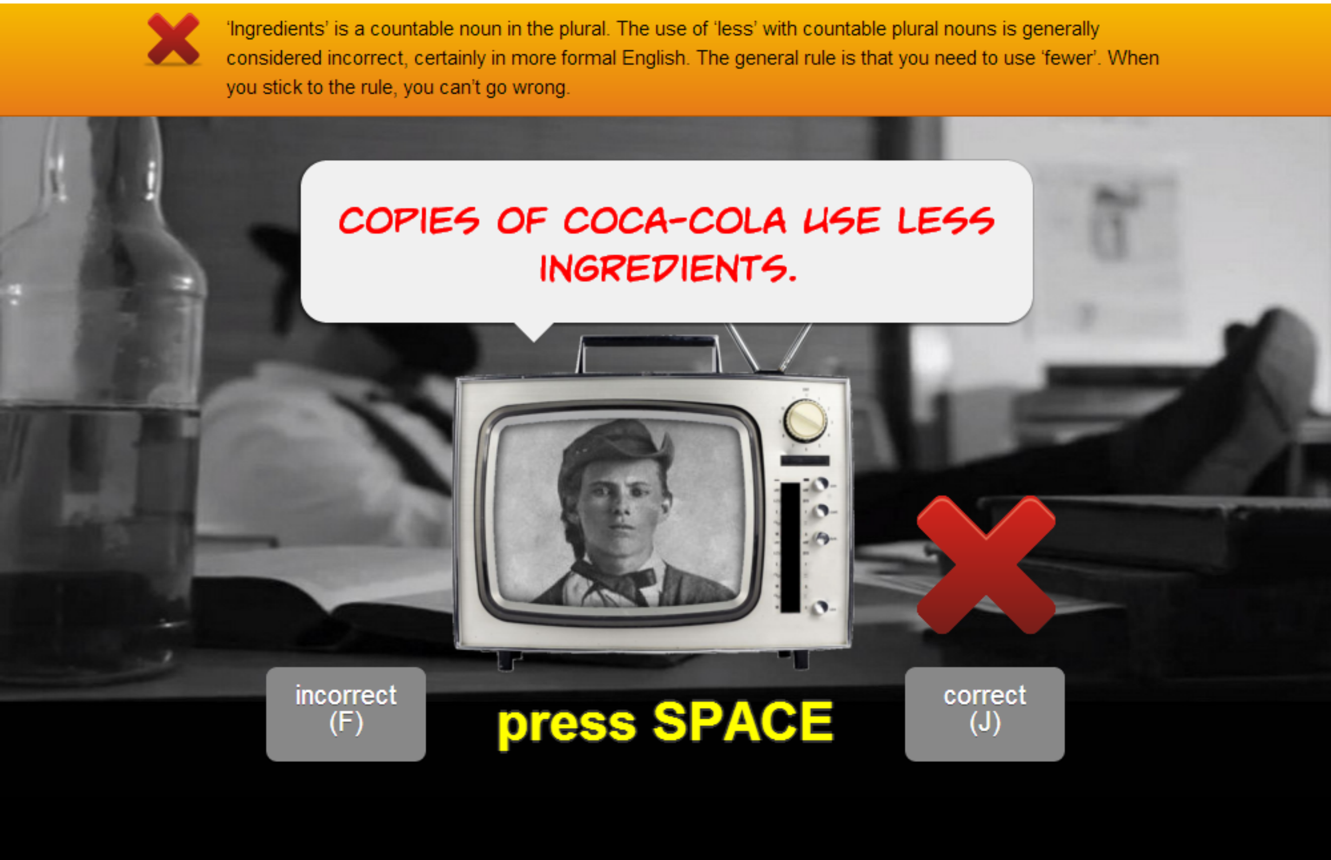
Adventures in red ink

Effectiveness of corrective feedback in digital game-based language learning

Proefschrift ingediend tot het behalen van de graad van
Doctor in de Taalkunde door
Frederik Cornillie

Promotor: Prof. dr. Piet Desmet
Copromotor: Prof. dr. Kris Van den Branden



'Ingredients' is a countable noun in the plural. The use of 'less' with countable plural nouns is generally considered incorrect, certainly in more formal English. The general rule is that you need to use 'fewer'. When you stick to the rule, you can't go wrong.

COPIES OF COCA-COLA USE LESS INGREDIENTS.

Katholieke Universiteit Leuven
Faculteit Letteren
Onderzoekseenheid Taalkunde

KU Leuven KULAK
Subfaculteit Letteren
iMinds - ITEC (Interactive Technologies)

Adventures in red ink

Effectiveness of corrective feedback

in digital game-based language learning

Proefschrift ingediend tot het behalen van de graad van
Doctor in de Taalkunde door

Frederik Cornillie

2014

Promotor
Prof. dr. Piet Desmet

Copromotor
Prof. dr. Kris Van den Branden

Lid van de begeleidingscommissie
Prof. dr. Steven L. Thorne

Juryleden
dr. Geraldine Clarebout
Prof. dr. Jozef Colpaert
dr. Elke Peters

Cover design
The sleeping detective is a still taken from the film noir project *The Big Smoke* (© 2012 Kenneth Gawne).
Icons are from the Coquette icon set.

Print
Uitgeverij University Press
Leegstraat 15
9060 Zelzate
België
Tel + 32 9 342 72 25
E-mail info@universitypress.be
Web www.universitypress.be

ISBN 978-94-6197-207-1

© 2014 Frederik Cornillie

All rights reserved. No part of this thesis may be reproduced or transmitted, in any form or by any means, electronic or mechanical, including photocopying, recording or otherwise, without written permission from the author or from the publishers holding the copyright of the published articles.

Table of contents

Abstract

Both in second and foreign language (L2) teaching environments and in digital games, feedback is considered indispensable as well as a powerful device to support learning. However, the state of affairs in the fields of second language acquisition and educational psychology shows that the effects of corrective (negative) feedback (CF) are not univocal, and suggests that the effectiveness of CF in digital game-based language learning is likely to depend on the following factors: the type of CF ('explicit' or 'implicit'), how 'learning' is measured (as the development of explicit or implicit L2 knowledge), and individual differences related to learners' receptivity to CF, namely perceived usefulness of CF and intrinsic motivation. The current PhD project investigates the complex interplay between these factors both from a theoretical point of view and empirically, and in a highly interdisciplinary way, combining insights from the literatures of second language acquisition, computer-assisted language learning (CALL; by itself a highly interdisciplinary undertaking), educational psychology and technology, motivational psychology, and game studies.

Subsequent to the general introduction (chapter 1) and the theoretical study (chapter 2), this dissertation presents the results of four empirical studies with prototypes of digital game-based experiences engineered for the instruction and practice of English as a L2. These were evaluated mainly in the context of secondary education.

The first empirical study (chapter 3) explores the role of individual differences vis-à-vis CF types in a 3D immersive game designed for the instruction of English pragmatics. It shows that learners (*N*=83) found explicit

CF more useful than more implicit CF, and that the perceived usefulness of explicit CF correlated positively with parameters of motivation.

The second empirical study (chapter 4) investigates learners' (N=36) use of explicit metalinguistic CF in a written interactive murder mystery, and cannot present any evidence that perceived usefulness predicted CF use; there was a strong positive association, however, between CF use and prior metalinguistic knowledge.

The third and fourth empirical studies focus on the effectiveness of grammar practice with CF in mini-games. The third study (chapter 5) found that vivid CF, adapted to the fantasy of the game concept, affected learners' (N=32) intrinsic motivation positively, which was related to their willingness to practise more.

The fourth empirical chapter (chapter 6) reports on the effects of grammar practice on L2 learning in a two-month study, comprising one month of practice. The results indicate that intensive practice supported by mini-games and CF helped learners (N=125; control group N=61) to develop L2 grammar knowledge that was useful for their performance on various transfer tasks (both near and far transfer). Moreover, the effects of explicit metalinguistic CF were, by and large, stronger than the effects of CF which did not include any metalinguistic explanation; this finding did not apply to learners' performance on a complex spoken language task.

The final chapter of this dissertation (chapter 7) discusses the main findings, and presents directions for future research. The results of this project bode well for the design of powerful technology-mediated language learning spaces which seek to engage learners by drawing on their gaming experience or interest, and which are aimed at supporting learners in mastering the conventions of formal language use.

Samenvatting in het Nederlands

Feedback is een essentieel en krachtig element in tweede- en vreemdetaalleeromgevingen en in games, en wordt verondersteld het leren optimaal te ondersteunen. Toch toont het huidige onderzoek naar tweedetaalverwerving, en naar leren en instructie in het algemeen, dat de effecten van correctieve (negatieve) feedback (CF) niet altijd even eenduidig zijn. De huidige stand van zaken van het onderzoek suggereert dat de effectiviteit van CF in taalleren ondersteund door middel van game-gebaseerde leeromgevingen mogelijks afhangt van de volgende factoren: het type CF ('explicit' of 'impliciet'), hoe 'leren' wordt gemeten (als de ontwikkeling van expliciete kennis, of van impliciete kennis van een tweede taal), en individuele verschillen tussen leerlingen die bepalen hoe ontvankelijk leerlingen zijn voor CF, namelijk hoe nuttig leerlingen dit vinden, en hun intrinsieke motivatie. Het huidige doctoraatsonderzoek legt zich toe op de complexe interactie tussen deze factoren, en dit zowel vanuit een theoretische invalshoek als op basis van empirisch onderzoek. Dit onderzoek is bovendien sterk interdisciplinair geïnspireerd, en combineert inzichten uit de literatuur rond tweedetaal-verwerving, computerondersteund taalonderwijs (CALL; op zich een sterk interdisciplinair onderzoeksveld), educatieve psychologie en technologie, motivatiepsychologie, en game studies.

Na de algemene inleiding (hoofdstuk 1) en de theoretische studie (hoofdstuk 2), presenteert deze dissertatie de resultaten van vier empirische studies uitgevoerd met prototypes van elektronische game-gebaseerde omgevingen ontworpen voor de instructie en het inoefenen van Engels als tweede/vreemde taal. Deze prototypes werden voornamelijk in de context van het secundaire onderwijs geëvalueerd.

De eerste empirische studie (voorgesteld in hoofdstuk 3) verkent de rol van individuele verschillen met betrekking tot verschillende types van CF in een immersief 3D spel ontworpen voor het aanleren van pragmatiek in het Engels. De studie toont aan dat de leerlingen (*N*=83) expliciete CF zinvoller vonden voor hun leerproces dan meer impliciete CF, en dat het gepercipieerde nut van expliciete CF positief samenhing met de motivatie van leerlingen.

