Languages are a fundamental part of people's lives and learning to communicate in several languages enables better intercultural contact and understanding. There is an urgent need to change the way foreign languages are taught, however, for Dutch FL education to be effective in reaching such goals.

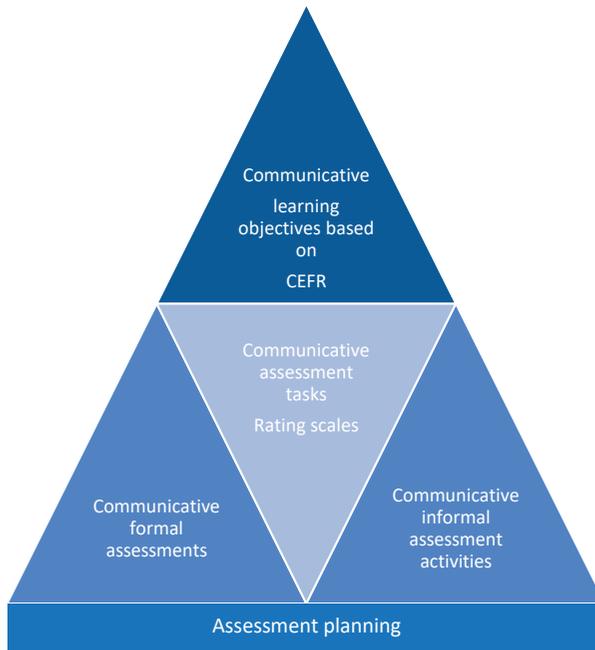
It has been well established that the main components of a curriculum - learning objectives, teaching activities and assessments - should be aligned to enable effective learning (Biggs, 1996). Assessments in particular play a crucial role, as they have an important washback effect on classroom practices and on the way students learn (Alderson & Wall, 1993; Green, 2007). Earlier studies on the washback effect of assessments in the context of FL education have mainly been conducted in centralized educational systems, investigating the effects of large-scale examinations developed by test experts (Chan, 2020; Rea-Dickins & Scott, 2007; Spratt, 2005). Fewer studies have looked at the effects of classroom-based assessments (CBA) developed and administrated by FL teachers themselves, as they are often used in the context of a decentralized system. The current project aimed to fill this gap by designing tools for the implementation of a communicative CBA program that could generate a positive washback effect on FL teachers' pedagogical choices.

The main objective of the current educational design-based research project was to investigate if and how communicative CBA programs could have a positive washback effect on teaching practices in Dutch lower form secondary FL education, thereby enhancing constructive alignment between the national CLT-oriented learning goals on the one hand and teaching and assessment activities on the other hand. More specifically, the aim of this research project was to implement communicative classroom-based assessment programs, developed by FL teachers themselves, as a steering tool to create a positive washback effect on teaching practices and ultimately on the way students learn. To ensure ecological validity of research outcomes (both in terms of gathered insights and developed materials), the research project was carried out in real school settings and in co-creation with practitioners. The whole project involved 52 FL teachers of English, French, German and Spanish, 26 language departments, 24 schools and about 1500 students.

The first phase of the research project consisted of a systematic inventory of the actual FL teaching and assessment practices in the research context (see Chapter 2). The main objective of this first study was to determine the degree of alignment between the current teaching and assessment practices on the one hand and the national CLT-oriented learning objectives on the other. To this end, classroom observations were conducted ( $N = 31$ ) in fifteen different schools, assessment materials of 10 language departments were analyzed and students ( $N = 473$ ) completed a digital survey. Teachers interviews ( $N = 21$ ) were then held to investigate which external (e.g., materials, school requirements) and conceptual factors (e.g., expertise, beliefs) had an influence on their pedagogical choices that could explain the observed degree of alignment.

Based on the results of the first study, the design phase of the project started (see Chapter 3). This second study aimed at identifying challenges and at providing clear theoretical and practical guidelines and tools for the development and implementation of a communicative CBA program. In a co-creation process that took place within a professional learning community consisting of 21 FL teachers from 15 Dutch secondary schools, we first determined which core components a communicative CBA program should include (see Figure 6.1) and formulated preliminary design principles for each of the components, focusing on both validity and feasibility. We then identified the challenges faced by the participants in the development and implementation (try-outs) of the different components of a communicative CBA program. Final design principles were formulated taking into account the challenges identified and solutions to overcome them (see Chapter 3, Table 3.2). Following these principles, we finally designed tools (formats and examples) that could be used by other secondary school FL teachers to develop and introduce a communicative CBA program at A2 level of the CEFR (the

most common level in the lower form). These tools were gathered in a toolbox available on a digital platform and in a teachers' guide (see Appendix B).



**Figure 6.1** Components of a constructively aligned communicative CBA program

In a third study, FL teachers ( $N = 32$ ) from 14 language departments of 10 schools used the tools designed in the previous research phase to develop and introduce their own communicative CBA program (see Chapter 4). In this implementation phase, the degree of treatment fidelity was first established by checking if the design principles defined as defined as the “key ingredients” of the intervention (Harn et al., 2013) were included in the different components of participants’ newly designed CBA program (see Chapter 3). Next came an investigation of how the implementation was experienced by the teachers and students involved through teacher logbooks, a student survey and interviews with students and teachers, in order to evaluate the implementation process and to formulate recommendations for future implementations.

Finally, in the fourth phase of the project, an evaluative study (see Chapter 5) was conducted in which the teacher activities and pedagogical choices of FL teachers teaching in both regular classes ( $N = 12$ ) and intervention classes ( $N = 12$ ) were compared. The regular classes applied a regular CBA program, largely based on assessments included in the textbooks used in the participating schools and similar to the CBA programs

analyzed in the previous inventory study (Chapter 2). In the intervention classes, on the other hand, a communicative CBA program was implemented. The aim of this study was to identify which observed and perceived washback effects the structural implementation of communicative CBA programs had on the pedagogical approach of the FL teachers involved. As part of a quasi-experiment *in situ*, the effects of the intervention on teaching practices were investigated through classroom observations in regular and intervention classes. Furthermore, the perceived effects of the intervention on teaching practices were assessed for both teachers and students through the qualitative analysis of teachers' logbooks and the quantitative analysis of student survey responses.

The main results of the various studies included in this dissertation are summarized and further discussed in this chapter.

## 6.2 Summary of the Main Findings

In the inventory study presented in Chapter 2, a lack of constructive alignment in the FL curricula of Dutch lower form secondary education was found, in particular between the national CLT-oriented objectives on the one hand, and the type of classroom-based assessments on the other. Observed classroom-based assessment activities focused mainly on knowledge of grammar rules and vocabulary out of context, and to some extent on reading skills, the only skill assessed at the national level at the end of secondary FL education. Only a few activities were aimed at developing productive skills and even fewer at developing speaking skills. Besides, assessment activities focusing on language skills were often similar to translation tasks and usually did not ask for an appropriate degree of free production at A2 level. In addition, only a few individual teachers or language departments used rating scales based on the CEFR can-do statements to evaluate students' performance on assessment tasks. These findings were echoed by the students surveyed, as they seldom reported the use of communicative activities to develop and assess their productive language skills.

The second part of the first study investigated which factors influence FL teachers' pedagogical choices and, as such, contributing to the observed lack of alignment. FL teachers in this study indicated that they lacked the time and expertise to develop their own materials, and that they therefore needed a textbook to save time and to guide their pedagogical choices. The results furthermore showed that teachers' pedagogical choices were sometimes linked to misconceptions about language learning, such as the idea that it is important to teach and assess knowledge of formal aspects of the language, such as specific grammar rules, separately and prior to the development of language skills (*focus-on-forms*; Long, 1991). This misconception has been identified in

other studies as an important factor of influence on pedagogical choices (Chang, 2011; Wang & Cheng, 2009). Whereas form-focused instruction can contribute to the development of language skills, it will only do so when provided in the context of meaningful communicative activities (Spada, 2011). In the decentralized Dutch educational system, where teachers have a considerable degree of autonomy in the selection, development and implementation of teaching and assessment activities, such misconceptions seem to be even more influential. The results of the inventory study furthermore showed that assessing the knowledge of formal language aspects outside a communicative context and the ability to translate sentences in highly guided tasks, instead of assessing students' ability to communicate in realistic situations, seemed to exert a negative washback effect on classroom activities that, as a result, focused mainly on language forms.

The results of the co-design study presented in Chapter 3 indicated that practical challenges during the design process were often related to the limited time allocated to develop, select, administer, and assess communicative assessment activities. The conceptual challenges identified essentially regarded a lack of expertise on how to operationalize specific CLT principles, such as unpredictability, focus-on-form and the integration of different language skills. Suggestions made to overcome these practical and conceptual challenges were taken into account in the formulation of the final design principles (see Table 3.2) and translated in the design of the tools aimed at helping FL teachers to develop and introduce a communicative CBA program suitable to their own practice (see Appendix B). The main suggestions for the development of a CBA program resulting from the collaboration project are summarized below:

1. Language skills should be assessed in an integrative way, combining receptive and productive skills, to reduce the number of formal assessments and to enhance authenticity and validity.
2. Assessment tasks should include new situations to ensure unpredictability, but should be based on well-known themes and speech acts to provide enough guidance to prepare students.
3. Rating scales should assess language skills in an integrated way. They should be descriptive, based on can-do statements belonging equally to each language skill. They should include a holistic and an analytical part to save time in scoring, to enable efficient feedback and ensure validity.
4. Informal assessments based on communicative activities should be systematically introduced to enhance students' motivation and to guide their learning process.

Chapter 4 reported on the implementation phase of the research project, in which the tools designed in the previous project phase (see Chapter 3) were used by 32 FL

teachers from 14 FL departments of 10 different schools to develop and introduce a communicative CBA program for their own context. First, the results showed an acceptable to high level of treatment fidelity with regard to the different “key ingredients” of the CBA programs for most language departments. The communicative character and the link to the CEFR were clearly included in the assessment materials and visible in the majority of the classroom observations.

Secondly, the investigation of how teachers involved perceived the implementation showed that the majority was positive about the intervention. They reported that the tools were inspiring and helped them develop their own CBA program. These results support the idea that an intervention with “built-in adaptations” (Harn et al., 2013, p.188) not only works as a source of inspiration but also offers flexibility, giving teachers the opportunity to adapt interventions to their own practice. The teachers involved furthermore stressed their enthusiasm about the new programs and noticed enthusiasm amongst their students as well. This latter aspect is certainly relevant, given the fact that pleasure and satisfaction stimulate motivation, which can be seen as the driving force behind FL learning (Ellis, 2019). Although teachers were generally positive about the implementation of their CBA programs, they also encountered some challenges. First of all, they reported a lack of experience with communicative assessment regarding CLT, among both teacher and students. They also mentioned the difficulty they had to keep track of individual students’ progress in large classes. Some of the participants also mentioned a lack of support from their school, which did not provide them with sufficient time to learn and experiment within the project. In the interviews, teachers also mentioned furthermore their lack of experience with the CEFR and the development of rating scales based on can-do statements in particular, resulting in extra workload when developing CBA. Finally, some of the participants expressed their concerns regarding a possible lack of “basic knowledge” of words and grammar rules since these formal aspects were not taught and assessed separately. Once again, these concerns reveal misconceptions regarding CLT and the introduction of a focus on form instead of on forms in the development of students’ communicative language skills (Ellis, 2015).

The majority of the students questioned were positive as well about the implementation of communicative CBA. They reported activities being more focused on the development of language skills and having more time to practice their language skills during the lessons, and thus being more confident to show they have attained the learning objectives in formal assessments. This is an important finding since self-confidence is known to contribute to the achievement of a higher level of language proficiency (MacIntyre et al., 1998).

Results of the final study on the observed and perceived effects of the structural implementation of a communicative CBA program on FL teaching practices, reported in Chapter 5, indicate that the implementation of such programs has a positive washback effect on teaching practices as they become more aligned with CLT learning goals and thus more communicative. Lessons observed in the intervention classes were all mainly focused on the development of language skills within a meaningful context, while in regular classes a large part of the lessons observed still focused mainly on forms (e.g., grammar rules out of context). Compared to the observed communicative teaching activities in regular classes, teaching activities in intervention classes adhered significantly more to communicative principles. These activities were more meaningful, authentic, and realistic, asking for more free production and taking a more integrated approach to teaching and learning language skills, matching real-life communication. Although there was no significant difference in the amount of target language use reported, teachers in intervention classes did tend to use the target language in a more conscious way to provide students with the input they need to improve their language skills. Finally, there was an increase in the use of informal formative assessment activities in intervention classes compared to regular classes. These activities were directly linked to CLT goals and assessment criteria. This is a positive aspect since using formative assessment activities aligned with learning goals gives students and teachers insights into students' learning progress and helps students to achieve learning objectives (Black & Wiliam, 2018).