De tweede empirische studie (hoofdstuk 4) onderzocht hoe leerlingen (*N*=36) gebruik maakten van expliciete grammaticale CF in een geschreven interactief moordmysterie. De verzamelde data kon de hypothese niet bevestigen noch weerleggen dat het gepercipieerde nut van CF het eigenlijke gebruik van CF positief beïnvloedt; het gebruik van CF hing wel sterk af van de grammaticale voorkennis van leerlingen.

De derde en vierde empirische studies legden zich toe op de effectiviteit van het inoefenen van grammatica ondersteund door CF in mini-games. De derde studie (hoofdstuk 5) toonde aan dat 'levendige' CF (aangepast aan de fantasie voorgesteld in het spelconcept) de intrinsieke motivatie van leerlingen (*N*=32) positief beïnvloedde, en dat intrinsieke motivatie positief samenhing met hun bereidheid om verder te oefenen.

Het vierde empirische hoofdstuk (hoofdstuk 6) beschrijft de effecten van het inoefenen van grammatica op tweedetaalverwerving in een studie die twee maanden duurde, en waarin leerlingen één maand intensief oefenden (*N*=125; met een controlegroep van *N*=61). De resultaten tonen aan dat leerlingen door intensief grammatica in te oefenen, ondersteund door mini-games en CF, grammatica-kennis ontwikkelden die nuttig was voor hun prestaties op verschillende transfertaken (zowel nabije transfer als verre transfer). Bovendien waren de effecten van expliciete grammaticale CF, in het algemeen, sterker dan de effecten van CF die geen grammaticale uitleg bevatte; dit effect was niet merkbaar in een complexe gesproken taaltaak.

Het laatste hoofdstuk van deze dissertatie (hoofdstuk 7) vat de voornaamste resultaten samen, en brengt een aantal pistes naar voren voor toekomstig onderzoek. De resultaten van dit project zijn nuttig voor het ontwerpen van krachtige taalleeromgevingen ondersteund door technologie, die aansluiting zoeken met de ervaring van leerlingen met games of met hun interesse voor dit medium, en die bedoeld zijn om leerlingen te ondersteunen in het beheersen van de conventies van formele registers van een tweede of vreemde taal.

Chapter I

Introduction: the adventure of learning a language, and the role of red ink

> *We are advocating an approach designed to engender engagement through the utilization of students' digital-literacy expertise and/or gaming experience or interest, but we seek also to provide encouragement for the development of gaming environments that provide feedback at the level of linguistic form and exposure to and movement toward awareness, and eventually mastery, of a wide range of communication genres, including those associated most closely with traditional literacies and "power genres" text conventions. To achieve this, we advocate the use of a three point sequence when designing video games: genuine player need, linguistic support and creative feedback.*
>
> Ravi Purushotma, Steven L. Thorne, & Julian Wheatley, 2008

1.1 Setting the scene

London, 2014. It hasn't been long since you last visited the Natural History Museum, but the place continues to thrill you. Its collection of dinosaurs, Archie the giant squid, the life-size replica of the blue whale—all are worth a regular return. However, your reason for today's visit is different. As you make your way up the stairs towards the entry, the curator walks out, looking worried.

"Bad news, Inspector," he says, "we've been burgled."

" I heard. Give me the details."

"No money or valuable objects are gone at first sight, but the situation may be far worse than that," the curator answers.

"Continue."

"Someone broke into the belly of the blue whale. The trapdoor is open again, after being sealed for over seven decades. The burglar must have been a firm believer of the urban legends surrounding the whale. Must have hoped to find something of value in the belly—though I doubt there was much in there."

"How much people have the key to the Large Mammals Hall?"

"Not many," the curator says, "two guards of the security firm, Tom, our chief of technical services, and myself of course."

"Time to question the security guards then!"

The alarm on your smartphone wakes you to reality. You switch off your computer, slip your phone into your pocket, and leave the house. It's time to go to school.

1.2 Language learning adventures in red ink

In this day and age, digital games are all around us—at least in many parts of more economically developed and wired countries. We carry them on us in our pockets, drink coffee from mugs printed with gaming heroes and heroines, and our children grow up with games from a very young age.

Educators have long dreamed of tapping the power of games in order to further learning and instruction in formal education, and language learning is no exception. The adoption and use of 'off-the-shelf' (i.e. non-educationally purposed) games in language classrooms, as well as the design of digital games specifically for language instruction are—considering the history of the academic field of Computer-Assisted Language Learning—nothing new under the sun (Cornillie, Thorne, & Desmet, 2012). What is changing at an ever-quickening pace, however, is the ways in which gaming technologies are creating new opportunities for language learners to interact meaningfully and communicatively through a second or foreign language with other users of that language (learners themselves, or native speakers) as well as other cultures. Much more slowly—but surely—gaming technologies are becoming available to language teaching innovators who believe there is a place for game-based learning environments specifically engineered for the purpose of language instruction, focused on 'power genres' of language, i.e. those linguistic genres that are most commonly associated with traditional schooling and formalistic communicative practices.

If the use of games and gaming mechanics in language learning is not to be the next fad—or the revival of an old one—then they need to prove effective for language development. There is little doubt that when playing off-the-shelf games, in informal contexts, language users including many learners communicate and/or use language in effective ways. Certainly in online games, players orient themselves towards (collaboratively) completing goals, and are often required to use language to meet those non-linguistic objectives. Moreover, evidence is accruing which suggests that in online game-mediated contexts, language use and discursive practices can be highly complex (e.g. Steinkuehler & Duncan, 2008; Thorne, Fischer, & Lu, 2012), and that—contrary to concerns about the loss of standard language—playful violations of standardized linguistic conventions ('bad' spelling or grammar) as encountered in online, game-oriented language use, and in text messaging (e.g. Wood, Kemp, & Waldron, 2014), are not necessarily detrimental to learners' linguistic performance in more formal communicative genres.

Nonetheless, in games designed primarily for the instruction of linguistic power genres, or when off-the-shelf games are integrated in classroom language teaching, you as a playing language learner would not only have to demonstrate that you can use the language in order to communicate effectively and achieve non-linguistic objectives, but also that you master the power genre for the instruction of which the environment was built, in all its aspects. In instructional environments, deviations from the conventions of that power genre, such as inaccurate ('incorrect') grammar or phrases that are inappropriate in a given situation, would probably not pass unnoticed. The question then becomes how to deal with such deviations in the most effective way. When we consider this issue in the field of (research on) language teaching and learning, we enter into lively debates on how 'corrective feedback' should be given.

Let's go back to our learner hero, and—assuming our hero is a boy—to his adventure in London's Natural History Museum. He was probably speculating about what was inside the whale's belly. Given his experience as an investigator, he undoubtedly knew which places to inspect first in search of evidence, and which potential suspects to question. Perhaps he was truly interested in the mystery, driven by a *genuine need* to solve the case.

But did our hero notice that he made a mistake against the rules of English grammar? Did he register that the virtual, non-player character (i.e. the curator) corrected his sentence? Making mistakes in a second or foreign language is perfectly natural when we learn a new language—even in our native language—and we often need a great deal of *linguistic support* to 'recover' from the most stubborn of errors.

The curator gave our hero linguistic support in the form of a 'recast', i.e. a reformulation of ungrammatical speech in correct grammar. This technique to give corrective feedback is often used in naturalistic language learning environments, for instance when parents correct children's speech. Communicative interaction in the language classroom also favours such implicit forms of feedback, because the focus is on message, not on grammar, as is typically the case when we play games.

But for some aspects of language that need to be mastered, such linguistic support may not be enough, and our learner hero will make the same error again and again. For the most persistent of errors, he may need more feedback. Perhaps he needs to be told more explicitly that something is wrong with the string of words *much people*. Perhaps he needs to be told that *people* refers to something that can be counted, despite the lack of –*s* at the end of the word. Perhaps he needs to be told that things which can be counted go with *many*, not with *much*.

Further, our hero may need to be told that the imperatives in his earlier utterances sound somewhat blunt, and that there are more appropriate linguistic devices in this particular situation to request more information on the burglary.

However, while our learner hero was thinking about his blue whale mystery, focusing on message and not on code, such extensive corrective feedback may soon call up the experience of a traditional grammar lesson, or remind him of that assignment which he got back from his teacher earlier that day, smothered in red ink. Would he appreciate such feedback in a game, which, after all, emphasizes ludic engagement? Would he pay attention to it? Would it make him feel good about himself if it didn't help him, and if he made the same mistake over and over again? Would it still make his adventure in language learning a pleasurable experience?

Would he not prefer the language learning adventure without red ink?

On the other hand, when we fail in good games, we are often rewarded for doing so. We don't get any points, nor do we 'level up', but are confronted with *creative feedback*, i.e. situations that we had not envisioned, rendered in appealing and spectacular ways. The result is that we remain engaged, and may even want to explore all the different ways in which we can fail in a given situation. Educators know that failure can give rise to powerful learning experiences. Perhaps, creative feedback as typically encountered in digital games can teach language educators a few things about how corrective feedback can be made interesting and fun, so that our learner hero remains engaged when he fails to produce appropriate responses, and is willing to continue on his quest.

1.3 Purpose of the research project and intended outcomes

The quest of this PhD research project, then, is to address the effectiveness of corrective feedback in digital game-based language learning. This will be done with particular reference to the potential tension between ludic engagement in experience and instruction of linguistic form, or between the adventure of learning a new language and of using its power to accomplish non-linguistic purposes, and the red ink that necessarily crops up in language education.

With a view to contributing to understanding of how grammar teaching and different types of corrective feedback aid second or foreign language development in ludic and meaning-focused instructional environments, this project was guided by the "three point sequence" advocated by Purushotma, Thorne, and Wheatley (2008) (introduced above), namely *genuine player need*, *linguistic support* and *creative feedback*. Using this lens for design, it subsequently addresses three empirical foci which, as we will argue, need to be considered in research on the design of effective and playful learning spaces for the instruction of a second or foreign language: namely learners' *perceptions* of different types of corrective feedback as well as of themselves as receivers of such feedback, their use of corrective feedback, and the effectiveness of corrective feedback and gaming mechanics for supporting learner motivation as well as second or foreign language grammar learning. The necessity for investigating these three empirical foci will be dealt with in detail from a theoretical point of view in the second chapter of this dissertation.

The results of this project may be relevant for language educators, teacher educators, applied linguists, designers of educational games for language learning, policy makers, and anyone interested in second or foreign language learning supported by educational games.

1.4 Contextual requirements for this research project

In this research project, data collection was carried out by means of three different prototypes of game-based learning environments specifically adapted to and engineered for language instruction. As noted above, the design of these prototypes was inspired by the "three point sequence" promoted by Purushotma, Thorne, and Wheatley (2008). This involved the design and development of technology-enhanced learning environments aimed at engaging learners in meaningful language processing (by means of interactive narratives), which were capable of generating and providing feedback at the level of linguistic form (by means of database and natural language processing technologies), and which were capable of rendering creative feedback sequences.

Designing game-based environments specifically for the purpose of language instruction and based on this three-point sequence is no trivial undertaking. This PhD research project would not have been possible without external support in relation to, first, technology development, and secondly, human-computer interaction design.

Therefore, this project interacted intensively with research and development projects involving partners from the game development industry and research groups concerned with (educational) gaming. This enabled the exploitation of commercially-oriented game development technology for research purposes, as well as of open-source technologies to build prototypes of game-based learning experiences.

Likely, the reader will note that the designs of the learning environments used in this project have their flaws. A point in case is that language practice was—overall—largely receptive in nature, and that the learning environment which involved productive practice could be improved in many ways. Ideally, environments for language practice can elicit oral language production and provide feedback on oral production, but despite advancements in the field of human language technology (particularly automatic speech recognition), we are not quite there yet. Yet, this project would not have been possible, if it were not at the forefront of technological innovation in CALL.

A second contextual requirement is of a methodological nature, and concerns the inter- and transdisciplinarity of the research and development setting. The creation of game-like educational experiences and their evaluation in terms of learning and engagement necessitate thinking outside of the box, and require bringing together expertise in many disciplines. This project built bridges between the academic fields of second language acquisition, computer-assisted language learning (CALL; by itself a highly interdisciplinary undertaking), educational psychology and technology, motivational psychology, and game studies, and also involved a dash of human-computer interaction design. The reader of this dissertation may note that the interaction between these various disciplines varies throughout the manuscript, and that this thesis does not present one grand theory of the phenomenon under investigation, but rather an eclectic synthesis of theories concerned with human development and engagement that vary in scope, which were are all used as tools in an attempt to increase our understanding of the subject.

1.5 Structure of the dissertation

This dissertation comprises three main sections. The first section (chapter 2) reflects a theoretical study, presenting an in-depth account of the major issues in the research literature, and describes the architecture of the current research project. The second section is composed of four chapters (from chapter 3 to 6), each of which reports on an empirical study conducted in digital game-based language learning environments. Each empirical chapter is preceded by a brief interlude that is intended to remind the reader of the empirical study in the overall architecture of the research project. The final section (chapter 7) summarizes the main findings of the empirical studies, discusses the limitations of the project, and suggests directions for future research.

Chapter II

Theoretical study and architecture of the research project

2.1 Introduction

This PhD project is concerned with the effectiveness of corrective feedback in digital game-based language learning. More specifically, and in more technical terms, it deals with the effectiveness of *corrective feedback* for the development of second language grammar knowledge in *tutorial* computer-assisted learning environments that share elements of instructional design with *task-based language teaching* and *game-based learning*. This chapter provides an overview of the relevant literature in the fields of second language acquisition (SLA), computer-assisted language learning (CALL), educational psychology, and to some extent also in game studies, defines the problem, and explains how this research project tackles the problem.

Section 2.2 sets the scope of this project. First, it provides definitions for each of the technical terms used above and by situating this project within the field of CALL, and then by explaining how this project builds on previous research on digital game-based language learning. It concludes with why the effectiveness of corrective feedback was chosen as the object of study. Section 2.3 presents a review of relevant background research, and identifies key factors that are likely to determine the effectiveness of corrective feedback in digital game-based language learning. Section 2.4, finally, identifies the central research questions as well as three research focuses on the basis of the three key factors identified in the background research, i.e. learners' perceptions (of corrective feedback types and of themselves as receivers of feedback), their use of corrective feedback, and the effectiveness of CF for supporting learner motivation and second language grammar learning. The section concludes with the architecture of this research project and an overview of the four empirical studies that are presented in the next four chapters.

2.2 Scope of the research project

2.2.1 General definitions

In the SLA literature, the term *corrective feedback* (CF) refers to all responses to learners' utterances in a second or foreign language (L2) that are not well-formed or are inappropriate in a given situation (e.g. R. Ellis, Loewen, & Erlam, 2006). CF may include a mere indication that an error has been made (also known as *knowledge of results/outcome feedback* or *verification feedback* in the more general literature on educational psychology; Shute, 2008), it can present the correct form, it can offer metalinguistic information about the nature of the error (such as a grammar rule), or it may combine these three types of information. CF corresponds to the notion of *negative feedback* used in educational psychology and particularly in research on concept learning (Schachter, 1991)—the term 'negative feedback' itself being opposed to the notion of *positive feedback* (i.e. information that a particular outcome was successful). Whereas in the SLA literature both terms seem to be in use (i.e. 'corrective feedback' and 'negative feedback'), perhaps the more specialized term 'corrective feedback' has come into existence because L2 learning is concerned with more than concept learning, namely with "the learning of distinctions in the coding system, which mediates between the speech stream and conceptual representations of the environment" (Carroll & Swain, 1993, p. 359). Whatever the domain-specific implications of the terminology used, CF/negative feedback can be seen as a form of learning support, especially in cognitive approaches to human learning (Boero & Novarese, 2012), intended to make learners aware of the errors they make, and to help them understand the nature of their errors when it provides metalinguistic information or reminds of such information given earlier. As will be argued in sections 2.3.1 and 2.3.2, CF is a complex research construct—there are many types of CF, which may all have different effects on how a L2 is learned.

Tutorial Computer-Assisted Language Learning (tutorial CALL) refers to language learning supported by computer programs "that include an identifiable teaching presence specifically for improving some aspect of language proficiency" (Hubbard & Bradin Siskin, 2004), and to the research discipline that is concerned with the design, development and evaluation of such programs. The use of tutorial CALL programs in language learning and teaching is traditionally distinguished from the use of technologies not specifically dedicated to the learning and teaching of languages, but which may nonetheless be usefully implemented in language learning and teaching, such as word processors or social software (Levy, 1997). Tutorial CALL can also be distinguished from the latter technologies in that it typically includes consistent and immediate CF, provided at runtime without the intervention of a teacher.

Next, whereas the term 'tutorial CALL' is sometimes used in a very narrow sense, namely to refer to drill-and-practice programs such as those popular in the early days of CALL which are intended primarily for the development of 'explicit L2 knowledge' (e.g. knowledge of grammar rules; see section 2.3.2 for a more comprehensive discussion of this concept), the current research project is concerned with tutorial CALL that shares elements of instructional design with *task-based language teaching* (TBLT) (R. Ellis, 2003; Van den Branden, Bygate, & Norris, 2009). TBLT is a communicative approach to language instruction, which defines a *task* as a pedagogical activity devised for L2 learning that involves any of the four language skills (or a combination of skills) and in which learners must use the L2 meaningfully in an attempt to obtain a certain non-linguistic (communicative) outcome. Examples of such tasks are resolving whether a picture you are holding is different from or the same as your neighbour's picture, or finding the killer in a whodunit by asking and responding to questions. While language learners are working on tasks and trying to achieve a non-linguistic goal, their attention is primarily focused on meaning (*focus on meaning*). In TBLT, tasks are distinguished from *exercises,* i.e. pedagogical activities which do not result in a non-linguistic outcome, but which are intended to help learners develop understanding of a specific linguistic aspect. Exercises do not entail a primary focus on meaning, but have a strong *focus on form*.

It is crucial to note that tasks are not exclusively aligned with meaning focus, and that exercises are not necessarily only form-focused. As both R. Ellis (2003) and Van den Branden *et al* (2009) point out, a relative degree of focus on form characterizes many task-based activities: the point is that the focus on form is skilfully embedded within an activity that predominantly focuses on meaning-making; conversely, exercises focus first and foremost on form, but may also include a certain meaning focus, e.g. vocabulary exercises. This being said, the distinction between tasks (primarily meaning-focused) and exercises (mainly form-focused) is key in TBLT (R. Ellis, 2003, pp. 2–9), and we will revisit its relevance for L2 learning in section 2.3.2.

We define *task-based tutorial CALL*, then, as tutorial CALL software that satisfies the following criterial features of authentic and complex language learning tasks as identified by R. Ellis (2003, pp. 9–10): namely [1] a workplan which [2] is intended to call primarily for meaning-focused language use, [3] which has a clearly defined non-linguistic outcome, [4] which engages cognitive processes such as reasoning and evaluating information, and [5] which involves any of the four language skills (or a combination thereof). We do not seek to comply with Ellis' sixth and final criterial feature, namely the involvement of processes that reflect those of real-world communicative language use, as human-computer interaction is at this point hardly capable of accurately simulating real-world communication. As a side-note, we refer readers interested in similar projects in task-based tutorial CALL (such as the spy-game *Spion* for German L2 learners) to Schulze (2010).

Game-based learning refers to the use of specific games or gaming features to support learning and teaching. Scholars in game-based learning have long distinguished between the pedagogical use of so-called *commercial off-the-shelf (COTS) games* (Van Eck, 2009) in learning and teaching contexts, and the design and evaluation of games specifically developed for instructional purposes (i.e. tutorial games). In what follows, we will use the term 'game-based' mainly to refer to the latter, namely to tutorial learning environments that use elements of game design, such as feedback, so as to enable and support learning (for an introduction of the notion of game design elements see section 2.2.2). This use of the term 'game-based' is consistent with Reinhardt & Sykes' (2012) distinction between game-based language learning and *game-enhanced* language learning (i.e. the use and implementation of vernacular, non-educationally purposed games in L2 teaching and learning).

The current research project is confined to tutorial CALL that is both task- and game-based. In the pedagogical literature, gaming has often been associated with TBLT (Baltra, 1990; García-Carbonell, Rising, Montero, & Watts, 2001; Mawer & Stanley, 2011; Purushotma et al., 2008; Taylor, 1990), and a recent survey among 50 experts in CALL, e-learning, and serious gaming confirms that there is a clear role for games that cater to task-based and primarily meaning-focused L2 use (De Grove, Cornillie, Mechant, & Van Looy, 2013). Hence, one of the key challenges in this project was to design tutorial CALL environments that would create opportunities for meaningful and task-based L2 use.