Most of the effects identified through classroom observations were also perceived as such by the teachers and students involved in the final study. In the majority of the participating language departments, teachers reported more focus on the development of language skills and particularly on productive skills, as well as a focus on form within communicative activities instead of a focus on forms out of context. Some of them also indicated using the target language more frequently. Most students in the intervention classes reported learning activities, summative assessments and formative assessment activities being mainly focused on the development of language skills, while according to students from regular classes the focus in their FL education was mainly on the knowledge of formal aspects of the language.

The conclusion that can be drawn from these combined findings is that the implementation of a CBA program aligned with CLT learning goals can help change a more traditional focus-on-forms teaching approach into a more communicative one, with a primary focus on the development of language skills and on students' ability to communicate in the FL.

## 6.3 Limitations

This research project was carried out in the context of authentic classroom practice, which enhanced the ecological validity of its outcomes. This also implies, however, that many external factors may have influenced the results.

First of all, teachers could only be observed once or twice at a given time during the school year. Observations were performed during a specific phase of their teaching program and could not fully reflect the way they teach throughout the school year. For this reason, surveys and student interviews were conducted to triangulate the results of the classroom observations. However, it is important to note that only half of the students involved filled in the survey and that interviews were conducted with a very small sample of six students all coming from the same school. Although the data were insightful and in accordance with the results gathered from the other data sources, they do not guarantee that students from other schools would have experienced the implementation in the same way. Moreover, the predominantly positive experiences reported by students in interviews may have been influenced by socially desirable behavior due to the interview setting with the researcher.

The implementation phase of the project (see Chapter 3) took place largely during the COVID-19 pandemic, which had an undeniable impact on this phase of the project, as teachers had to adapt materials to online teaching and were forced to make pragmatic choices, such as reducing the number of planned formal assessments. More generally, the restrictive measures linked to the health crisis had an impact on the methodological choices. The unexpected first lockdown in the second half of the 2019-2020 school year prevented classroom observations prior to the implementation of the new CBA programs. Therefore, the original research plan of the final study had to be adapted by asking participating teachers to teach in at least two classes from the same grade in the same school year: an intervention class with the implementation of a communicative CBA program, and a regular class with a regular CBA program. Classroom observations in intervention and regular classes were therefore conducted in 2020-2021 during the same school year. As FL teachers generally do not have much time to prepare their lessons, it is possible that teaching materials and activities were used from one class to the other, since these were parallel classes. Due to a second lockdown during the 2020-2021 school year, between December 2020 and February 2021, some lessons were held online, and some schools worked with partial classes. In this particular context, not all classroom observations could be carried out, some participants were unable to provide all the data requested and others dropped out of the project during the school year. In addition, covid restrictions prevented classroom observations with two researchers, and the timeframe was too short to obtain approval from the ethics committee to make

video recordings of the lessons. The fact that results of the student surveys were consistent with those of the classroom observations, however, seems to confirm the validity of observation findings, and as such compensates somewhat for the lack of interrater reliability measures.

All these external factors, largely linked to the pandemic, had an unavoidable impact on the project. It is interesting to note, however, that despite these special circumstances, the large majority of the participating teachers followed the project through to the end and were enthusiastic about it.

## 6.4 Suggestions for Educational Practice

The results of this research project have shown that the washback-potential of assessment on teaching practices applies not only to high stakes external examinations, but also to classroom-based assessments. To obtain curriculum alignment and hence effective FL teaching and learning, it is essential to ensure that FL assessment is aligned with communicative learning objectives. In the Netherlands, it is therefore important that CBA programs are reconsidered and developed more in line with the national communicatively oriented learning outcomes to be achieved.

To compensate for a lack of time and expertise, particularly regarding curriculum design and CBA, FL teachers, like those involved in this research project, often rely on ready-made teaching and assessment materials included in textbooks. Unfortunately, these materials do not always put enough emphasis on communicative activities, and often include assessments that mainly and directly assess knowledge of formal aspects of the language outside a communicative context. Ideally, educational publishers should include more communicative teaching and assessment activities in FL textbooks. In any case, FL teacher education programs should explicitly train prospective teachers to use textbooks selectively and focus on the communicative activities they contain, to select CLT teaching and assessment materials from other sources, and/or to develop these themselves.

Besides, it is important that FL teachers are aware of the different interpretations of CLT, which teaching practices are related to these interpretations and which ones are most suitable to teach a second, third or even fourth language in the context of secondary education. Therefore, methodology courses in pre-service and in-service language teacher training programs should focus more on how to introduce CLT in large classes with few teaching hours, on how to give appropriate form-focused instruction and feedback to students within communicative situations or tasks, and finally on how to

develop and introduce an appropriate communicative CBA program. More generally, in daily teaching practice, teachers should be allocated more time for curriculum design and for the selection, development and introduction of communicative assessment activities.

Currently, secondary education attainment targets for FL education in the Netherlands are being revised. Teachers, teacher educators and researchers are invited to take part in this process. To ensure that the newly formulated learning objectives will subsequently be translated into teaching practices properly, it is essential that they are precisely formulated and accompanied with clear guidelines on how to translate them into teaching as well as assessment programs. The results of this research project could also be taken into account in the redevelopment of national examination programs. They should emphasize the assessment of language skills in an integrative way and give prominence to productive skills, and to speaking skills in particular. This could, in the light of the results obtained, create a positive washback on classroom practices and, as such, place oral communication at the center of foreign language teaching. This is important as speaking is a fundamental skill; being able to speak an FL not only facilitates personal and professional success through day-to-day interactions but also contributes to cross-cultural interactions and a better understanding of other perspectives.

## 6.5 General Conclusion

This dissertation is the product of a practitioner research project that was facilitated by Utrecht University of Applied Sciences where I work as a FL teacher educator. This PhD grant gave me the opportunity to explore and investigate a problem I faced in my daily educational practice, and to generate impact while conducting research. I chose a problem that I encountered as a secondary school FL teacher many years ago, which my students still face, and which is a recurring topic of discussions in my methodology classes. Although the present research project is not directly linked to my own teaching in higher education, the fact that I teach future teachers made that the project also had concrete, short-term effects on my own teaching practice. In the methodology courses we teach, for example, more attention is now given to curriculum design, constructive alignment, communicative classroom-based assessment, and formative evaluation. Of course, I still have discussions about difficulties that come with CLT implementation, but I am confident that I can rely on my research findings to provide my students with evidence-based tools and insights, as well as with concrete examples of good practices that I gathered during the research process. Moreover, topics pertaining to the question of how to introduce CLT in practice are now part of the training that me and my colleagues provide to FL teachers supervising our students in our partner schools. I have

also had the opportunity to present the insights and tools of my study not only to other researchers, but also to many FL teachers, language departments, teacher educators, publishers, testing specialists and curriculum developers. Their feedback and discussions have improved both my research as well as their practices. Conducting research in practice has given me the opportunity to see many different school environments and practices. I am grateful to all the participating teachers who opened the doors of their classrooms and shared their doubts and questions openly. All classroom observations, working sessions and interviews with teachers I got to perform during the project, proved to be more than just valuable instruments for data collection; they were inspiring encounters that gave me the opportunity to develop professionally and personally.

The overall findings as presented in this dissertation indicate that the implementation of a communicative CBA program can generate a positive washback effect on teaching practices and enhance constructive alignment in FL curricula. The studies in this dissertation have demonstrated that, with the right tools providing enough flexibility, FL teachers can implement a communicative CBA program that positively changes the way they teach. Results showed that these positive effects relate to both the selection of communicative teaching and assessment activities as well as to the way teachers introduce them in practice. Furthermore, results showed that both teachers and students were positive about the implementation of a communicative CBA program and the effect they perceived on teaching and learning activities.

# Bibliography

- Ahmed, M. (2016). Communicative language teaching: A practical scenario in the context of Bangladesh. *Advances in Language and Literary Studies*, 7(5), 97-104. <https://doi.org/10.7575/aiall.v.7n.5p.97>
- Al Hinai, M. K., & Al Jardani, K. S. (2021). Washback in language testing: An exploration with a focus on a specific EFL context in Oman. *International Journal of English Linguistics*, 11(1).
- Alderson, C., & Wall, D. (1993). Does washback exist? *Applied Linguistics*, 14(2), 115-129. <https://doi.org/10.1093/applin/14.2.115>
- Allal, L. (2016). The co-regulation of student learning in an assessment for learning culture. In D. Laveault & L. Allal (Eds.), *Assessment for learning: Meeting the challenge of implementation* (pp. 259–273). Springer International Publishing.
- Allen, D. (2016). Investigating washback to the learner from the IELTS test in the Japanese tertiary context. *Language Testing in Asia*, 6(7), 1-20. <https://doi.org/10.1186/s40468-016-0030-z>
- ALTE (2011). *Manual for language test development and examining – For use with the CEFR*, Language Policy Division, Council of Europe. <https://rm.coe.int/1680667a2b>
- Anderson, J. (2017). A potted history of PPP with the help of *ELT Journal*. *ELT journal*, 71(2), 218-227. <https://doi.org/10.1093/elt/ccw055>
- Bachman, L. F. (2002). Some reflections on task-based language performance assessment. *Language Testing*, 19(4), 453-476.
- Bachman, L. F. (2010). Second language assessment. *International Encyclopedia of Education*, 4, 140-147.
- Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice: Designing and developing useful language tests* (Vol. 1). Oxford University Press.
- Bakkenes, I., Vermunt, J. D., & Wubbels, T. (2010). Teacher learning in the context of educational innovation: Learning activities and learning outcomes of experienced teachers. *Learning and instruction*, 20(6), 533-548.
- Barnes, M. (2017). Washback: Exploring what constitutes “good” teaching practices. *Journal of English for Academic Purposes*, 30, 1-12.
- Berry, V., Sheehan, S., & Munro, S. (2019). What does language assessment literacy mean to teachers? *ELT Journal*, 73(2), 113–123. <https://doi.org/10.1093/elt/ccy055>
- Biggs, J. (1996). Enhancing teaching through constructive alignment. *Higher Education*, 32(3), 347-364. <https://doi.org/10.1007/BF00138871>
- Biggs, J. B. & Tang, C. (2007). *Teaching for quality learning at university*. 3rd Edn., McGraw Hill Education & Open University Press.
- Binkhorst, F., Poortman, C. L., & Van Joolingen, W. R. (2017). A qualitative analysis of teacher design teams: In-depth insights into their process and links with their outcomes. *Studies in Educational Evaluation*, 55, 135-144. <https://doi.org/10.1016/j.stueduc.2017.10.001>
- Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education*, 5(1), 7-74. <http://dx.doi.org/10.1080/0969595980050102>
- Black, P., & Wiliam, D. (2018). Classroom assessment and pedagogy. *Assessment in Education: Principles, policy & practice*, 25(6), 551-575. <https://doi.org/10.1080/0969594X.2018.1441807>
- Broadfoot, P., & Black, P. (2004). Redefining assessment of L2 development? The first ten years of assessment in education. *Assessment in Education*, 1(1), 7–27.
- Brown, J. D. (2005). *Testing in language programs: A comprehensive guide to English language assessment*. McGraw-Hill.