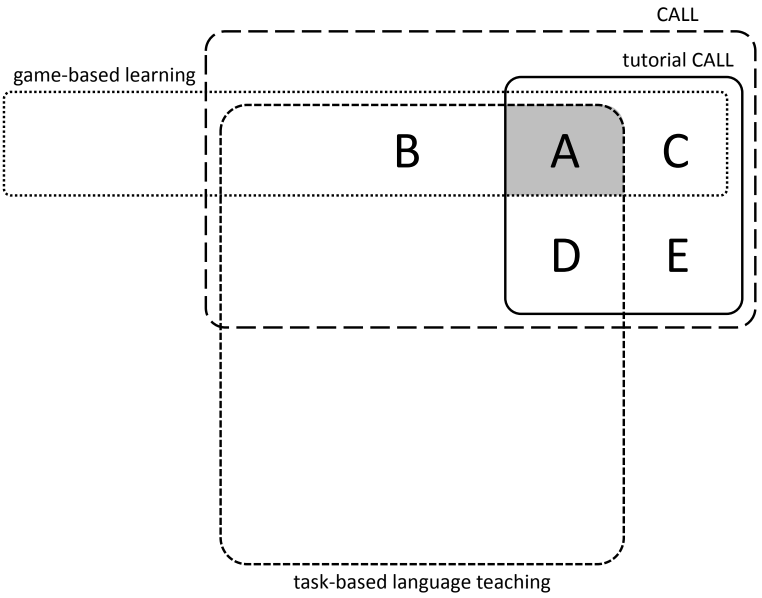


Figure II-1: situation of our research project within the field of CALL

Figure II-1 situates the current research project in the field of CALL. Our focus on task- and game-based tutorial CALL (zone A in the figure) (e.g. Johnson, 2007; Neville, Shelton, & McInnis, 2009) makes this project related to, but different from, research on language learning in games not specifically designed for language learning (i.e. COTS games) (zone B) (e.g. deHaan, Reed, & Kuwada, 2010; Thorne, Fischer, & Lu, 2012). Secondly, our research focus excludes game-based language learning environments which are not task-based, such as *mini-games* (i.e. games that are limited in scope) that are not embedded in meaning-focused tasks but that are focused solely on form (zone C) (for discussion see Cornillie & Desmet, forthcoming). Finally, our focus excludes task-based tutorial CALL that is related to certain aspects of gaming, but that are not game-based (zone D). This comprises research that is mainly situated in the field at the intersection of tutorial CALL, natural language processing, and artificial intelligence, such as research on L2 learning in interaction with conversation agents (chatbots) (e.g. Petersen, 2010; Zacharski, 2003). Finally, our project is related to but separate from research on 'classical' drill-and-practice style tutorial CALL (zone E).

A final element that confines the focus of the current research project is the mode of interaction in the technology-supported learning environment. Ideally, game-like tutorial CALL activities can be developed that revolve around interaction in spoken language (e.g. Ehsani & Knodt, 1998; Engwall & Bälter, 2007; Johnson, 2007; Strik, Cornillie, Colpaert, van Doremalen, & Cucchiarini, 2009). However, given the technological challenges associated with processing learner speech, generating and providing appropriate and useful feedback on speech, and implementing speech-recognition-based technologies in authentic classrooms, our research project makes use of designed environments in which learners interact primarily through the written word.

2.2.2 Previous research on digital game-based language learning

Research into digital game-based language learning (DGBLL)—i.e. the area of CALL that is concerned with the use of gaming environments for language learning and teaching, and which includes the design and evaluation of game-7based tutorial environments (zones A and C in Figure II-1) as well as the evaluation of COTS games implemented in L2 learning settings (zone B)—has been around since the early days of CALL, but arguably the related empirical research is still in its infancy (Cornillie, Thorne, et al., 2012). This gap in the empirical research was already noted more than a decade ago by Philip Hubbard, a pioneering theorist, researcher and practitioner in CALL. In 2002, he stated that "the majority of [the previous] research focused on demonstrating the validity of the general approach rather than specific elements of its implementation (Gale, 1989), and that is still the case with more recent studies" (Hubbard, 2002, p. 211). Even though Hubbard's statement applies only to a specific genre of tutorial CALL games (i.e. interactive participatory dramas), it can be extended to tutorial CALL gaming and to CALL research on gaming in general. To date, very little is known about what specific affordances of games actually contribute to language learning and teaching, and where and how the research on gaming departs from other areas in CALL and SLA. In line with calls for more theoretically rigid and methodologically sound empirical research on games in the broader domain of educational technology (Tobias, Fletcher, Dai, & Wind, 2011), Hubbard's claim may be interpreted as a call for investigating the added instructional value of what could be referred to as *game design elements*, i.e. features that are part and parcel of game design, such as conflict, fantasy, assessment, and feedback (e.g. Bedwell, Pavlas, Heyne, Lazzara, & Salas, 2012). Investigating game-based language learning environments in terms of the added value of their features—rather than making wholesale comparisons between games and other instructional media—has the benefit that research can start from and contribute to questions, methodologies and findings in other areas of CALL research. This approach has been adopted most notably in e.g. research into multimedia CALL (Chapelle, 1998).

The need for researchers to consider specific aspects and elements of game-based learning environments rather than investigating games as holistic entities was also emphasized in recent reviews of the more general research on game-based learning (Bedwell et al., 2012; Vandercruysse, Vandewaetere, & Clarebout, 2012). In addition to noting a discrepancy between the gaming elements listed in the theoretical literature and the ones used in the empirical research literature, the review of Vandercruysse, Vandewaetere, & Clarebout (2012) touches on the difficulty of identifying and describing such elements exhaustively.

While acknowledging the challenges involved in the construction of a coherent and comprehensive framework on gaming elements, the current research project will focus on one specific feature of game-based learning environments, namely feedback. We choose feedback because it is an indispensable feature of game design (e.g. Aldrich, 2005; Becker, 2007; McGonigal, 2011; Prensky, 2001; Rigby & Ryan, 2011; Salen & Zimmerman, 2004), and since it has been shown developmentally useful in a wide range of SLA research (e.g. Aljaafreh & Lantolf, 1994; R. Ellis et al., 2006; Long, 2007; Ranta & Lyster, 2007; Schulze, 2003; Sheen, 2010) as well as in the more general educational research (e.g. Black & Wiliam, 1998; Hattie & Timperley, 2007).

The need for researching feedback as an element in tutorial CALL games also arises from a selection of empirical findings in the CALL gaming literature. First, the research on COTS games has focused, among other issues, on interactivity and on the use of linguistic materials for supporting L2 play. In a recent study, deHaan, Reed, & Kuwada (2010) found that the interactivity in a commercially available English language rhythm video game induced cognitive load and reduced vocabulary recall. Given that both cognitive theories of instructed SLA (e.g. Gass & Mackey, 2008) and multimedia learning theories (e.g. Mayer, 2001) have stressed the beneficial nature of interaction—in which feedback evidently has a crucial role to play—this raises the question as to which forms of game interactivity might support, rather than hinder, L2 development. Since corrective feedback is explicitly intended to provide support in language learning tasks, it may be a feature of such a beneficial form of interactivity—Sims (1997) reserves the word 'support interactivity' for the latter. Next, two studies with COTS games present evidence that learners' use of supportive materials during L2 play results in higher vocabulary learning gains (Miller & Hegelheimer, 2006; Ranalli, 2008). This suggests that supporting learning processes during the meaningful activity of play, for instance through feedback that focuses on linguistic issues, may be effective. Finally, in comparison with the research on COTS games, the research on tutorial CALL games is scant, but one study in this area is particularly relevant to the current research project. In a study with a game designed for the instruction of Spanish L2 pragmatics, Sykes (2009) found little improvement in learners' pragmatic strategy use on post-tests, and suggested that the integration of different types of feedback, ranging from implicit to explicit, is crucial for improving the effectiveness of CALL games, because such feedback might help learners to notice differences between their interlanguage and the L2. In a nutshell, first findings from the empirical research on DGBLL suggest that there might be some positive role for feedback as a form of interactive and developmentally useful support in tutorial CALL games.

In the next section, we will theorize how CF could be effective in DGBLL. Given that the focus of this research is on CF in task-based tutorial CALL (see above), the conceptual framework will borrow mainly from the literature on CF in instructed SLA. In addition, considering the specific nature of feedback in digital gaming environments, the framework will also be informed by the literature that intersects educational psychology, motivational psychology, and—to some extent—game studies.

2.3 Background research

In the SLA literature, research on CF constitutes a substantial body of work, potentially because the findings of CF research have strong implications for practice: knowing when, how, and which errors to correct could help teachers make informed and effective decisions about error correction in the classroom (Pica, 1994). This assumption applies even more to the design of tutorial CALL software, in which CF can be given more consistently than in classroom L2 teaching and learning, and potentially *ad infinitum*.

Three recent meta-analyses show that CF has significant and durable effects on L2 development. Lyster & Saito (2010) found a medium-sized mean effect (Cohen's d = .74) on immediate post-tests drawn from 15 studies on oral CF in classroom learning. Russell & Spada's (2006) meta-analysis includes the results from 15 studies on CF in grammar learning tasks (both oral and written, but excluding computer-assisted learning), and reports a large mean effect size for CF as measured by immediate post-tests (Cohen's d = 1.16). Li's (2010) meta-analysis reports a medium-sized mean effect of CF on immediate post-tests (Cohen's d between .61 and .64 depending on how the meta-analysis was carried out), based on the results of 33 primary studies. All three meta-analyses show that the effects observed on immediate post-tests did not decline significantly on delayed post-tests. Mackey & Goo (2007) also examined the effects of CF in a meta-analysis on L2 interaction research, but failed to isolate the contribution of CF from the effects of classroom interaction as a whole.

In addition, there is some evidence that providing CF with high consistency may enhance its effectiveness. Mackey & Goo's (2007) meta-analysis of interaction, while admittedly not focusing exclusively on CF, revealed much larger effect sizes on immediate post-tests for laboratory studies (Cohen's d = .96) than for classroom-based studies (d = .57). Along similar lines, Li (2010) reports significantly larger effects for laboratory studies. Taking into account that in laboratory environments, CF is given with high consistency in comparison with more naturalistic classroom learning, this is promising for tutorial CALL, in which CF can equally be given in highly consistent ways.

These meta-analyses suggest that CF can be quite powerful for L2 development and that the developmental effects of CF can be durable, and that there is a particularly promising role for consistent CF in tutorial CALL settings.

These findings are in line with the more general educational research on feedback. In this domain, meta-analyses have reported mean effect sizes between .74 and 1.13 for 'feedback about the task', a construct very similar to CF in SLA (Hattie & Timperley, 2007; Hattie, 2009).

While CF can be very powerful in general, current meta-analyses also point at great variability between empirical studies. This suggests that CF may be differentially effective contingent upon a host of factors such as the type and timing of CF, the linguistic focus of CF, how the effects of feedback on L2 development are measured, length of the treatment, the nature of the instructional setting (naturalistic vs. laboratory), and individual differences between learners. In the remainder of this section, we will argue that effectiveness research on feedback in DGBLL needs to address the complex interplay between the following variables: 1) the type of CF offered (section 2.3.1); 2) the instruments used to measure L2 development that may be attributed to instruction with CF (section 2.3.2); and 3) individual difference factors (section 2.3.3), more particularly learners' perceptions of CF and their self-perceptions. The final part of this section (2.3.4) will identify affordances of game-based feedback for learning, which may open up new avenues for effectiveness research on CF as well as for L2 pedagogy.

2.3.1 Type of CF

A first variable that might explain differences in the effectiveness of CF is the type of CF. As to the effects of this variable, the results of current meta-analytic research are inconsistent and inconclusive, due in great part to the fuzziness of CF typology (i.e. to the many different ways in which CF has been operationalized) (Long, 2007; Lyster & Saito, 2010; Mackey & Goo, 2007). One notable result, however, is that in oral interaction, 'prompts' (i.e. utterances which signal the error but do not provide the correct form, including metalinguistic clues) are significantly more effective than recasts (i.e. correct reformulations of erroneous utterances) (Lyster & Saito, 2010). In addition, singular experimental studies show that CF types including metalinguistic explanation are more beneficial than CF in the form of recasts (Carroll & Swain, 1993; R. Ellis et al., 2006) and than CF that simply informs the learner that an error was committed (Carroll & Swain, 1993). Next, there is some evidence that explicit recasts are more effective than implicit recasts in terms of noticing (Philp, 2003; Sachs & Suh, 2007) and post-test performance (Loewen & Philp, 2006).

Hence, by and large, this research suggests that 'explicit' CF, possibly accompanied by metalinguistic instruction, is likely to facilitate L2 development more than 'implicit' CF. This is in line with Norris' and Ortega's (2000) seminal meta-analysis on the effects of implicit versus explicit instruction. However, as noted above, a fundamental problem with CF research is that comparison of CF types is methodologically challenging, because quite different CF types are often conflated under the categories 'explicit' or 'implicit'. Recasts, for instance, are generally treated as instances of implicit CF, because they focus both on form and on meaning (Lyster, 1998), whereas they can be very explicit. Hence, for future empirical research, it is crucial that 'explicit CF' and 'implicit CF' are carefully defined.

A good starting point for the current research project is Lyster & Saito's (2010) two-dimensional model of CF types in oral, classroom-based interaction, which disentangles the degree of explicitness of CF (i.e. its perceptual salience) from its function, which may either be *output-prompting* (signalling the error but not giving away the correct form) or *input-providing* (providing a correct reformulation or the correct answer) (see Figure II-2). To this model, we have added the general types of information given in CF (knowledge of results, correct answer, metalinguistic information), as provided in R. Ellis *et al.*'s (2006) definition. Further, because Lyster & Saito's (2010) model focuses on CF in oral, classroom-based interaction, we have added the corresponding CF types for tutorial CALL (i.e. human-computer interaction), based on Heift (2004).

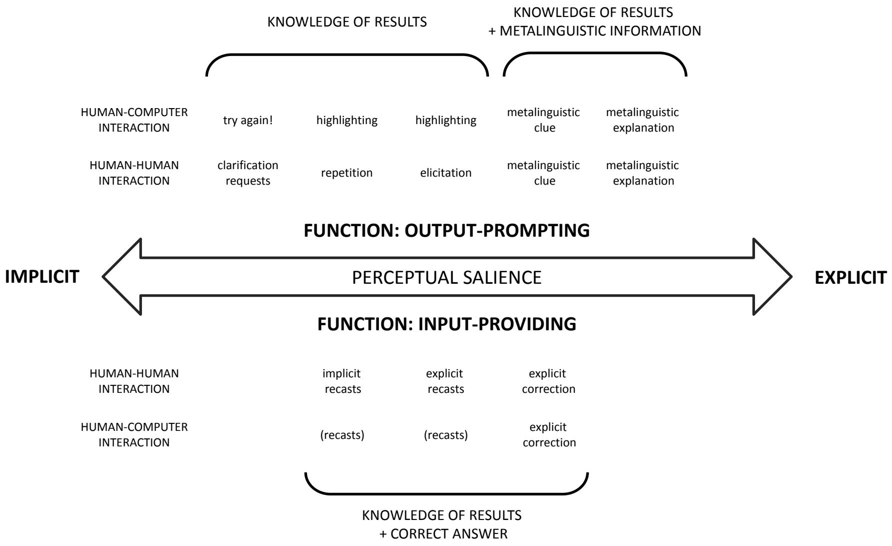


Figure II-2: typology of CF (based on R. Ellis et al., 2006; Heift, 2004; Lyster & Saito, 2010)

Due to the medium, however, CF in human-computer interaction is not equivalent to CF in human-human interaction (including such interaction mediated by computer technology). Generally speaking, the typology of CF in human-computer interaction is less nuanced than the types of CF in face-to-face L2 learning environments. Some CF types afforded by face-to-face settings do not have identical counterparts in human-computer interaction, such as clarification requests, repetitions, and elicitations. Further, more communicative types of CF, such as recasts, are especially difficult to realize in human-computer interaction, because of technological challenges (however, see Morton & Jack, 2005). Next, it could also be argued that output-prompting feedback in human-computer interaction, in contrast with face-to-face learning, is almost invariably explicit in terms of perceptual salience, as is clear in the case of the "try again!" CF type. Finally, it has to be noted that computer-mediated communication, a specific type of human-human interaction but mediated by technology, makes it possible to visually enhance specific types of feedback, such as written recasts (Sachs & Suh, 2007).

The distinction between output-prompting CF and input-providing CF is theoretically important for the discipline of SLA, as the former type provides only negative evidence about the L2, whereas the latter provides both negative and positive evidence (see Table II-1). Positive and negative evidence refer to information about which utterances and linguistic analyses are possible (grammatical) and impossible (ungrammatical) in the L2. The prototypical example of negative evidence is CF that includes a grammar rule with examples of ungrammatical constructions in the L2.

Table II-1: evidence provided in input-providing vs. output-prompting CF

	input-providing CF	**output-prompting CF**
positive evidence	yes	no
negative evidence	yes	yes

One notable position in the SLA literature, which is predominantly informed by theories based on Universal Grammar (UG), is that positive evidence alone is sufficient to trigger language acquisition (both in L1 and L2 acquisition). This strand of research starts from the observation that children normally achieve perfect mastery of their L1 (Carroll, 1995; Gregg, 2001; Schwartz, 1993), whereas negative evidence (grammar explanations and CF) is largely absent in the environment of L1 learners—however, see e.g. Farrar (1992) for counterevidence that implicit CF such as recasting is available to and useful for L1 learners. Hence, such research tries to explain how the same innate linguistic universals presumed to be at work in L1 acquisition—i.e. a special 'Language Acquisition Device', triggered by abundant amounts of positive evidence—can help L2 learners to avoid making previous errors, in particular errors due to overgeneralization. Thus, in UG-based theories of SLA, negative evidence is irrelevant (Carroll, 1995). If it can be shown, however, that positive evidence alone is not enough for L2 learning—which seems plausible in light of the research which shows that errors may persist (or 'fossilize') in L2 learners' speech even when these learners have been long immersed in settings that provide abundant positive evidence (e.g. Ranta & Lyster, 2007)—and that explicit and extensive negative evidence (such as CF comprising rule explanation) is sufficient and necessary for triggering changes in a learner's interlanguage, then this might constitute proof that L2 acquisition is somehow different from L1 acquisition. Hence, the difference between input-providing CF (involving both negative and positive evidence) and output-prompting CF (involving negative evidence alone) is crucial for theory building in the field of SLA.

Lastly, the type of CF, more particularly the distinction between input-providing CF and output-prompting CF, is also relevant from a game design perspective. It has been noted that in COTS games, feedback rarely gives away "the 'answer'" (Becker, 2007, p. 35). This may be explained by the observation that the design of 'good' COTS games favours a learning paradigm that is popularly known as *experiential learning* (García-Carbonell et al., 2001; Kiili, 2005) or, to use a more technical term, as *constructivism*. This view of the human mind stresses the capacity of the brain as a pattern-recognizer, and learning is seen as a sequence of being exposed to concrete experiences (i.e. 'data'), making reflective observations, construing mental generalizations and hypotheses about the experience, and testing these hypotheses through active experimentation (Kiili, 2005). In the most extreme version of this view on learning, this process even unfolds without much awareness—interestingly, parallels have been drawn between such a view of learning in games and L1 acquisition (Koster, 2005), which is also assumed to rely to a great extent on implicit learning (see section 2.3.2 for a discussion of this concept).

The key point here is not that learning in games is implicit—for output-prompting CF invites learners to explicitly analyse linguistic forms (e.g. N. C. Ellis, 2005a)—but rather that games put the individual's agency centre stage: "exploration and experimentation are actively supported in most good games" (Becker, 2007, p. 41). Moreover, COTS games typically take the experiential learning approach to the extreme, and as a result, rather than giving 'correct answers' away in feedback, they stimulate players to find solutions themselves by means of trial and error, through motivating game design elements such as *positive failure feedback* (see section 2.3.4), and by gradually increasing the level of learning support on an as-needed basis. The assumption underlying this design principle seems to be that—to borrow a phrase from Socio-Cultural

Theory of SLA—“over-assistance [i.e. giving too much feedback] decreases the student’s [player’s] agentive capacity” (Lantolf & Thorne, 2006, p. 277). Hence, if instructional designers of game-based language learning want to adhere to the experiential learning design of COTS games—which may be necessary to maintain the learner’s agentive capacity—it seems advisable to apply input-providing CF sparingly in digital game-based language learning, or at least to only offer it after output-prompting feedback types have been exhausted.