- Burke, B. M. (2011). Rituals and beliefs ingrained in world language pedagogy: defining deep structure and conventional wisdom. *Journal of Language Teaching and Research*, 2(1), 1. <https://doi.org/10.4304/jltr.2.1.1-12>
- Burner, T. (2023). A review of formative language assessment research and implications for practitioners. In S.W. Chong & H. Reinders (Eds.), *Innovation in Learning-Oriented Language Assessment. New Language Learning and Teaching Environments*. Palgrave Macmillan. [https://doi.org/10.1007/978-3-031-18950-0\\_2](https://doi.org/10.1007/978-3-031-18950-0_2)
- Celce-Murcia, M. (2007). Rethinking the role of communicative competence in language teaching. *Intercultural language use and language learning*, 41-57.
- Chan, K. L. R. (2020). Washback in education: A critical review and its implications for language teachers. *Journal of Foreign Language Education and Technology*, 5(1), 108-124.
- Chang, M. (2011). EFL teachers' attitudes toward communicative language teaching in Taiwanese college. *Asian EFL Journal*, 53(1), 17-34. <https://doi.org/10.5539/elt.v4n2p3>
- Chapman, C., & Muijs, D. (2014). Does school-to-school collaboration promote school improvement? A study of the impact of school federations on student outcomes. *School effectiveness and school improvement*, 25(3), 351-393.
- Chen, Q., Zhang, J., & Li, L. (2021). Problematizing formative assessment in an undeveloped region of China: voices from practitioners. *Educational Assessment, Evaluation and Accountability*, 33, 649-673.
- Cheng, L. (2005). Changing language teaching through language testing: A washback study (Vol. 21). Cambridge university press.
- College voor Toetsen en Examens. (2019). *Examenprogramma moderne vreemde talen vwo vanaf CE 2019* [Exam program foreign languages in pre-academic stream]. [https://www.examenblad.nl/examenstof/syllabus-2019-moderne-vreemde-5/2019/vwo/f=/mvt\\_versie\\_4\\_vwo\\_2019\\_nader\\_vastgesteld.pdf](https://www.examenblad.nl/examenstof/syllabus-2019-moderne-vreemde-5/2019/vwo/f=/mvt_versie_4_vwo_2019_nader_vastgesteld.pdf)
- Commissie kwaliteit schoolexaminering. (2019). Schoolexaminering in het voortgezet onderwijs: naar een zorgvuldig verloop en een betere kwaliteitsbewaking. Inspectie van het onderwijs. <https://www.onderwijsinspectie.nl/binaries/onderwijsinspectie/documenten/rapporten/2020/01/31/schoolexamineringinhetvoortgezetonderwijs/Rapport+Schoolexaminering+in+het+voortgezet+onderwijs.pdf>
- Corbin, J. M., & Strauss, A. (1990). Grounded theory research: Procedures, canons, and evaluative Criteria. *Qualitative Sociology*, 13(1), 3-21. <https://doi.org/10.1007/BF00988593>.
- Corbin, J. & Strauss, A. (2008). Strategies for qualitative data analysis. In *Basics of qualitative research (3rd ed.): Techniques and procedures for developing grounded theory*, 65-86. SAGE Publications, Inc. <https://dx.doi.org/10.4135/9781452230153.n4>
- Coskun, A. (2011). Investigation of the application of communicative language teaching in the English language classroom: A case study on teachers' attitudes in Turkey. *Journal of Linguistics and Language Teaching*, 2(1) 85-109.
- Council of Europe (2011). Manual for language test development and examining. *For use with the CEFR*. Council of Europe.
- Council of Europe (2020). *Common European Framework of Reference for Languages: Learning, teaching, assessment. Companion volume*. Council of Europe. <https://www.coe.int/en/web/common-european-framework-reference-languages>
- Criado, R. (2013). A critical review of the presentation-practice-production model (PPP) in foreign language teaching. In R. Monroy (Ed.), *Homenaje a Francisco Gutiérrez Díez* (pp. 97-115). Edit.um. ISBN: 978-84-15463-55-9

- Cumming, A., Kantor, R., Baba, K., Eouanzoui, K., Erdosy, U., & James, M. (2006). *Analysis of discourse features and verification of scoring levels for independent and integrated prototype written tasks for the new TOEFL* (TOEFL Monograph Series MS30). Princeton, NJ: Educational Testing Service.
- Curriculum.nu. (2019). *Leergebied Engels/ MVT: Voorstel voor de basis van de herziening van de kern-doelen en eindtermen van de leraren en schoolleiders uit het ontwikkelteam Engels/moderne vreemde talen*. [English/modern foreign languages curriculum: Proposal for the revision of the core learning objectives and attainment targets]. Available at: <https://www.curriculum.nu/voorstellen/engels-mvt/>
- Davison, C. (2019). Using assessment to enhance learning in English language education. *Second Handbook of English Language Teaching*, 433-454.
- Davies, A. (1999). Dictionary of language testing (Vol. 7). *Cambridge University Press*.
- De Neve, D., Leroy, A., Struyven, K., & Smits, T. (2022). Supporting formative assessment in the second language classroom: An action research study in secondary education. *Educational Action Research*, 30(5), 828-849.
- Dixon, D. D., & Worrell, F. C. (2016). Formative and summative assessment in the classroom. *Theory into practice*, 55(2), 153-159. <https://doi.org/10.1080/00405841.2016.1148989>
- Dönszelmann, S., Kaal, A., De Graaff, R., & Beishuizen, J. (2019). Doeltaal-Leertaal; De invloed van een concrete doeltaal didactiek op vaardigheidsprestaties in de reguliere mvt-les. *Levende Talen Tijdschrift*, 20(4), 3-13.
- Dos Santos, L. M. (2020). The discussion of communicative language teaching approach in language classrooms. *Journal of Education and e-Learning Research*, 7(2), 104-109.
- Durlak, J. A. (1998). Why program implementation is important? *Journal of Prevention & Intervention in the Community*, 17(2), 5-18. [https://doi.org/10.1300/J005v17n02\\_02](https://doi.org/10.1300/J005v17n02_02)
- Ellis, R. (2010). Second language acquisition, teacher education and language pedagogy. *Language Teaching*, 43(2), 182-201.
- Ellis, R. (2015). The importance of focus on form in communicative language teaching. *Eurasian Journal of Applied Linguistics*, 1(2), 1-12.
- Ellis, N. C. (2019). Essentials of a theory of language cognition. *The Modern Language Journal*, 103(1), 39-60. <https://doi.org/10.1111/modl.12532>
- Fasoglio, D., De Jong, K., Pennewaard, D., Trimbos, B. & Tuin, D. (2015). *Moderne vreemde talen: Vakspecifieke trendanalyse 2015*. [Modern foreign languages: Subject-specific trend analysis 2015]. Stichting Leerplanontwikkeling. <https://slo.nl/@4300/moderne-vreemde/>
- Firoozi, T., Razavipour, K., & Ahmadi, A. (2019). The language assessment literacy needs of Iranian EFL teachers with a focus on reformed assessment policies. *Language Testing in Asia*, 9(2). <https://doi.org/10.1186/s40468-019-0078-7>
- Fischer, J. (2020). The underlying action-oriented and task-based approach of the CEFR and its implementation in language testing and assessment at university. *Language Learning in Higher Education*, 10(2), 301-316. <https://doi.org/10.1515/cercles-2020-2021>
- Fox, R., Corretjer, O., & Webb, K. (2019). Benefits of foreign language learning and bilingualism: An analysis of published empirical research 2012–2019. *Foreign Language Annals*, 52(4), 699-726.
- Fulcher, G. (2012). Assessment literacy for the language classroom. *Language Assessment Quarterly*, 9(2), 113-132. <https://doi.org/10.1080/15434303.2011.642041>
- Gan, L. & Lam, R. (2022). A review on language assessment literacy: Trends, foci, and contributions, *Language Assessment Quarterly*, 19(5), 503-525. <https://doi.org/10.1080/15434303.2022.128802>

- Gebril, A. (2010). Bringing reading-to-write and writing-only assessment tasks together: A generalizability analysis. *Assessing Writing*, 15(2), 100-117.
- Genesee, F. (2004). What do we know about bilingual education for majority language students? In T. K. Bhatia and W. Ritchie (Eds.), *Handbook of bilingualism and multiculturalism* (pp. 547-576). Blackwell.
- Grabe, W., & Zhang, C. (2016). Reading-writing relationships in first and second language academic literacy development. *Language Teaching*, 49(3), 339-355.
- Graus, J., & Coppen, P. A. (2016). Student teacher beliefs on grammar instruction. *Language Teaching Research*, 20(5), 571-599. <http://dx.doi.org/10.1177/1362168815603237>
- Graus, J., & Coppen, P. A. (2017). The interface between student teacher grammar cognitions and learner-oriented cognitions. *The Modern Language Journal*, 101(4), 643-668. <https://doi.org/10.1111/modl.12427>
- Green, A. (2007). Washback to learning outcomes: A comparative study of IELTS preparation and university pre-sessional language courses. *Assessment in Education*, 14(1), 75-97. <https://doi.org/10.1080/09695940701272880>
- Green, A. (2013). Washback in language assessment. *International Journal of English Studies*, 13(2), 39-51. <https://doi.org/10.6018/ijes.13.2.185891>
- Guerrettaz, A. M., & Johnston, B. (2013). Materials in the classroom ecology. *Modern Language Journal*, 97, 779-796.
- Hakim, L. N., & Tasikmalaya, U. P. (2018). Washback effect in language testing: *What do we know and what is its effect?* *Jurnal Forum Didaktik* 2(1), 59-68.
- Harding, L. (2014). Communicative language testing: Current issues and future research. *Language Assessment Quarterly*, 11(2), 186-197.
- Harding, L., & Brunfaut, T. (2020). Trajectories of language assessment literacy in a teacher researcher partnership: Locating elements of praxis through narrative inquiry. In M. E. Poehner & O. Inbar-Lourie (Eds.), *Towards a re-conceptualization of second language classroom assessment* (pp. 61-81). Springer International Publishing. <https://doi.org/10.1007/978-3-030-35081-9>
- Harn, B., Parisi, D., & Stoolmiller, M. (2013). Balancing fidelity with flexibility and fit: What do we really know about fidelity of implementation in schools? *Exceptional Children*, 79(2), 181-193. <https://doi.org/10.1177/0014402913079002051>
- Harsch, C., & Martin, G. (2013). Comparing holistic and analytic scoring methods: Issues of validity and reliability. *Assessment in Education: Principles, Policy & Practice*, 20(3), 281-307.
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81-112.
- Hidri, S. (2016). Conceptions of assessment: Investigating what assessment means to secondary and university teachers. *Arab Journal of Applied Linguistics*, 1(1), 19-43. <https://doi.org/10.1234/ajal.v1i1.313>
- Hill, K., & McNamara, T. (2012). Developing a comprehensive, empirically based research framework for classroom-based assessment. *Language Testing*, 29(3), 395-420. <https://doi.org/10.1177/02655322114283>
- Hill, K. (2017). Understanding classroom-based assessment practices: A precondition for teacher assessment literacy. *Papers in Language Testing and Assessment*, 6(1), 1-17.
- Howatt, A. (1984). *A history of English language teaching*. Oxford University Press.
- Hymes, D. (1972). On communicative competence. In J. B. Pride & J. B. Holmes (Eds.), *Sociolinguistics: Selected readings* (pp. 269-293). Penguin.

- Isaacs, T. (2016). Assessing speaking. In D. Tsagari & J. Banerjee (Eds.), *Handbook of second language assessment* (pp. 131–146). De Gruyter Mouton.
- Jiang, Y., & Wang, J. (2018). A study of cultural empathy in foreign language teaching from the perspective of cross-cultural communication. *Theory and Practice in Language Studies*, 8(12), 1664–1670.
- Jin, Y. (2010). The place of language testing and assessment in the professional preparation of foreign language teachers in China. *Language Testing*, 27(4), 555–584.
- Kissau, S., Rodgers, M. & Haudeck, H. (2015). Practicing what they Preach? A Comparison of teacher candidate beliefs and practices. *International Journal of Language Studies*, 9(4), 29–54.
- Krashen, S. (1984). *The input hypothesis: Issues and implications*. Longman.
- Krijgsman, C., Borghouts, L., Mainhard, T., Van Tartwijk, J., & Haerens, L. (2023). Toetsing en motivatie. Synergie en balans in formatief handelen en summatief toetsen beïnvloeden vanuit een systemisch perspectief. *Pedagogische Studiën*, 100(3). <https://doi.org/10.59302/ps.v100i3.17635>
- Kroll, J. F., & Dussias, P. E. (2017). The benefits of multilingualism to the personal and professional development of residents of the US. *Foreign Language Annals*, 50(2), 248–259.
- Kuang, Q. (2020). A review of the washback of English language tests on classroom teaching. *English Language Teaching*, 13(9), 10–17.
- Lam, R. (2019). Teacher assessment literacy: Surveying knowledge, conceptions, and practices of classroom-based writing assessment in Hong Kong. *System*, 81, 78–89. <https://doi.org/10.1016/j.system.2019.01.006>
- Larsen-Freeman, D. & Anderson, M. (2011). *Techniques and principles in language teaching*. Oxford University Press.
- Lee, J., Butler, Y. G., & Peng, X. (2021). Multiple stakeholder interaction to enhance preservice teachers' language assessment literacy. *Languages*, 6(4), 213. <https://doi.org/10.3390/languages6040213>
- Lee, Y. W. (2006). Dependability of scores for a new ESL speaking assessment consisting of integrated and independent tasks. *Language Testing*, 23(2), 131–166.
- Lendrum, A., & Humphrey, N. (2012). The importance of studying the implementation of interventions in school settings. *Oxford Review of Education*, 38(5), 635–652. <https://doi.org/10.1080/03054985.2012.734800>
- Leung, C., Davison, C., East, M., Evans, M., Liu, Y., Hamp-Lyons, L., & Purpura, J. (2018). Using assessment to promote learning: Clarifying construct theories and practices. In J. McE. Davis, J. M. Norris, M. E. Malone, T. H. McKay, & Y. A. Son (Eds.), *Useful assessment and evaluation in language education* (pp. 75–94). Georgetown University Press.
- Levin, B. (2004). Making research matter more. *Education Policy Analysis Archives*, 12(56), 1–22. <https://doi.org/10.14507/epaa.v12n56.2004>
- Lewkowicz, J., & Leung, C. (2021). Classroom-based assessment. *Language Teaching*, 54(1), 47–57. <https://doi.org/10.1017/S0261444820000506>
- Lightbown, P.M., Halter, R., White, J.L., & Horst, M. (2002). Comprehension-based learning: The limits of “do it yourself.” *Canadian Modern Language Review*, 58, 427–464.
- Lightbown, P. M., & Spada, N. (2013). *How languages are learned* (4th ed.). Oxford University Press.
- Little, D. (2007). The Common European framework of reference for languages: Perspectives on the making of supranational language education policy. *The Modern Language Journal*, 91(4), 645–655. [https://doi-org.proxy.library.uu.nl/10.1111/j.1540-4781.2007.00627\\_2.x](https://doi-org.proxy.library.uu.nl/10.1111/j.1540-4781.2007.00627_2.x)
- Littlewood, W., and Yu, B. (2011). First language and target Language in the foreign language classroom. *Language Teaching*, 44(1), 64–77. <https://doi.org/10.1017/S0261444809990310>

- Long, M. H. (1991). Focus on form: A design feature in language teaching methodology. In K. de Bot, R. Ginsberg, & C. Kramsch (Eds.), *Foreign language research in cross-cultural perspective* (pp. 39–52). John Benjamins.
- Long, M. (2009). Methodological principles for language teaching. In M. Long and C. Doughty (Eds.), *The Handbook of language teaching* (pp. 373-394). Wiley-Blackwell.
- Lyster, R. (2007). *Learning and teaching languages through content: A counterbalanced approach*. John Benjamins.
- MacIntyre, P. D., Clément, R., Dörnyei, Z., & Noels, K. A. (1998). Conceptualizing willingness to communicate in a L2: A situational model of L2 confidence and affiliation. *The Modern Language Journal*, 82(4), 545-562. <https://doi.org/10.1111/j.1540-4781.1998.tb05543.x>
- Maulana, R., Helms-Lorenz, M., & Van de Grift, W. (2017). Validating a model of effective teaching behaviour of pre-service teachers. *Teachers and Teaching*, 23(4), 471–493, <https://doi.org/10.1080/13540602.2016.1211102>
- McMillan, J. H. (Ed.). (2013). *SAGE handbook of research on classroom assessment*. Sage.
- Meijer, D., & Fasoglio, D. (2007). *Handreiking schoolexamen moderne vreemde talen havo/vwo: Duits, Engels, Frans*. [Guidelines for the school exams for modern foreign languages in secondary education: German, English, and French]. SLO.
- Michel, M., Vidon, C., De Graaff, R. & Lowie, W. (2021). Language learning beyond English in the Netherlands: A fragile future? *European Journal of Applied Linguistics*, 9(1), 159-182. <https://doi-org.proxy.library.uu.nl/10.1515/eujal-2020-0020>
- Moonen, M., Stoutjesdijk, E., De Graaff, R. & Corda, A. (2013). Implementing CEFR in secondary education: Impact on FL teachers' educational and assessment practice. *International Journal of Applied Linguistics*, 23(2), 226-246. <https://doi.org/10.1111/ijal.12000>
- Morgan, C. (1996). The teacher as examiner: the case of mathematics coursework, *Assessment in Education*, 3(3), 353–376.
- Morrow, C. K. (2018). Communicative language testing. In J.I. Lontos (Ed.), *The TESOL Encyclopedia of English Language Teaching*, (pp. 342–350). Wiley.
- Muñoz, A. P., & Álvarez, M. E. (2010). Washback of an oral assessment system in the EFL classroom. *Language Testing*, 27(1), 33-49. <https://doi.org/10.1177/0265532209347148>
- Muranoi, H. (2007). Output practice in the L2 classroom. In R. M. DeKeyser (Ed.), *Practice in a second language. Perspectives from applied linguistics and cognitive psychology* (pp. 51-84). Cambridge University Press.
- Naghdipour, B. (2017). Incorporating formative assessment in Iranian EFL writing: A case study. *The Curriculum Journal*, 28(2), 283-299.
- National Institute for Curriculum Development SLO (2015). Curriculumspiegel Deel A: Generieke trendanalyse. Stichting Leerplanontwikkeling. <file:///C:/Users/charline.rouffet/Downloads/curriculumspiegel-2015-deel-a.pdf>
- National Institute for Curriculum Development (SLO). (2013, July 20). Criteria voor de kwaliteit van een toets. *Checklist kwaliteit toetsen*. SLO. <https://www.slo.nl/handreikingen/havo-vwo/shared/kwaliteit-schoolexamen/criteria-kwaliteit/>
- Nishino, T. (2011). Japanese high school teachers' beliefs and practices regarding communicative language teaching. *JALTJournal*, 33(2), 131-155.
- North, B., Figueras, N., Takala, S., Van Avermaet, P., & Verhelst, N. (2009). *Relating language examinations to the Common European Framework of Reference for Languages: Learning, teaching, assessment (CEFR). A manual*. Council of Europe. <https://www.coe.int/en/web/language-policy/home>

- North, B. (2016). Developing CEFR illustrative descriptors of aspects of mediation. *International Online Journal of Education and Teaching (IOJET)*, 3(2), 132-140.
- Pan, Y. C. (2009). Review of washback and its pedagogical implications. *VNU Journal of Foreign Studies*, 25 (4), 257-263.
- Pan, Y. C. (2020). Taiwan university students' perceptions of summative and formative classroom assessment in English courses. *TESOL International Journal*, 15(2), 46-64.
- Perrone, M. (2011). The effect of classroom-based assessment and language processing on the second language acquisition of EFL students. *Journal of Adult Education*, 40(1), 20-33.
- Plakans, L. (2021). Integrated Skills Assessment. In H. Mohebbi, C. Coombe (Eds), *Research questions in language education and applied linguistics* (pp 329-332). Springer Texts in Education. [https://doi.org/10.1007/978-3-030-79143-8\\_58](https://doi.org/10.1007/978-3-030-79143-8_58)
- Plomp, T., & Nieveen, N. (2009). *An introduction to educational design research*. National Institute for Curriculum Development (SLO). [https://research.utwente.nl/files/14472302/Introduction\\_20to\\_20education\\_20design\\_20research.pdf](https://research.utwente.nl/files/14472302/Introduction_20to_20education_20design_20research.pdf)
- Poehner, M. E. (2009). Dynamic assessment as a dialectical framework for classroom activity: Evidence from second language (L2) learners. *Journal of Cognitive Education and Psychology*, 8(3), 252-268.
- Preece, P. F. W. & Skinner, M. C. (1999) The national assessment in science at Key Stage 3 in England and Wales and its impact on teaching and learning, *Assessment in Education*, 6(1), 11-26.
- Purpura, J. E. (2008). The impact of large-scale and classroom-based language assessments on the individual. In *Language testing matters: Investigating the wider social and educational impact of assessment—Proceedings of the ALTE Cambridge Conference* (pp. 301-325).
- Rea-Dickins, P. (2001). Mirror, mirror on the wall: Identifying processes of classroom assessment. *Language Testing*, 18(4), 429-462.
- Rea-Dickins, P., & Scott, C. (2007). Washback from language tests on teaching, learning and policy: Evidence from diverse settings. *Assessment in Education*, 10(2), 1-7.
- Reeves, T. C. (2006). Design research from a technology perspective. *Educational Design Research*, 1(3), 52-66.
- Richards, J. C. (2006). *Communicative language teaching today*. Cambridge University Press.
- Robinson, P. (2001). Individual differences, cognitive abilities, aptitude complexes and learning conditions in second language acquisition. *Second Language Research*, 17(4), 368-392.
- Rouffet, C., Van Beuningen, C., & De Graaff, R. (2021). Designing a communicative foreign language assessment program for Dutch secondary schools: A design-based research project within a professional learning community. In *Collated Papers for the ALTE 7th International conference, Madrid* (pp. 179-183)
- Rouffet, C (2023). *Handleiding communicatief toetsen en formatief handelen*. Hogeschool Utrecht. <https://www.hu.nl/onderzoek/publicaties/handleiding-communicatief-toetsen-en-formatief-handelen>
- Rouffet, C., Van Beuningen, C., & De Graaff, R. (2023). Constructive alignment in foreign language curricula: An exploration of teaching and assessment practices in Dutch secondary education, *The Language Learning Journal*, 51(3), 344-358. <https://doi.org/10.1080/09571736.2022.2025542>
- RUG (Rijksuniversiteit Groningen). (2019). ICALT3: *International comparative analysis of learning and teaching*. Rijksuniversiteit Groningen. <https://www.rug.nl/gmw/lerarenopleiding/onderzoek/psychometrisch>
- Rukthong, A., & Brunfaut, T. (2020). Is anybody listening? The nature of second language listening in integrated listening-to-summarize tasks. *Language Testing*, 37(1), 31-53.

- Saglam, A. L. G., & Farhady, H. (2019). Can exams change how and what learners learn? Investigating the washback effect of a university English language proficiency test in the Turkish context. *Advances in Language and Literary Studies*, 10(1), 177-185. <https://doi.org/10.7575/aialc.all.v.10n.1p.177>
- Saito, H., & Inoi, S. I. (2017). Junior and senior high school EFL teachers' use of formative assessment: A mixed-methods study. *Language Assessment Quarterly*, 14(3), 213-233.
- Sarab, A. M. R., Monfared, A., & Safarzadeh, M. M. (2016). Secondary EFL school teachers' perceptions of CLT principles and practices: An exploratory survey. *Iranian Journal of Language Teaching Research*, 4(3), 109-130.
- Savignon, S. J. (2017). Communicative competence. In J.I. Liantas (Ed.) *The TESOL Encyclopedia of English Language Teaching* (pp. 1-7). John Wiley. <https://doi.org/10.1002/9781118784235.eelt0047>
- Scheerens, J. (2016). Recapitalization, implications for educational policy and practice and future research. In J. Scheerens (Ed.) *Opportunity to learn, curriculum alignment and test Preparation: A research review* (pp. 121-135). SpringerBriefs in Education, Springer. <https://doi.org/10.1007/978-3-319-43110-9>
- Schnabel, P., Ten Dam, G., Douma, T., Van Eijk, R., Tabarki, F., Van der Touw, A. & Visser, M. (2016). *Ons Onderwijs 2032: Eindadvies*. Platform for Education. <https://dspace.library.uu.nl/handle/1874/346309>
- Shen, C. (2002). Revisiting the relationship between students' achievement and their self-perceptions: A cross-national analysis based on TIMSS 1999 data. *Assessment in Education*, 9(2), 161-184.
- Shepard, L. A., Penuel, W. R., & Pellegrino, J. W. (2018). Using learning and motivation theories to coherently link formative assessment, grading practices, and large-scale assessment. *Educational Measurement: Issues and Practice*, 37(1), 21-34.
- Spada, N., & Tomita, Y. (2010). Interactions between type of instruction and type of language feature: A meta-analysis. *Language Learning*, 60, 1-46.
- Spada, N. (2011). Beyond form-focused instruction: Reflections on past, present, and future research. *Language Teaching*, 44(2), 225-236. <https://doi.org/10.1017/S0261444810000224>
- Spada, N. (2014). Instructed second language acquisition research and its relevance for L2 teacher education. *Education Matters: The Journal of Teaching and Learning*, 2(1).
- Spada, N., & Lightbown, P. M. (2022). In it together: Teachers, researchers, and classroom SLA. *The Modern Language Journal*, 106(3), 635-650. <https://doi.org/10.1111/modl.12792>
- Spratt, M. (2005). Washback and the classroom: The implications for teaching and learning of studies of washback from exams. *Language Teaching Research*, 9(1), 5-29.
- Stecher, B. M., Barron, S. L., Chun, T., & Ross, K. (2000). *The effects of the Washington State education reform on schools and classrooms* (CSE Technical Report No. 525). National Center for Research on Evaluation, Standards, and Student Testing, University of California.
- Stoll, L., Bolam, R., McMahon, A., Thomas, S., Wallace, M., Greenwood, A. & Hawkey, K. (2006). *Professional learning communities: Source materials for school leaders and other leaders of professional learning*. Innovation Unit, DfES, NCSL and GTC.
- Stoyhoff, S. (2012). Looking backward and forward at classroom-based language assessment. *ELT journal*, 66(4), 523-532.
- Sultana, N. (2018). Investigating the relationship between washback and curriculum alignment: A literature review. *Canadian Journal for New Scholars in Education*, 9(2), 151-158.
- Swain, M. (2005). The output hypothesis: Theory and research. In E. Hinkel (ed.), *Handbook of research in second language teaching and learning* (pp. 471-483). Routledge.