Output-prompting feedback types, on the other hand, may be crucial in experiential, game-based language learning. First, explicit indications of failure (i.e. ‘knowledge of results’)—which is a core aspect of the design of good COTS games (see section 2.4)—may help to make learners aware that a specific hypothesis in their interlanguage is erroneous. Such awareness will likely prompt learners to make an effort to retrieve a correct form or metalinguistic knowledge (a grammar rule) from memory, or perhaps induce a rule themselves—the latter strategy resonates well with active experimentation and reasoning. Learners who are unable to (re-)produce such a rule and who have the necessary knowledge to deal with metalinguistic information may be helped with explicit feedback in the form of metalinguistic clues and, one step further, with metalinguistic explanations.

In the game-based learning literature, there is no consensus on the availability of explicit meta-information in the actual design of COTS games. Gee (2007) notes that “overt verbal information is offered—and often lots of it—‘just in time’ (when it is needed and can be used) or ‘on demand’ (when the player is ready for it and knows why it is needed)” (p. 156). Reichle (2012), on the other hand, observes a “relative scarcity of commercial games that actively cultivate or demand a meta-awareness of game rules” (p. 147), and Juul (2013) notes that “many games do not communicate [the causes of our failure] directly” (p. 52). There is, however, empirical evidence that in the broader ecologies of COTS games, players “spend many hours of statistical, logical, and strategic analysis outside of actually playing the game [i.e. in briefing and de-briefing sessions]” (Rigby & Ryan, 2011, p. 9). They overcome problems by engaging in highly scientific reasoning (Steinkuehler & Duncan, 2008), they generate linguistically complex meta-texts on wiki-pages (Thorne et al., 2012), and even in-game episodes of expert-learner feedback focused on problematic L2 usage have been recorded (Thorne, 2008).

To summarize, in digital game-based language learning, active experimentation with the L2, and intensive interaction with a range of explicit and particularly output-prompting feedback types (without or with cultivation of metalinguistic awareness), may likely have a significant impact on L2 development. In light of the available evidence on the differential effects of CF types, pushing learners through explicit CF to revise their hypotheses about the L2 and to modify their output may be more effective than providing them with CF that 'gives away' target language forms (Lyster & Saito, 2010; Ranta & Lyster, 2007, pp. 152–153). Therefore, this project focuses mainly on feedback types that are explicit in terms of perceptual salience and that are output-prompting. In the next section, we will discuss the relevance of CF for learning processes, more particularly the types of L2 knowledge which CF is thought to develop.

2.3.2 CF, the development of explicit and implicit L2 knowledge, and transfer

Next to differences between CF types, the variability in effect sizes of experimental studies on CF may also be due to differences with respect to how 'learning' was measured on post-tests. In other words, the type of outcome measure may be a moderating variable in studies on the effectiveness of CF. Russell and Spada (2006) did not examine this variable in their meta-analysis on CF, but pointed to its importance by referring to Norris and Ortega's (2000) meta-analysis on the effects of L2 instruction, who "report that effect sizes were greater in studies that used more controlled, test-like outcome measures and that smaller effects were observed on both free response and grammaticality judgment outcome measures" (p. 155-156). Li's (2010) meta-analysis on CF did not find statistically significant differences between different types of outcome measures. Lyster & Saito's (2010) meta-analysis on oral CF, however, found significantly larger effect sizes on post-tests involving free constructed-response measures than on post-tests that required learners to produce metalinguistic judgments. The implications of this finding will be discussed later in this section, but for now it suffices to note that SLA research has by and large measured 'learning' either as the development of explicit L2 knowledge, or as the development of implicit L2 knowledge, and that the effects of instruction (including feedback) may be smaller or larger depending on the type of knowledge that was measured.

Explicit L2 knowledge (also known as *metalinguistic* or *declarative knowledge*) is defined as knowledge about an L2 that is available to awareness and can be verbalized (Dörnyei, 2009, p. 143; R. Ellis, 2004), such as the knowledge of lexical units, linguistic concepts and grammar rules. The construct *implicit L2 knowledge* is less transparent, and is currently most accurately characterized "simply as knowledge that is not explicit" (Dörnyei, 2009, p. 143). Such knowledge is considered "tacit and intuitive" and independent of the knowledge of linguistic rules (R. Ellis, 2009a, p. 11). Implicit L2 knowledge is also considered to be involved in on-line/automatic parsing and productive L2 skills: a key characteristic is that it can be rapidly accessed, whereas the retrieval of explicit L2 knowledge is considered to be more effortful and slower (Dörnyei, 2009).

The dichotomy between explicit and implicit L2 knowledge is especially important in psychologically oriented strands of SLA research. The distinction is of crucial theoretical importance to mainstream SLA. The explicit-implicit knowledge dichotomy is related to an assumption that underpins a great deal of the SLA literature, namely that there is a fundamental distinction between *acquisition* or *implicit learning* on the one hand and *(explicit) learning* on the other hand. Implicit learning can be defined as "learning without awareness of what is being learned" (DeKeyser, 2005, p. 314), and involves unconscious processes of abstraction (of principles or rules), drawing on exemplar-based input-processing, as in first language acquisition. Explicit learning, on the contrary, is learning that does involve awareness: “input processing with the conscious intention to find out whether the input information contains regularities and, if so, to work out the concepts and rules with which these regularities can be captured” (Hulstijn, 2005, p. 131).

Explicit and implicit learning processes are often related to the kind of knowledge that they produce, namely explicit knowledge and implicit knowledge, respectively (Hulstijn, 2005, pp. 131–132), but one of the most challenging problems in SLA theory concerns the question whether explicit knowledge acquired through explicit learning can contribute to—or ‘convert into’—implicit knowledge needed for on-line parsing and productive L2 skills. This question is also known as the “interface question”, which is at the heart of current SLA theory (N. C. Ellis, 2005a, p. 307), and has important implications for classroom practice when we consider that L2 pedagogies rely to some extent on explicit L2 knowledge such as metalinguistic terminology and explanations. The so-called *non-interface position*, as maintained by e.g. Krashen’s (1981) Monitor Theory or accounts of SLA based on Universal Grammar (e.g. White, 2008), claims that explicit and implicit L2 learning processes are completely separated, and that explicit knowledge has no role to play in acquisition/implicit language learning. Conversely, theories based on the assumption that there is an interface between explicit and implicit knowledge attribute some positive role to conscious processes and explicit knowledge for the development of implicit knowledge. Depending on the assumed strength of this interface, the latter theories sometimes labelled either as either weak-interface or strong-interface theories of SLA, and include e.g. connectionist (e.g. N. C. Ellis, 2005b) and interactionist accounts on SLA (e.g. Gass & Mackey, 2008), but arguably the prime example of an interface theory is Skill Acquisition Theory.

Skill Acquisition Theory (SAT) (DeKeyser, 2008) is a strong-interface theory of language learning based on a general theory of cognition known under the acronym ACT-R (Adaptive Control of Thought-Rational; see Anderson et al., 2004 for a comprehensive review). In a nutshell, SAT proposes that declarative knowledge about an L2 assists the acquisition of implicit knowledge through the subsequent processes of proceduralization and automatization. Proceduralization is gradual, and requires a high degree of awareness in the early stages of the acquisition of a specific rule, and, in later stages, continued practice in order to automatize the acquired rule. The result of this process is procedural (implicit) knowledge which learners can access without much effort and speedily during complex, communicative tasks. SAT thus posits a strong interface between explicit and implicit learning processes, and implies a general decrease of conscious and focused attention (i.e. from more explicit learning processes to more implicit learning). Practice and feedback are considered important instructional catalysts of automatization, as they aid in reducing the error rate and in increasing the speed with which learners respond (DeKeyser, 2001, pp. 145–146). Hence, increased accuracy and shorter response times are seen as measures of learners' development of implicit L2 knowledge (R. Ellis et al., 2009).

CF is not only an essential instructional feature from the specific perspective of SAT, it also occupies a central position in the interface debate in the SLA literature. As CF is essentially language about language, it is inherently metalinguistic in nature (Birdsong, 1989; Carroll, 2001). To be perceived as feedback on previous performance, a learner needs to interpret CF as such (e.g. "my teacher tells me that something was wrong with the grammar of what I have just said"). Such interpretation and thinking has been hypothesized to "take place in that part of the functional architecture dedicated to inferencing, thinking and the construction of mental models of the on-going discourse" (Carroll, 1995, p. 76).

Hence, the processing of CF requires at least a minimal degree of consciousness, namely awareness at the level of noticing (Schmidt, 1990), and entails the involvement of explicit learning processes. Weak-interface theories of L2 development attribute a positive role to CF that helps learners to 'notice the gap' between their production and the input. From such a point of view, recasts are particularly important, as they provide both negative and positive evidence (for overviews of the research on recasts from a weak-interface perspective, see Long, 2007, chapter 4, and Goo & Mackey, 2013).

Further, CF that only provides negative evidence (i.e. output-prompting CF; see above) is thought to result primarily in explicit L2 knowledge, viz. knowledge *that* (knowledge of results) and/or *why* (metalinguistic knowledge) a particular utterance is not possible in the L2 (Schwartz, 1993). The latter implies a form of consciousness that transcends noticing, namely awareness at the level of understanding (Schmidt, 1990). The processing of such CF involves highly explicit learning processes and is likely to result first and foremost in explicit L2 knowledge. Therefore, such CF is mainly relevant in strong-interface views of SLA, and is considered less useful in weak-interface theories of SLA, which attach great importance to the provision of positive evidence in CF.

Against the backdrop of this theoretical framework, Lyster and Saito's (2010) meta-analysis on oral CF is particularly noteworthy. As noted above, they found significantly higher effects on post-tests that may be assumed to tap primarily into learners' implicit L2 knowledge (i.e. free constructed-response measures) than on post-tests that favour explicit L2 knowledge (viz. metalinguistic judgments). This suggests that CF in oral tasks facilitates especially the development of implicit L2 knowledge—not explicit L2 knowledge, in contrast to what SLA theory on CF predicts (see above). In explaining their findings, the researchers pointed at the similarity between the conditions of the instructional tasks and the conditions of the post-tests, which is claimed to facilitate transfer of the knowledge acquired through instruction with CF to the post-test. Or, in other words, the instructional tasks inculcated transfer-appropriate L2 processing (Lightbown, 2008).