- Tang, S. Y., Wong, A. K., & Cheng, M. M. (2012). Professional learning in initial teacher education: Vision in the constructivist conception of teaching and learning. *Journal of Education for Teaching*, 38(4), 435-451.
- Taylor, L. (2005). Washback and impact. *ELT Journal*, 59(2), 154-155.
- Toksöz, I., & Kılıçkaya, F. (2017). Review of journal articles on washback in language testing in Turkey. *Lublin Studies in Modern Languages and Literature*, 41(2), 184-204. <http://dx.doi.org/10.17951/lsml.2017.41.2.184>
- Tsagari, D., & Vogt, K. (2017). Assessment literacy of foreign language teachers around Europe: Research, challenges, and future prospects. *Papers in Language Testing and Assessment*, 6(1), 41-63.
- Tsagari, D. (2020). Language assessment literacy: Concepts, challenges, and prospects. In S. Hidri (Ed.), *Perspectives on language assessment literacy: Challenges for improved student learning* (pp.13–32). Routledge. <https://doi.org/10.4324/9781003016083>
- Turner, C. E., & Purpura, J. E. (2016). 16. Learning-oriented assessment in second and foreign language classrooms. In D. Tsagari & J. Baneerjee (Eds.), *Handbook of second language assessment* (pp. 255-274). De Gruyter Mouton.
- Van Batenburg, E. S., Oostdam, R. J., Van Gelderen, A. J., Fukkink, R. G., & De Jong, N. H. (2020). Evaluating opportunities in Dutch EFL course books for developing pre-vocational learners' oral interactional ability. *Language Teaching Research*, 24(4), 434-455.
- Van den Akker, J. (1999). Principles and methods of development research. In J. J. H. v. d. Akker, R. Branch, K. Gustafson, N. M. Nieveen, & T. Plomp, T. (Eds.), *Design approaches and tools in education and training* (pp. 1-14). Springer.
- Van den Akker, J. (2013). Curricular development research as specimen of educational design research. In P. Tjeerd & N. Nienke (Eds.), *Educational design research*. Part a: An introduction (pp. 53–70). SLO.
- Van Til, A., Beeker, A., Fasoglio, D., & Trimbos, B. (2011). *Toetsen en beoordelen met het ERK*. Cito/SLO.
- Vogt, K., & Tsagari, D. (2014). Assessment literacy of foreign language teachers: Findings of a European study. *Language Assessment Quarterly*, 11(4), 374-402. <https://doi.org/10.1080/15434303.2014.960046>
- Voogt, J., Westbroek, H., Handelzalts, A., Walraven, A., McKenney, S., Pieters, J., & De Vries, B. (2011). Teacher learning in collaborative curriculum design. *Teaching and Teacher Education*, 27(8), 1235-1244.
- Wall, D. (2013). Washback. In Fulcher, G., & Davidson, F. (Eds.), *The Routledge handbook of language testing* (pp. 93-106). Routledge.
- Wallace, M. R. (2009). Making sense of the links: Professional development, teacher practices, and student achievement. *Teachers College Record*, 111(2), 573-596.
- Wang, H. & Cheng, L. (2009). Factors affecting teachers' curriculum implementation. *Linguistics Journal* 4 (2), 135–166.
- Webster-Stratton, C., Reinke, W. M., Herman, K. C., & Newcomer, L. L. (2011). The incredible years teacher classroom management training: The methods and principles that support fidelity of training delivery. *School Psychology Review*, 40(4), 509-529. <https://doi.org/10.1080/02796015.2011.12087527>
- West, L., & Verspoor, M. (2016). An impression of foreign language teaching approaches in the Netherlands. *Levende Talen Tijdschrift*, 17(4), 26-36.
- Westhoff, G. (2007). Challenges and opportunities of the CEFR for reimagining foreign language pedagogy. *The Modern Language Journal*, 91(4), 676-679.

- Wigglesworth, G. (2008). Task and performance based assessment. In N. H. Hornberger (Ed.), *Encyclopedia of language and education* (pp. 111-122). Springer. [https://doi.org/10.1007/978-0-387-30424-3\\_171](https://doi.org/10.1007/978-0-387-30424-3_171)
- William, D., & Thompson, M. (2007). Integrating assessment with instruction: What will it take to make it work? In C. A. Dwyer (Ed.). *The future of assessment: Shaping teaching and learning* (pp. 53–82). Lawrence Erlbaum Associates.
- Yan, X., Zhang, C., & Fan, J. J. (2018). "Assessment knowledge is important, but...": How contextual and experiential factors mediate assessment practice and training needs of language teachers. *System*, 74, 158-168. <https://doi.org/10.1016/j.system.2018.03.003>
- Zhu, Y., & Shu, D. (2017). Implementing foreign language curriculum innovation in a Chinese secondary school: An ethnographic study on teacher cognition and classroom practices. *System*, 66, 100-112. <https://doi.org/10.1016/j.system.2017.03.006>
- 10voordeleraar.(2017).*Kennisbases*.[Nationalstandardsforteachertraining]<https://10voordeleraar.nl/kennisbases/publicaties>

# Appendix A

## Observation Instrument for Communicative FL Teaching



<https://www.hu.nl/-/media/hu/documenten/onderzoek/projecten/observation-instrument-communicative-modern-foreign-language-teaching-feb-2023.ashx>

# Appendix B

Handleiding Communicatief toetsen en formatief handelen



<https://www.hu.nl/onderzoek/publicaties/handleiding-communicatief-toetsen-en-formatief-handelen>

# Nederlandse samenvatting

## Inleiding

Het leren van verschillende talen zorgt niet alleen op persoonlijk niveau voor meer communicatieve mogelijkheden en meer arbeidskansen (o.a. Fox et al., 2019; Kroll et al., 2017), maar zorgt op maatschappelijk niveau ook voor een beter intercultureel begrip (Jiang & Wang, 2018). Vooral het laatste wordt steeds belangrijker in de huidige internationaal georiënteerde en multiculturele samenleving. De Raad van Europa benadrukt daarom het belang van meertalig onderwijs dat taaldiversiteit waardeert en benut en het leren van meerdere talen bevordert.

De afgelopen decennia hebben veel negatieve veranderingen gebracht voor het talenonderwijs in Nederland, met name voor andere schooltalen dan Engels. Het niveau is omlaaggegaan en de motivatie is gedaald. Leerlingen die hun hele schoolcarrière lessen Frans, Duits of Spaans hebben gevolgd, blijken maar beperkt in staat om in deze moderne vreemde talen (mvt) te communiceren. Mvt-leraren willen het talenonderwijs daarom aantrekkelijker en effectiever maken. Ze zouden graag een meer communicatieve benadering willen toepassen in hun onderwijs, met onder andere veel aandacht voor de ontwikkeling van de productieve vaardigheden, maar ze vinden dit moeilijk te realiseren in de praktijk.

Huidige uitgangspunten van de mvt-vakken in het voortgezet onderwijs in Nederland zijn gebaseerd op de principes van communicatief talenonderwijs. Eindtermen zijn geformuleerd vanuit een communicatief perspectief, gebaseerd op het Europees Referentiekader (ERK). De mvt-onderwijspraktijk in Nederland richt zich daarentegen veelal op kennis van grammaticaregels en woordenschat buiten context in plaats van op de ontwikkeling van taalvaardigheden, met name in de onderbouw (Fasoglio et al., 2015; West & Verspoor, 2016). Leraren noemen als voornaamste reden hiervoor dat ze zich aan moeten passen aan eisen van hun sectie of school. Ze voelen zich genoodzaakt een bepaald aantal hoofdstukken per schooljaar te behandelen en/of een vastgesteld aantal (methode)toetsen af te nemen die voornamelijk de kennis van grammaticaregels en woordenschat toetsen.

Om effectief leren mogelijk te maken moeten, volgens de *constructive alignment* theorie van Biggs (1996), de belangrijkste componenten van een curriculum – leerdoelen, onderwijsactiviteiten en toetsing – op elkaar zijn afgestemd (zie Figuur 1). Vooral de toetsing speelt hierin een cruciale rol, omdat het een belangrijk teruglageffect heeft op het vakdidactisch handelen van leraren en op de manier waarop leerlingen leren (zgn. *washback effect*; Alderson & Wall, 1993; Green, 2007). Als mvt-leraren vooral de

niet-communicatieve toetsen uit de leergangen gebruiken met bijvoorbeeld focus op grammatica, bestaat het risico dat de communicatieve leerdoelen naar de achtergrond verdwijnen en de didactiek onvoldoende communicatief wordt. Het structureel aanbieden van een communicatief toetsprogramma waarin de nadruk ligt op de productieve vaardigheden, zou de keuze voor een meer communicatieve didactiek (leer- en onderwijsactiviteiten) positief kunnen beïnvloeden en kunnen zorgen voor meer consistentie in het curriculumontwerp.

Eerdere studies naar het *washback effect* van toetsen in de context van mvt-onderwijs zijn voornamelijk uitgevoerd in gecentraliseerde onderwijssystemen. In deze studies werden de effecten van grootschalige nationale examens, ontwikkeld door toetsexperts, onderzocht (Chan, 2020; Rea-Dickins & Scott, 2007; Spratt, 2005). Weinig studies hebben gekeken naar de effecten van toetsen die zijn ontwikkeld en afgenomen in de klas door mvt-leraren zelf. Het is echter belangrijk om inzicht te krijgen in de effecten van deze toetsing, vooral in een onderwijssysteem zoals in Nederland, waarin alleen een deel van de examinering op nationaal niveau plaatsvindt en waar leraren een grote mate van autonomie hebben in de manier waarop ze leerdoelen in de klas toetsen.

Het probleem is dat het ontwikkelen en organiseren van communicatieve toetsen niet gemakkelijk is in de praktijk. Schrijven, rapporteren of spreken zijn complexe handelingen en de beoordeling hiervan vereist van leraren veel tijd, aandacht en expertise. Er is geen sprake van eenduidig goede of foute antwoorden, zoals bij meer kennisgerichte toetsen vaak wel het geval is. Wat de taak nog uitdagender maakt, is dat de opvatting over toetsen de laatste jaren veranderd is. De nadruk ligt niet langer uitsluitend op de summatieve functie van toetsen, gericht op het vaststellen van het niveau van de kennis en vaardigheden van leerlingen aan het einde van het leerproces, maar ook steeds meer op de formatieve functie. Formatieve activiteiten bieden leerlingen de mogelijkheid om te oefenen zonder formele consequenties voor hun school of toekomstige academische carrière en geven zowel leerlingen als leraren inzicht in het leerproces (Black & William, 1998; Dixson & Worrell, 2016). Een effectief communicatief toetsprogramma zou dus niet alleen summatieve toetsen moeten bevatten, maar ook formatieve activiteiten, met als uiteindelijk doel het leren van de leerlingen te bevorderen. In Nederland worden leraren in het voortgezet onderwijs zich steeds meer bewust van het belang van formatief handelen. De daadwerkelijke invoering in de praktijk is echter nog beperkt (Krijgsman et al., 2023; Commissie Kwaliteit Schoolexaminering, 2019).

In dit ontwerpgericht promotieonderzoek is onderzocht welke tools mvt-leraren nodig hebben om een communicatief toetsprogramma te kunnen implementeren met een adequate balans tussen summatieve en formatieve toetsactiviteiten. Het uiteindelijke doel was om inzicht te krijgen in hoe deze implementatie een communicatieve didac-

tiek kan bevorderen. Om de ecologische validiteit te waarborgen, werd dit onderzoek uitgevoerd in echte schoolomgevingen en in co-creatie met mvt-leraren uit de praktijk. In totaal zijn er 52 leraren Engels, Frans, Duits en Spaans, 26 talensecties, 24 scholen en ongeveer 1500 leerlingen betrokken geweest.

## **Opzet en belangrijkste bevindingen**

Dit proefschrift bestaat uit vier verschillende studies.

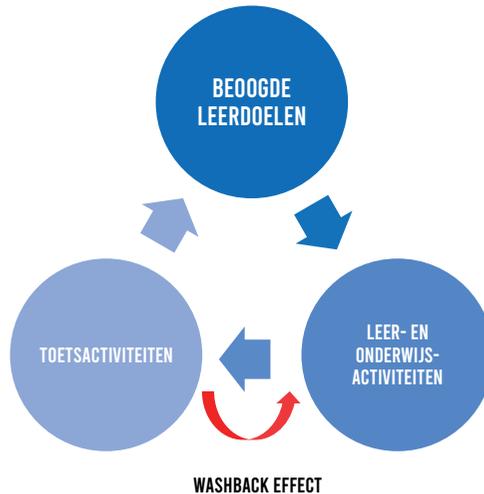
### ***Studie 1: Inventarisatie (hoofdstuk 2)***

Ten eerste is er een inventarisatiestudie uitgevoerd om de mate van samenhang in vreemdetalencurricula voor de onderbouw vast te stellen en de factoren te identificeren die van invloed kunnen zijn op vakdidactische keuzes.

Om de huidige situatie ten aanzien van constructive alignment (Biggs, 1996) tussen communicatieve leerdoelen enerzijds (nl. kerndoelen) en de ingezette didactiek en toetsing anderzijds in de betrokken scholen gedetailleerd in kaart te brengen, werd gebruik gemaakt van vier verschillende methodes: lesobservaties ( $N = 31$ ), analyse van lesmateriaal en toetsen ( $N = 43$ ), interviews met mvt-leraren ( $N = 21$ ) en een leerlingenenquête ( $N = 473$ ). De lesobservaties werden uitgevoerd met behulp van een vakdidactisch observatie-instrument dat in het lectoraat Didactiek van de Moderne Vreemde Talen ontwikkeld en gevalideerd is. Dit instrument bevat indicatoren die principes van communicatief talenonderwijs waarneembaar maken en daarmee de kwaliteit van het vakdidactisch handelen van mvt-leraren meetbaar maken. Analyse van lesmateriaal en toetsen vond plaats op basis van een checklist met criteria voor de mate van communicativiteit van leermiddelen (Van Til et al., 2011). In interviews zijn we met deelnemende leraren in gesprek gegaan over de mate waarin zij samenhang in hun onderwijsontwerp realiseren en waar ze hun vakdidactische keuzes op baseren.

De resultaten uit de inventarisatiestudie, gepresenteerd in hoofdstuk 2, laten een gebrek aan samenhang zien in de onderbouw-curricula van het mvt-onderwijs, met name tussen de landelijke communicatieve kerndoelen en eindtermen aan de ene kant en de toetsactiviteiten aan de andere kant. Geobserveerde toetsactiviteiten waren voornamelijk gericht op kennis van grammaticaregels en woordenschat buiten context, en op leesvaardigheid, de enige vaardigheid die op nationaal niveau wordt getoetst aan het einde van het voortgezet onderwijs. Weinig toetsactiviteiten waren gericht op de ontwikkeling van de productieve vaardigheden en in het bijzonder op de ontwikkeling van spreekvaardigheid. Geobserveerde toetsactiviteiten die wel gericht waren op de ontwikkeling van de productieve taalvaardigheden waren vaak vergelijkbaar met vertaaltaken en vroegen meestal niet om een passende mate van vrije productie op A2-niveau (het meest getoetste niveau in de onderbouw). Bovendien gebruikten slechts

een paar individuele leraren of talensecties rubrics gebaseerd op de *can-do statements* van het ERK. Deze resultaten komen overeen met de resultaten van de leerlingenenquête.



**Figuur 1** Constructive alignment, met washback effect (naar Biggs, 1996)

Wat betreft de factoren die hun vakdidactische keuzes beïnvloeden, gaven de geïnterviewde leraren aan dat ze de tijd en expertise misten om hun eigen materialen te ontwikkelen, en dat ze daarom een leergang nodig hadden. Vrijwel alle mvt-leraren in Nederland gebruiken inderdaad een leergang en bijbehorende toetsen als uitgangspunt voor het vormgeven van hun onderwijs (Fasoglio et al., 2015). De leergang dient dus de facto als curriculum (Guerrettaz & Johnston, 2013). Recente leergangen bieden veel authentiek en interactief materiaal, in lijn met de communicatieve benadering, maar ook nog steeds veel traditionele grammatica-, vertaal- en invuloefeningen. De toetsen die bij de leergangen horen, toetsen vaak nauwelijks de communicatieve taalvaardigheid. De productieve vaardigheden, en met name spreekvaardigheid, krijgen daardoor relatief weinig aandacht in toetsprogramma's. Uit de resultaten bleek verder dat de vakdidactische keuzes van leraren soms gekleurd zijn door misvattingen over communicatief taalonderwijs. Een aantal deelnemers benadrukte het belang van het opdoen van 'basiskennis' (zoals grammaticaregels en vocabulaire buiten context) voorafgaand aan het ontwikkelen van taalvaardigheden (vgl. *focus-on-forms*; Long, 1991). Deze misvatting is eerder in andere contexten geïdentificeerd als een belangrijke factor van invloed op vakdidactische keuzes (Chang, 2011; Wang & Cheng, 2009). In het gedecentraliseerde Nederlandse onderwijssysteem, waar leraren een aanzienlijke mate van autonomie hebben bij het selecteren, ontwikkelen en uitvoeren van toetsactiviteiten, lijken dergelijke misvattingen nog meer invloed te hebben op de lespraktijk.

**Studie 2: Ontwerp (hoofdstuk 3)**

Gebaseerd op de resultaten van de eerste studie, zoals gepresenteerd in hoofdstuk 2, werd vervolgens een ontwerpgericht onderzoek uitgevoerd binnen een professionele leergemeenschap (PLG). Een groep van 21 mvt-leraren van de meest aangeboden moderne vreemde talen in Nederland (Engels, Frans, Duits en Spaans) van 15 verschillende scholen nam deel aan deze PLG. Aan de hand van theorie- en praktijkervaringen (Bakkenes et al., 2010) werden in co-creatie met mij als onderzoeker en lerarenopleider, kerncomponenten voor de ontwikkeling van een communicatief toetsprogramma geïdentificeerd: communicatieve toetsopdrachten, summatieve toetsen, formatieve toetsactiviteiten, rubrics en een planning met de verdeling van toetsen over het schooljaar. Vervolgens hebben we voorlopige ontwerpprincipes geformuleerd voor elk van deze componenten, gericht op zowel validiteit als haalbaarheid. De uitdagingen waarmee de deelnemers werden geconfronteerd bij de ontwikkeling en implementatie van (deel van) het communicatieve toetsprogramma, werden geïdentificeerd en verwerkt in de definitieve ontwerpprincipes. Aantekeningen over het verloop van elke PLG-sessie en de uitwisselingen, opmerkingen en vragen van de deelnemers werden door mij als onderzoeker vastgelegd in een logboek. Feedback van de deelnemers werd gebruikt om de ontwerpprincipes aan te scherpen.

Vanuit de definitieve ontwerpprincipes zijn vervolgens tools ontworpen waarmee andere mvt-leraren een eigen communicatief toetsprogramma kunnen ontwikkelen. De ontwikkelde tools omvatten: een voorstel voor een planning met de verdeling van toetsen over het schooljaar (aan te passen voor de eigen schoolcontext), een checklist met criteria waar communicatieve toetsen aan moeten voldoen, toetsmatrijzen voor het samenstellen van summatieve toetsen, voorbeeld van summatieve toetsen en bijbehorende rubrics, voorbeelden van communicatieve formatieve activiteiten, en voorbeelden van communicatieve (toets)opdrachten. Uiteindelijk zijn de ontworpen tools verzameld in een toolbox die beschikbaar is gesteld op een digitaal platform en in een handleiding.

De resultaten van de co-designstudie, zoals gepresenteerd in hoofdstuk 3, laten zien dat praktische uitdagingen tijdens het ontwerpproces inderdaad vaak gerelateerd waren aan de beperkte tijd die toegewezen was aan het ontwikkelen, selecteren, uitvoeren en beoordelen van communicatieve toetsactiviteiten. De conceptuele factoren die geïdentificeerd werden, hadden vooral te maken met een gebrek aan expertise over hoe specifieke principes van communicatief taalonderwijs, zoals onvoorspelbaarheid, focus-on-form en de integratie van verschillende taalvaardigheden, toegepast konden worden.

### Studie 3: Implementatie (hoofdstuk 4)

In een derde deelstudie is vervolgens onderzocht hoe de ontworpen tools uit de voorgaande studie werden gebruikt door andere leraren ( $N = 32$ ) uit veertien talensecties van tien nieuwe scholen en hoe de implementatie van de communicatieve toetsprogramma's die zij met de tools ontwikkelden, werd ervaren door zowel leraren zelf als hun leerlingen.

De implementatie werd geëvalueerd op basis van het curriculummodel van Van den Akker (2013), zoals beschreven in Figuur 2. We hebben eerst geobserveerd of het ontwerp werd *uitgevoerd* zoals *beoogd* en gekeken in hoeverre de componenten van de ontwikkelde communicatieve toetsprogramma's overeenkwamen met de ontwerpprincipes. Vervolgens hebben we door middel van logboeken en interviews onderzocht hoe de implementatie door leraren werd ervaren (*geïnterpreteerd*). Daarnaast hebben we gekeken hoe de implementatie door leerlingen werd *ervaren* met betrekking tot de relatie tussen doelen, leeractiviteiten en toetsing. Dit werd gedaan met een leerlingengenquête ( $N = 423$ ), aangevuld met leerlingenterviews ( $N = 6$ ).

<b>Beoogd</b>	<i>Denkbeeldig</i>	Visie (onderliggende ideeën van het curriculum).
	<i>Geschreven</i>	Bedoelingen van het curriculum zoals uitgewerkt in documenten en materialen.
<b>Uitgevoerd</b>	<i>Geïnterpreteerd</i>	Het curriculum zoals geïnterpreteerd door de gebruikers (met name de leraren).
	<i>In actie</i>	Het daadwerkelijk proces van lesgeven en leren, dus de uitvoering.
<b>Bereikt</b>	<i>Ervaren</i>	De leerervaringen van deelnemers (met name de leerlingen).
	<i>Geleerd</i>	De leerresultaten van de deelnemers.

**Figuur 2** Curriculumniveaus (Van den Akker, 2013, p.56)

Resultaten laten voor de meeste talensecties een acceptabel tot hoog niveau van *treatment fidelity* zien met betrekking tot de verschillende ontwerpprincipes of 'hoofdingrediënten' (Harn et al., 2013) zoals zichtbaar in de ontwikkelde toetsprogramma's. Het communicatieve karakter en de koppeling aan het ERK waren duidelijk opgenomen in de toetsactiviteiten en rubrics én waren zichtbaar in de meeste informele toetsactiviteiten zoals die tijdens de lessen geobserveerd werden. Uit deze studie bleek ook dat de meerderheid van zowel de betrokken leraren als de leerlingen de implementatie als positief heeft ervaren. Leraren gaven aan dat de tools inspirerend waren en hen hielpen bij het ontwikkelen van hun eigen toetsprogramma. Deze resultaten ondersteunen het idee dat een interventie die voldoende flexibiliteit biedt, niet alleen als inspiratiebron werkt maar ook goed toepasbaar is in de eigen lespraktijk.