The issue of *transfer* is an important one, both from the perspective of SAT and in game-based learning environments. SAT claims that the implicit knowledge that is implied in the development of L2 skills through proceduralization and automatization is highly skill-specific as well as dependent on other conditions of the task (DeKeyser, 2007b). For instance, L2 learners who have developed implicit knowledge in writing tasks may not be able to apply this knowledge in speaking tasks. Or, learners who master a particular grammatical aspect in form-focused written exercises may not be able to use this knowledge in more complex (authentic) written tasks in which the primary focus is on meaning, because they also have to express personal meaning. This is why, in TBLT, the distinction between *tasks* and *exercises* (see section 2.2.1) is key. Exercises may help to develop knowledge that is useful for L2 development, but because learners may exclusively focus on form in such exercises, the processing conditions may be different from the processing conditions of genuine and highly contextualized communicative tasks.

In game-based learning environments, transfer of knowledge gained during in-game activities to tasks performed outside the game setting, such as language classrooms, might be impeded for multiple reasons. First, the ways in which language and language use are mediated through game genres and interfaces (deHaan, 2005; Jordan, 1992) may create conditions that are too different from out-of-game L2 use, which may hinder learners in realizing transfer. For instance, the way in which a role-playing game gives form to communication between the player's character and non-player, computer-directed characters through 'point-and-click dialogues' hardly resembles communicative L2 use, and players who are good at performing such dialogues in games may not be able to use this knowledge in genuine communicative contexts, including collaboration with other (human) players in an online game. Secondly, games are typically based on self-referential ontologies (Hubbard, 1991, p. 221; Phillips, 1987, p. 276), and are often detached from 'real life'—Gee writes that games set up specific 'semiotic domains' (2003). This also may hinder transfer. Third, the specific mechanics and sensorimotor operations involved in play may create specific types of cognitive load (deHaan et al., 2010) that are absent in genuine communicative contexts. All these factors involved in gaming constitute task-specific conditions that may result in the development of specialized skills, and, hence, may hinder transfer to 'real-life' communicative tasks.

Reichle (2012) presents a counterargument, arguing, from the perspective of SAT, for research into how language acquisition and the development of skills used in playing video games (i.e. mastering the mechanics of gameplay) are coupled through proceduralization. He claims that "computer games that maximize opportunities for implicit learning could potentially lead to more nativelike processing" (p. 151). It seems that the games that qualify as such are particularly immersive games that focus on real language use in communicative tasks, rather than mini-games in which there is a strong focus on form. However, the problem with such immersive games, from the perspective of transfer, is that language is highly contextualized. And the more contextualized language becomes, the more different it becomes from other contexts of use, thus potentially hampering transfer.

Abstract knowledge, however, may help to bridge the differences between the conditions of tasks: "knowledge that is overly contextualized can reduce transfer; abstract representations of knowledge can help promote transfer" (Bransford, Brown, & Cocking, 1999, p. 41, as cited in DeKeyser, 2007, p. 6). SAT thus offers a role to CF comprising metalinguistic explanation. As such CF potentially helps learners to develop abstract knowledge about linguistic constructions, it may enable them to realize transfer from one task to another. The question is, however, whether learners will be receptive to such CF in digital game-based language learning (see section 2.3.3.1).

In sum, these arguments call for the consideration of task conditions involved in gameplay, particularly the kind of L2 processing that is involved (focus on meaning, focus on form, or both), and how and which type of CF is given. These factors are likely to determine whether explicit or implicit L2 knowledge will be developed and how this knowledge will transfer to other language learning tasks.

2.3.3 CF and individual differences: the mediating role of learner perceptions

In sections 2.3.1 and 2.3.2, we argued on the basis of meta-analytic research that the type of CF and how learning is measured are two key variables in the effectiveness of CF. A third set of variables that might explain the mixed results of effectiveness research on CF comprises various learner characteristics, usually called *individual differences* (IDs) in the academic literature. ID factors may explain why feedback does not always reinforce behaviour, and may shed light on how and why learners accept, modify, or reject feedback (Hattie & Timperley, 2007; Kulhavy, 1977). Particularly, IDs could "influence learners' receptivity to error correction and thus [mediate] the effectiveness of the feedback" (Sheen, 2011, p. 129).

In SLA, three broad categories of IDs have been identified, related to three psychological systems that are thought to interact—and are often difficult to separate—namely cognition (including meta-cognition), motivation, and affect/emotion (Dörnyei, 2009). Despite pressing claims to investigate CF effectiveness in relation with ID factors (e.g. DeKeyser, 1993), the empirical research was rather scant until quite recently, as revealed by Russell & Spada's (2006) meta-analysis. To date, there is evidence from primary studies that language analytic ability (Sheen, 2007, 2011), anxiety (DeKeyser, 1993; Havranek & Cesnik, 2001; Sheen, 2008, 2011), attitudes towards CF (Havranek & Cesnik, 2001; Sheen, 2007, 2011), and prior extrinsic motivation (DeKeyser, 1993) somehow explain differences in CF effectiveness. However, taking into account the complex nature of CF (see section 2.3.1 above), it seems crucial for theory construction to consider ID factors in relation with the type of CF, rather than with CF as a whole, as "it is quite possible that different types of CF are mediated differentially by different individual factors" (Sheen, 2011, p. 130).

The current research project focuses on how the effectiveness of CF types is mediated by ID factors that are related to learners' perceptions, namely the perceived usefulness of these CF types, and learners' perceptions of themselves as goal-directed individuals. Both variables instantiate in the learners' interaction with his or her learning environment and are, hence, specific to particular circumstances and therefore unstable. This poses considerable methodological challenges (Dörnyei, 2009), which will be taken up in sections 2.4.2 and 2.4.3 of this chapter, which outlines the overall architecture and methodology of this research project. The former variable, perceived usefulness, is meta-cognitive in nature (Luyten, Lowyck, & Tuerlinckx, 2001); the latter factor, comprising learners' perceptions about themselves, seems closer to what is commonly understood as 'motivation'.

2.3.3.1 Perceived usefulness of CF as a predictor of CF use

Perceived usefulness (PU) of instructional interventions, such as CF, is a variable that is associated with meta-cognition, and is formed as the result of the interaction between the instructional environment and the learners' *instructional knowledge*, i.e. "students' conceptions about the relationship between instructional interventions and learning" (Luyten et al., 2001, p. 205). PU constitutes a *belief* (e.g. "Feedback aids learning"), and is intricately related to, but somewhat different from, an *attitude* (e.g. "I want the teacher to correct all errors"), in that the former has stronger factual support whereas the latter is more related to affect (Dörnyei, 2005; Sheen, 2011). The construct of PU originates in expectancy-value theory (Fishbein & Ajzen, 1975) and is the most central variable in Technology Acceptance Model (TAM) (e.g. Davis, 1989; Venkatesh, Morris, Davis, & Davis, 2003), which posits that users' behaviour (i.e. their use of technology) can be predicted to a significant extent by how useful they find a particular system.

Although the TAM is a general model of technology adoption, it has also been applied in educational technology contexts, as it fits in nicely with a perspective on learning and instruction known as the Cognitive Mediational Paradigm (Winne, 1987). This paradigm presupposes that learners' perceptions of instructional design features (such as CF) mediate learning processes, such as learners' use of and attention to such features during learning tasks, and may as such affect learning outcomes: "an assumption common to many contemporary theories of learning and instruction [is] that learners' perceptions of tasks and cues mediate forms of engagement and, in turn, affect performance" (D. L. Butler & Winne, 1995, p. 253). Evidently, learners' use of CF is a *conditio sine qua non* for it to be effective—as Black & William note in their review of feedback in classroom learning, "for assessment to be formative the feedback information has to be used" (1998, p. 16)—so perceived usefulness of CF is likely to mediate its effectiveness.

In instructed L2 environments, there is ample evidence that language learners find CF generally helpful in a wide range of tasks and settings (including tutorial CALL) (e.g. Chenoweth, Day, Chun, & Luppescu, 1983; Enginarlar, 1993; Havranek & Cesnik, 2001; Nagata, 1993; Radecki & Swales, 1988; Schulz, 2001). Also, learners' *preferences*, a construct that seems to sit somewhere between beliefs and attitudes, have been examined: learners prefer feedback that contains metalinguistic explanations rather than less detailed 'knowledge of results' feedback (Nagata, 1993) and recasts (correct reformulations of erroneous utterances) (Kim & Mathes, 2001). In addition, some research indicates that learners would like to be corrected more than their teachers think is good for them, particularly in communicative settings, where the focus is on meaningful and fluent interaction in the L2 (Magilow, 1999; Raimes, 1991; Schulz, 2001).

The consistent finding that learners generally value CF and that they prefer detailed metalinguistic CF raises issues for the design of effective game-based CALL. As noted in section 2.2.1, the use of games and the application of gaming principles in language learning and teaching contexts is mainly guided by the communicative approach of TBLT. In TBLT, a focus on language as meaningful communication through complex tasks precedes a focus on isolated language forms (R. Ellis, 2003). Consequently, in some forms of TBLT (e.g. Bullard, 1990; Willis & Willis, 2007), the provision of extensive forms of CF including metalinguistic explanations, which entails an explicit and analytical focus on linguistic form, is typically deferred until the post-task/debriefing phase, so as to invoke meaningful use of the L2 and to "prevent students from 'regurgitating' pre-selected expressions and grammatical structures" during communicative interaction (Jager, 2009, p. 200). Or, in R. Ellis' (2003, p. 8) terms, during tasks, learners need to *use* rather than *display* language. During communicative, meaning-focused tasks, focus-on-form is often realized implicitly and without analytical rigour, e.g. through more implicit recasts (Lyster & Ranta, 1997). Hence, one might say that the objective of the communicative approach is to create conditions that favour learning processes which would result in the development of implicit L2 knowledge rather than explicit L2 knowledge. This explains why teachers who adopt a communicative approach usually abstain from giving elaborate and metalinguistic CF during authentic and communicative tasks.