Tot slot benadrukten de betrokken leraren hun enthousiasme over het nieuwe programma en merkten ze ook enthousiasme bij hun leerlingen. Dit laatste aspect is zeker relevant, gezien het feit dat plezier en voldoening de motivatie stimuleren, wat gezien kan worden als drijvende kracht achter het leren van een vreemde taal (Ellis, 2019). De meerderheid van de bevroegde leerlingen meldde dat de leer- en toetsactiviteiten meer gericht waren op de ontwikkeling van taalvaardigheden, dat ze meer tijd hadden om kennis in praktijk te brengen en dat ze actiever waren tijdens de lessen. De geïnterviewde leerlingen benadrukten ook dat ze meer zelfvertrouwen hadden tijdens de formele toetsen omdat ze meer tijd hadden om te oefenen met informele communicatieve toetsactiviteiten tijdens de lessen. Deze laatste bevinding is belangrijk, want hoe hoger het niveau van zelfvertrouwen, hoe meer leerlingen zich durven uit te drukken in de vreemde taal en hoe groter de kans dat ze een hoger niveau van taalvaardigheid bereiken (MacIntyre et al., 1998).

Hoewel leraren de implementatie van de toetsprogramma's in veel opzichten beschouwden als een positieve ervaring, meldden ze ook een aantal belemmeringen. De belangrijkste aandachtspunten waren een gebrek aan ervaring bij leerlingen en leraren met de toepassing van communicatieve principes en de moeilijkheid om de voortgang van individuele leerlingen bij te houden in grote klassen. Sommige deelnemers noemden ook een gebrek aan facilitering van de school, die hen onvoldoende tijd gaf om te leren en te experimenteren binnen het project. In de interviews noemden leraren ook een gebrek aan ervaring met het ERK. Dit gebrek aan ervaring bleek te leiden tot extra werkdruk, omdat het meer tijd kostte dan normaal om zowel rubrics als geschikte communicatieve toetsactiviteiten te ontwikkelen.

Tot slot deelden sommige leraren hun zorgen over de overgang naar de bovenbouw en 'reguliere' klassen, en het mogelijk gebrek aan 'basiskennis' van woorden en grammaticaregels bij leerlingen, omdat deze onderdelen in de nieuwe aanpak niet geïsoleerd werden behandeld en getoetst. Deze zorgen laten opnieuw zien dat mvt-leraren misvattingen hebben over communicatief taalonderwijs en beperkt inzicht in hoe een *focus-on-form* in plaats van *focus-on-forms* tot betere en duurzamere leerresultaten kan leiden (Ellis, 2015).

#### **Studie 4: Effecten (hoofdstuk 5)**

Uiteindelijk is in de laatste projectfase een effectstudie uitgevoerd, waarin werd onderzocht hoe de structurele implementatie van een communicatief toetsprogramma het vakdidactische handelen van leraren beïnvloedde. Deze studie werd uitgevoerd in de tien scholen en veertien talensecties die ook deelnamen aan de implementatiefase van het project.

Voor deze laatste studie hebben we de lespraktijk in reguliere klassen (controlegroep) vergeleken met de lespraktijk in interventieklassen, waarin het ontworpen communicatieve toetsprogramma geïntroduceerd was. Gegevens zijn verzameld door middel van lesobservaties in reguliere klassen ( $N = 12$ ) en in interventieklassen ( $N = 12$ ), logboeken van het lerarenteam van elke deelnemende talensectie ( $N = 14$ ), en enquêtes onder leerlingen in reguliere klassen ( $N = 330$ ) en in interventieklassen ( $N = 423$ ).

De resultaten van de laatste studie (hoofdstuk 5) naar de geobserveerde en ervaren effecten van de structurele introductie van een communicatief toetsprogramma op het vakdidactisch handelen van mvt-leraren, geven aan dat de introductie van dergelijke toetsprogramma's een positief terugslageffect (*washback*) heeft op de lesactiviteiten. De geobserveerde lessen in de interventieklassen waren allemaal voornamelijk gericht op de ontwikkeling van taalvaardigheden en met name de productieve vaardigheden binnen een betekenisvolle context, terwijl in de reguliere klassen een groot deel van de geobserveerde lessen nog steeds voornamelijk gericht was op kennis van vormaspecten, zoals grammaticaregels buiten een betekenisvolle context.

Vergeleken met de geobserveerde communicatieve lesactiviteiten in reguliere klassen, sloten de lesactiviteiten in interventieklassen significant beter aan bij communicatieve principes en leerdoelen. Deze activiteiten waren betekenisvoller, authentieker en realistischer, vroegen om meer vrije productie en hanteerden een meer geïntegreerde aanpak. Wat betreft het gebruik van de doeltaal zagen we geen significant verschil, maar leraren in interventieklassen leken de doeltaal wel bewuster te gebruiken om leerlingen de input te geven die ze nodig hadden om hun taalvaardigheid te verbeteren. Tot slot zagen we dat er in de interventieklassen meer gebruik werd gemaakt van informele toetsactiviteiten dan in de reguliere klassen. Deze activiteiten waren direct gekoppeld aan de communicatieve leerdoelen en beoordelingscriteria. Dit is een positief aspect, omdat het gebruik van formatieve activiteiten die goed zijn afgestemd op de leerdoelen, bijdraagt aan de leervorderingen van leerlingen (Black & Wiliam, 2018).

De meeste geobserveerde effecten werden ook als zodanig ervaren door de betrokken leraren en leerlingen. De meerderheid van de leraren gaaf aan meer aandacht te hebben voor de ontwikkeling van taalvaardigheden en in het bijzonder van productieve vaardigheden, evenals een *focus-on-form* binnen communicatieve activiteiten in plaats van een *focus-op-forms* buiten context. Sommigen gaven ook aan de indruk te hebben de doeltaal vaker te gebruiken.

De meeste leerlingen in de interventieklassen meldden dat leeractiviteiten, formele toetsen en informele toetsactiviteiten vooral gericht waren op de ontwikkeling van taalvaardigheden, terwijl volgens leerlingen uit reguliere klassen de focus vooral lag op

de kennis van grammaticaregels en losse woorden. Deze gecombineerde bevindingen laten zien dat de implementatie van een toetsprogramma dat afgestemd is op communicatieve leerdoelen, kan helpen om een meer traditionele vormgerichte aanpak bij te sturen richting een meer communicatieve aanpak.

## Conclusie

De resultaten van de studies die deel uitmaken van dit proefschrift, geven aan dat de implementatie van een communicatief toetsprogramma de samenhang in mvt-curricula kan verbeteren en een positief terugslag-effect op de lesactiviteiten kan genereren. Dit onderzoek heeft aangetoond dat aanpassingen in de dagelijkse toetspraktijk kunnen leiden tot concrete positieve effecten op de vakdidactische aanpak van leraren en daarmee uiteindelijk op de beoogde leeropbrengsten van leerlingen. De resultaten laten zien dat deze positieve effecten betrekking hebben op zowel de selectie van communicatieve onderwijsactiviteiten als op de manier waarop leraren deze in hun lespraktijk introduceren. Over het algemeen bleek uit de resultaten ook dat zowel leraren als leerlingen positief waren over de introductie van een communicatief toetsprogramma.

Daarnaast heeft dit onderzoek ook direct invloed gehad op mijn eigen onderwijspraktijk in de mvt-lerarenopleiding. In de leerlijn vakdidactiek die recent opnieuw ontwikkeld is voor alle moderne vreemde talen, wordt nu bijvoorbeeld meer aandacht besteed aan curriculumontwerp, *constructive alignment*, communicatief toetsen en formatief handelen. Bovendien zijn thema's over de invoering van communicatief vreemdetalenonderwijs in de praktijk nu onderdeel van trainingen die gevolgd worden door mvt-leraren die onze studenten begeleiden op onze partnerscholen. Verder zijn de inzichten en praktische opbrengsten van het onderzoek gepresenteerd, niet alleen aan andere onderzoekers, maar ook aan veel mvt-leraren (al dan niet in opleiding), talensecties, lerarenopleiders, uitgevers, toetsexperts en leerplanontwikkelaars. De feedback en discussies naar aanleiding van deze ontmoetingen hebben zowel mijn onderzoek als hun praktijk verbeterd.

## Aanbeveling voor de onderwijspraktijk

Omdat veel leraren leergangen als basis gebruiken voor het vormgeven van hun onderwijs, zouden educatieve uitgevers meer communicatieve les- en vooral toetsactiviteiten moeten opnemen in mvt-leergangen. Lerarenopleidingen zouden zeker ook een rol kunnen spelen door toekomstige leraren te trainen in het selectief gebruiken van leergangen met meer focus op de communicatieve activiteiten die ze bevatten, het selecteren en aanpassen van communicatieve leer- en toetsactiviteiten uit andere bronnen en/of het zelf ontwikkelen hiervan.

Daarnaast is het belangrijk dat mvt-leraren de principes van communicatief taalonderwijs goed kennen en weten hoe ze deze kunnen toepassen in de context van het voortgezet onderwijs in Nederland. Daarom zouden cursussen vakdidactiek in de mvt-lerarenopleidingen zich nog meer moeten richten op hoe communicatief taalonderwijs te introduceren in grote klassen met weinig lessen, en op hoe vormgerichte instructie toe te passen binnen communicatieve leertaken. Tenslotte zouden toekomstige mvt-leraren moeten leren hoe ze een communicatief toetsprogramma kunnen ontwikkelen en introduceren.

Over het algemeen zouden leraren in de dagelijkse onderwijspraktijk tijd moeten krijgen om zich te professionaliseren en te laten begeleiden bij het ontwikkelen en uitvoeren van communicatieve toetsing. Dit laatste punt lijkt van cruciaal belang, gezien het aangetoonde *washback-effect* van toetsen. Om mvt-leraren te begeleiden bij deze complexe taak, zou curriculumdocumentatie niet alleen moeten verwijzen naar het ERK, maar in ieder geval ook de belangrijkste principes van communicatief taalonderwijs explicieter moeten vermelden, zoals het belang van integratie van taalvaardigheden, het vergroten van onvoorspelbaarheid in communicatieve taken en het gebruik van vormgerichte instructie binnen betekenisvolle communicatieve situaties.

Momenteel worden de eindtermen voor het voortgezet onderwijs in Nederland herzien, ook voor het mvt-onderwijs. Leraren, lerarenopleiders en onderzoekers zijn uitgenodigd om deel te nemen aan dit proces. Om ervoor te zorgen dat de nieuw geformuleerde kerndoelen en eindtermen vervolgens goed worden vertaald naar de onderwijspraktijk, is het essentieel dat ze duidelijke richtlijnen bevatten over hoe ze kunnen worden vertaald naar een communicatief toetsprogramma. De resultaten van dit onderzoek zouden ook meegenomen kunnen worden bij de ontwikkeling van nationale examenprogramma's. Deze nieuwe programma's zouden de nadruk moeten leggen op het op een geïntegreerde wijze toetsen van taalvaardigheden en zouden de productieve vaardigheden, en spreekvaardigheid in het bijzonder, meer centraal moeten stellen. Op basis van dit onderzoek kunnen we zeker verwachten dat dit positief resultaat zal hebben op de aantrekkelijkheid, het belang en de leeropbrengsten van het vreemdetalenonderwijs.

# Dankwoord/Remerciements

Hoe kan ik in een paar woorden alle mensen bedanken die mij hebben geholpen, gesteund, begeleid, vertrouwd, getroost, aangemoedigd, gemotiveerd, naar me hebben geluisterd, van goede advies hebben voorzien of gewoon er waren voor mij tijdens deze promotiejaren? Niet dus, daarom meer dan duizend. Ik zal proberen niemand te vergeten, maar ik wil sowieso alvast iedereen bedanken die ook simpelweg heeft gevraagd hoe het met mijn proefschrift ging.

Allereerst wil ik mijn begeleiders, Rick en Catherine, speciaal bedanken voor de geweldige begeleiding en steun die ze me gegeven hebben. Rick, ik herinner me onze eerste ontmoeting toen ik solliciteerde voor een positie als vakdidactica in je nieuwe lectoraat. Ik was onder de indruk van deze hoogleraar met zoveel ervaring. Vanaf de eerste werksessie met de onderzoeksgroep ontdekte ik de mens achter de titel: een heel attent en toegankelijk persoon die het verband weet te leggen tussen het leven en het werk door zijn persoonlijke ervaringen te delen en interesse te tonen in die van anderen. Dit traject zou heel anders zijn gelopen zonder de menselijke dimensie die jij zo goed weet te creëren. Ik vraag me af of er veel hoogleraren zijn die tijdens de kerstvakantie, in het weekend of op regenachtige doordeweekse dagen, persoonlijk langsgaan bij promovendi en collega's om zelfgemaakte kerstkoekjes, bloemen of glutenvrije taartjes te geven om een nieuwe mijlpaal te vieren. Bedankt natuurlijk ook voor je deskundigheid en de verdiepende gesprekken die mij op weg hebben geholpen. Catherine, zoals je weet, was ik aanvankelijk nogal afgeschrikt door je rode pen. Ondanks je bemoedigende zinnen in je mails: *'Je bent er bijna!'* of *'Wat een werk heb je verricht!'* was de hoeveelheid feedback in mijn document in de bijlage toch altijd even schrikken. Ik merkte echter al snel dat je rode pen inderdaad (vaak) goed werkte! De precisie waarmee je mijn teksten van feedback voorzag, hielp me echt om vooruit te komen en ik weet inmiddels dat als je uiteindelijk schreef: *'Het ziet er goed uit!'* dat dat ook echt zo was. Ik heb geen vergelijkingsmateriaal, maar ik kan alleen maar wensen dat andere promovendi zo'n betrokken, deskundige en efficiënte begeleider krijgen. Ik wil je ook in het bijzonder bedanken voor je luisterende oor en attente appjes die me door de dalen van het promoveren heen hebben geholpen.

Kristina, mijn geweldige mede-promovenda en collega, speciale dank aan jou voor je waardevolle feedback op mijn Engelse schrijfvaardigheid, je bijdrage bij analyses, de samenwerking tijdens cursussen en onze koffieafspraken tijdens de drukke lesdagen om te sparren over onze projecten. Ik heb zo veel bewondering voor je werk en hoe je op een laagdrempelige manier met dramatechnieken studenten en leerlingen over hun spreekangst weet te tillen. Ik ben blij dat je mijn paranimf wilde zijn.

Ik wil graag ook mijn andere collega's van de eerste onderzoeksgroep bedanken, Oliver, Karin, Tobey, Mariska, Irene en Marleen. Ik vond het geweldig om met jullie samen te werken en na te denken over de beste manier om vreemde talen te onderwijzen en onze studenten te helpen de kloof tussen theorie en praktijk te overbruggen. We kunnen trots zijn op het werk dat verricht is en op de samenwerking tussen de talen die ontstond en concreet resulteerde in een sterke gezamenlijke vakdidactische leerlijn. Ik wil ook de leden van de huidige onderzoeksgroep bedanken, die de prachtige taak hebben om meertaligheid te promoten en hiervan de waarde en rijkdom te laten zien. Jantien en Linda, bedankt dat jullie mij de kans geven om deel te blijven uitmaken van deze groep en betrokken te zijn bij boeiende projecten.

Dit onderzoek was niet mogelijk geweest zonder de betrokken leraren die meegedaan hebben aan de Professionele Leergemeenschap en soms tot in de late uurtjes hebben meegedacht en mee ontwikkeld. Er was altijd een fijne sfeer. De domtorentjes hebben hier zeker aan bijgedragen, maar vooral de manier waarop jullie met een open en positieve houding hebben meegedaan aan het project. Dank ook aan alle andere participanten, leraren en leerlingen voor jullie beschikbaarheid en openheid tijdens lesobservaties en interviews.

Ik wil graag ook een aantal mensen bij de UU, het Kenniscentrum en het Instituut Archimedes speciaal bedanken. Marjolein, Lyske, Pauline, Maaïke, Kirsten, jullie hebben mij fantastisch ondersteund. Nynke, bedankt voor het helpen met de delicate taak om onderwijs en onderzoek te combineren, op een efficiënte en eerlijke manier. Bedankt ook in het bijzonder aan mijn toenmalige teamleider bij IA, geweldige collega en *conseillère particulière* die mij vanaf het begin haar vertrouwen heeft gegeven en mij daarna quasi dagelijks heeft aangemoedigd en gesteund, en nog steeds doet: Marleen, *un grand merci* (bedenk hier een bloem). Veel dank ook aan mijn andere collega's van de vakgroep Frans, Irene, Frits, Nathalie en Maaïke, die een deel van mijn taken hebben moeten opvangen de afgelopen jaren. Bedankt ook aan mijn andere collega's bij IA voor de fijne samenwerking en hun enthousiasme en steun voor dit project. Tobey, thank you so much for having taken the time to read this entire dissertation in just a few days before the final check. I hope this sentence is written correctly! En niet te vergeten een speciale dank aan mijn studenten voor hun belangstelling in dit onderwerp en de interessante en inzichtelijke gesprekken tijdens de colleges.

Veel andere mensen buiten de professionele sfeer hebben een belangrijke rol gespeeld tijdens mijn promotietraject en ik wil ze ook hiervoor speciaal bedanken. Johanna en Maartje, mijn vriendinnen sinds mijn eerste jaar in Nederland. Wij hebben inmiddels zoveel leuke etentjes, drankjes en koffie uurtjes achter de rug. Bedankt dat jullie er altijd voor mij zijn en altijd bereid waren, ondanks de drukte van het leven, om mij te helpen

met mijn proefschrift. Jullie expertise, jij Maartje als onderzoeker en jij Johanna als leraar Duits en vakdidactica, hebben mij veel gebracht. Johanna, bedankt dat je mijn paranimf wilde zijn. Het was vanzelfsprekend om jou te vragen, de persoon met wie ik Nederlands heb geleerd en met wie ik de eerste stapjes in het onderwijs heb gezet. Saskia, bedankt voor je interesse en beschikbaarheid. Zo fijn dat ik van je ervaring als onderzoeker kon profiteren. Ik weet hoe druk je het hebt, maar je had altijd tijd voor mij even tussendoor om naar een van mijn vraagstukken te kijken. Bedankt ook voor de lekkere en gezellige etentjes op de dinsdagavond. Altijd super hectisch met de kinderen, maar waar ook altijd ruimte was, en nog steeds is, om te sparren over onderzoek en onderwijs. Ik wil graag ook mijn schoonfamilie bedanken voor hun oprecht interesse in mijn voortgang en het onderwerp. Zo leuk dat je mijn artikel in LTM wilde lezen en delen Annemijn!

Quelques mots de remerciements en français, adressés tout d'abord à mes copains d'Utrecht, les grands comme les petits, qui ont fait que la plupart de mes weekends et vacances n'étaient pas que dédiés à ma thèse mais aussi joyeusement ponctués par apéros, dîners, ballades en forêt, restos, terrasses et campings. Merci d'être si présents et d'avoir aussi partagé avec moi votre expérience de chercheurs et de prof. J'en ai de la chance de vous avoir ! Un merci tout particulier à Hélène qui après une longue journée de travail s'est prêtée au jeu pour essayer nos tâches communicatives dans son meilleur néerlandais. Heureusement qu'on avait du vin blanc (situation authentique oblige) ! J'aimerais bien sûr aussi remercier mes parents et mes soeurs, qui m'ont toujours encouragée et soutenue. Sophie, merci pour tes supers bons conseils. Marj, merci d'avoir pris le temps de réfléchir avec moi au design de mes projets. Et même si c'est en l'air, j'aimerais aussi remercier ma grand-mère, qui aurait été sans doute celle que j'aurais appelée en premier pour lui annoncer chaque nouvelle étape. Enfin, un énorme merci à mes petits chats, Chloé et Oscar qui ne m'ont absolument pas demandé comment allait ma thèse mais qui ont bien fait en sorte que je joue, danse, coure dehors, lise à haute voix les sept tomes d'Harry Potter, construisse des cabanes, chante, câline, console et bien d'autres activités sans aucun lien avec le travail.

En Machiel... Ik had dit absoluut niet kunnen doen zonder jou. Je bent gewoon een toffe peer (dit mag je uitspreken zoals je wil).

# About the Author

Charline Rouffet was born in Nemours, France, on October 28<sup>th</sup>, 1983. She studied Literary Studies at the University of Nîmes and University of Paris-IV Sorbonne. In 2005, she obtained her Master of Arts in Comparative Literature from the University of Paris-IV Sorbonne, which she partly completed at the Vrije Universiteit Amsterdam (VU) through the Erasmus Program. She then decided to stay in the Netherlands. After earning a Master of Arts in French Language and Culture (2006; cum laude) from the VU, she completed a Master's degree in Education from the Academic teacher training program at the same university in 2007. Her Master's research project focused on the use of the target language in the context of TaBaScO (Task Based School Organisation for the Acquisition of Language), a project that supports a task-oriented approach to language learning. She then worked as a French teacher in the lower and upper forms of several secondary schools. Since 2010, she has been working as a foreign language teacher educator in the Bachelor's and Master's teaching programs at University of Applied Sciences Utrecht (HU). In these programs she currently teaches French literature, language skills, methodology courses and supervises Master's practical research projects. Since 2016, she has been affiliated with the HU Research group Multilingualism and Education at the Research Center for Learning and Innovation. In September 2018, she was awarded a HU research Grant (*HU promotievoucher*) and became a PhD candidate at Utrecht University, with prof. dr. Rick de Graaff as promotor and dr. Catherine van Beuningen (Amsterdam University of Applied Sciences) as co-promotor. Her doctoral research project explores the influence of communicative assessments on foreign language teachers' pedagogical approaches and aims at transforming FL teaching through classroom-based assessment programs.

# List of Publications

- Rouffet, C. (2020). Communicatief toetsen binnen het vreemdetalenonderwijs. *Remediaal*, 20(2-3), 38-41.
- Rouffet, C., Van Beuningen, C., & De Graaff, R. (2021). Designing a communicative foreign language assessment program for Dutch secondary schools: A design-based research project within a professional learning community. In *Collated Papers for the ALTE 7th International conference, Madrid* (pp. 179-183).
- Rouffet, C., Van Beuningen, C., & De Graaff, R. (2022). Samenhang in vreemdetalencurricula: een verkenning van onderwijs- en toetspraktijken in het voortgezet onderwijs. *Levende Talen Tijdschrift* 23(3), 3-17.
- Rouffet, C., Van Beuningen, C., & De Graaff, R. (2023). Constructive alignment in foreign language curricula: An exploration of teaching and assessment practices in Dutch secondary education. *The Language Learning Journal*, 51(3), 344-358.
- Rouffet, C., Van Beuningen, C., & De Graaff, R. (2023). Toetsen om te leren: effecten van een communicatief toetsprogramma moderne vreemde talen. *Levende Talen Magazine*, 110(3), 24-29.
- Rouffet, C., De Kleyn, I., IJzerman, M., Van Beuningen, C., & De Graaff, R. (2023). Een vakdidactisch observatie-instrument voor communicatief vreemdetalenonderwijs: brug tussen theorie en praktijk, *Levende Talen Tijdschrift*, 24(2), 3-15.
- Rouffet, C., Van Beuningen, C., & De Graaff, R. (under revision). The implementation of a foreign language classroom-based assessment program in Dutch lower form secondary education.



In today's foreign language (FL) education, FL teachers universally recognize the importance of fostering students' ability to communicate in the FL. However, existing assessments often do not sufficiently evaluate this crucial aspect. Instead, assessments primarily focus on formal language knowledge in isolation, disconnected from real-world communicative contexts. This misalignment between assessment practices and communicative objectives, which is particularly prevalent in the lower form of secondary education in the Netherlands, hampers effective FL teaching.

The aim of this PhD research project was to gather insight into the potential of assessments to steer FL teaching practices. To this end, tools for developing communicative classroom-based assessment (CBA) programs were designed and implemented in practice, in close collaboration with FL teachers. The first study reported upon in this dissertation (Chapter 2) investigates which factors contribute to the identified lack of alignment between learning goals and assessment practices. The second study (Chapter 3) then shows how CBA-tools were co-designed to overcome the challenges FL teachers face when developing assessments. The third study (Chapter 4) explores how these tools were used by other FL teachers (who did not take part in the co-design process) to implement a communicative CBA program in their own context. In the final study (Chapter 5), effects of the newly designed CBA programs on teaching practices are investigated, taking both teachers' and students' perspectives into account. Findings reveal that assessing FL competencies in a more communicative way can transform teaching practices, placing communicative abilities at the heart of FL education.

Charline Rouffet is a foreign language teacher educator and a researcher at the University of Applied Sciences Utrecht, in collaboration with Utrecht University.