One might argue that if, in general, explicit and extensive (metalinguistic) CF is more effective and perceived more useful than more implicit types of CF, this will also be the case in game-based learning environments. However, this cannot be presupposed: task conditions may have a strong impact on learner beliefs. Considering the links between game-based and task-based learning on the level of attention to meaningful language use, learners might think that explicit and elaborate CF is not useful in game-like environments, especially if such CF is given immediately, as the primary focus is on meaningful interaction in the L2. Or, if learners perceive of the task mainly as an activity intended for amusement, rather than as a learning activity, they might be unwilling to accept extensive CF, or they may not notice its availability, or might refrain from using it because it feels like cheating. Or, if CF does not help learners to achieve in-game objectives and does not improve their performance on those objectives, then it might not be considered useful. These questions relate to the notion of *calibration* (D. L. Butler & Winne, 1995; Winne, 2004): for instruction to be effective, learners' perceptions of instructional cues such as CF need to be calibrated with respect to the actual benefit such cues bring. This warrants the investigation of the perceived usefulness of CF types in game-like learning environments, as this ID factor might mediate the effectiveness of CF through learners' receptivity to and actual use of CF.

2.3.3.2 Learners' self-perceptions: perceived competence and immersion as components of intrinsic motivation

In a review of the effects of formative feedback on learning, Black & William (1998) emphasize that not only formal properties and cognition-related aspects of feedback matter, but that its effectiveness depends equally on "the broader context of assumptions about the *motivations and self-perceptions* of students within which it [formative feedback] occurs [emphasis added]" (p. 17). At first sight, the assumption that the developmental effects of feedback are in some way mediated by the ID factor 'motivation' is not different from the assumption that other instructional features or instruction in general are mediated by 'motivation'. But the situation for feedback is specific, as feedback may itself determine how learners orient themselves towards goals. For, feedback is not only intended to increase a learner's understanding through cognitive processes, but also to influence their handling of the task through affective processes, "such as increased effort, motivation, or engagement" (Hattie & Timperley, 2007, p. 82). Serious games designer Marc Prensky writes that "games are good at [negative feedback] because they give players the motivation to keep trying" (2001, p. 159). Or conversely, feedback may have unintended side-effects on learners' 'motivation'. As Hattie & Yates (2014) phrase it: "[Learners] are sensitive to the climate under which criticism is given. Often, what a teacher intends as helpful critical feedback turns to personal ego evaluation in the eyes of the receiver" (p. 65). Or, too much feedback may—to borrow that phrase from Lantolf & Thorne (2006) again—"[reduce] the student's agentive capacity" (p. 277). Computer-generated feedback may exacerbate this issue, since technology is not (yet) capable of making educated guesses about learners' mental and emotional states—if it will ever be.

In what follows, we will theorize how the effects of CF on L2 development in game-based learning could be mediated by *intrinsic motivation* as seen from the perspective of Self-Determination Theory (SDT) (Ryan & Deci, 2000). In the SLA literature, various views have been proposed on what drives L2 learners, each with different definitions and operationalizations of the complex construct of 'motivation' (see Dörnyei, 2003, 2005 and Ushioda & Dörnyei, 2009 for overviews). The current research project focuses on intrinsic motivation: an orientation/type of motivation that refers to behaviour which is performed because these are inherently interesting or enjoyable and not because they lead to a separable outcome (i.e. an outcome not related to the content of the task, such as an extrinsic reward). This project capitalizes on the work done on intrinsic motivation in the paradigm of SDT for the following reasons: it is one of the most advanced and influential theories of human motivation and has been applied in many contexts, including instructional settings; the theory has received some support from SLA researchers (Noels, Pelletier, Clément, & Vallerand, 2000); its instruments have proven reliable and valid; and it has recently been applied to gaming contexts. However, before we zero in on the relation between SDT (with a focus on intrinsic motivation), gaming and CF, we will first take a broader view on the SLA research that deals with CF and learners' self-perceptions.

In SLA research on CF, learners' self-perceptions have been primarily operationalized as *(language) anxiety*. Some studies consider anxiety as a more or less stable learner characteristic that determines the extent to which learners will benefit from CF (DeKeyser, 1993; Havranek & Cesnik, 2001; Sheen, 2008). A different perspective on the mediating role of anxiety is one that is more situated: it views anxiety as the result rather than the cause of poor language learning, and conjectures that CF may itself induce anxiety and subsequently could influence learners' receptivity to correction. Truscott (1996), for instance, argues that CF has harmful effects on L2 development because it induces stress and reduces learners' self-confidence, and that, even if learners find CF helpful (see section above), such anxiety results in ineffective learning. Krashen (1998) shares this position, and hypothesizes that CF raises an *affective filter*, which is supposed to block acquisition. Sheen (2011) found some support for this hypothesis. In this study, oral CF (in particular oral CF that contained metalinguistic information) provoked language anxiety and had a significant negative effect on learners' post-test scores, whereas this effect was not observed for written CF types. This indicates that "the medium of the CF affects whether anxiety plays a mediating role" (Sheen, 2011, p. 150): in oral classroom settings, CF may threaten a learner's face whereas this may be less so in the context of written CF, which is more private.

In tutorial CALL settings, CF constitutes a human-computer interaction, and is hence private and less face-threatening. Since anxiety seems to have been conceptualized as a more socially determined individual difference (i.e. as the inhibition to speak in the L2 in front of an entire class), it may not come into play. Still, computer-mediated CF is not 'neutral'. It is reasonable to believe that learners may experience typical tutorial 'knowledge of results' CF such as "Wrong!" or "No!" as a form of "mild social punishment" (i.e. negative reinforcement) (Schulze, 2003, p. 442), and that such seemingly harmless messages could have equally devastating effects on the motivation of learners. In accordance with perception theory, learners' perceptions may intensify "salient cues which are evaluated as negative" (G. L. Robinson, 1991, p. 192), such as immediate and abundant tutorial CF. The result of this is that learners experience failure at a particular task more intensely—as Hattie & Yates (2014) put it in their recent chapter on feedback: "bad is stronger than good"—and that they remember the feeling of failure longer than the actual content of the task.

As CF is an indication of error, learners may perceive of it mainly as a measure of their performance, rather than as an opportunity to learn. If learners focus on this aspect of the CF, abundant and explicit CF may undermine *perceived competence*, a construct very closely related to the notion of *self-efficacy* (Bandura, 1997). As specified by self-determination theory (SDT), perceived competence is one of the three main constructs involved in intrinsic motivation, which is considered the ideal type of motivation: "a natural wellspring of learning and achievement that can be systematically catalyzed or undermined by parent and teacher practices" (Ryan & Deci, 2000, p. 55). To date, there is consistent evidence that feedback which directs attention away from the task to the self, such as normative feedback (e.g. grades) undermines intrinsic motivation (R. Butler, 1987; Deci, Koestner, & Ryan, 1999) as well as actual performance (Kluger & Denisi, 1996). So, if learners interpret tutorial CF in game-based environments mainly as normative, it might harm their self-esteem or perceptions of competence, and hence affect their intrinsic motivation. Or, conversely, if learners find in such CF "information that is useful and nonjudgmental" (Rigby & Ryan, 2011, p. 19), it may motivate them to work through that information, solve challenging in-game problems, and enhance their (perceptions of) competence. Wu (2003), for instance, found that L2 learners developed competence if they got instructional support and evaluation that emphasized self-improvement on L2 tasks which were moderately challenging, which in turn instilled high levels of intrinsic motivation.

Next to perceived competence, the research conducted within SDT points to one additional factor associated with learners' (self-)perceptions that seems relevant in game-based learning environments, namely *(perceived) immersion.* Recently, SDT has been applied to the study of intrinsic motivation in gaming environments, including game-based learning (Rigby & Przybylski, 2009; Rigby & Ryan, 2011; Ryan, Rigby, & Przybylski, 2006). This has yielded an extension of the original model that is particularly interesting for the study of CF in game-like environments. The main tenet of SDT is that humans are most intrinsically motivated when three basic needs are simultaneously satisfied: the need for feeling competent (*competence need*), the need for choosing and acting autonomously (*autonomy need*), and the need for interacting with others (*relatedness need*). Because of the specificity of digital gaming environments, the extension of this model of need satisfaction towards games (coined Player Experience of Need Satisfaction; PENS) includes two additional factors, viz. intuitive controls, which relates to the perceived ease of use of games, and (perceived) immersion.

The latter construct, *(perceived) immersion* (or *presence*), refers to the sense of 'feeling there' in virtual environments, and covers physical presence (related to the perceptual fidelity of the simulation), emotional presence (associated with the interplay between emotions and the potential to take action upon these), and narrative presence (the extent to which players feel they are part of a story) (Rigby & Ryan, 2011). High immersion may have cognitive benefits, as it is the prerequisite for creating situated experiences that might help learners recall learning problems and achieve transfer (e.g. Mantovani & Castelnuovo, 2003). In addition, we argue that the construct of perceived immersion might shed light on the utility of CF in educational games. For, if learners are immersed in playful experience, they might find the provision of immediate CF, in particular metalinguistic commentaries, disruptive. If they do, then they could disregard the feedback, or even abandon the learning experience altogether.

To summarize this subsection, we argue that studies into the utility of language-focused feedback (i.e. CF) in game-based CALL would benefit from taking into account two ID variables that are associated with learners' perceptions of themselves as individuals who engage in game-like activities, viz. perceived competence and perceived immersion. To our knowledge, neither of these variables has yet been included in effectiveness research on CF in the SLA literature—Sheen's (2011) investigation of the effects of CF on language anxiety comes closest. What is more, although perceived competence/self-efficacy has been examined in SLA settings (e.g. Noels, 2001; Noels et al., 2000; Xiaoli Wu, Lowyck, Sercu, & Elen, 2012; Xinyi Wu, 2003), including game-based settings (Zheng, Young, Brewer, & Wagner, 2009), immersion constitutes a more novel area for research (for theoretical discussions see Schwienhorst, 2002; deHaan, 2008), and this factor seems particularly relevant for computer-mediated learning environments that simulate (authentic) L2 tasks, such as fully immersive games. Both perceived competence and immersion are relevant for this research because they are alleged to be components of intrinsic motivation in games, which in its turn is considered a predictor of the time and effort players spend on gaming. If that motivational power can be harnessed to create volitional and self-sustainable educational experiences, learners' may spend more time interacting with L2 environments, in turn increasing the frequency of specific linguistic constructions in the input, which has been shown to be a significant catalyst of L2 development (N. C. Ellis, 2002).

2.3.4 Affordances of game-based feedback for learning

So far, the theoretical framework of this study has focused on the interplay between corrective feedback, L2 development and individual differences as theorized in the literature on SLA, educational psychology, and motivational psychology. In addition to these bodies of research, one might argue that the investigation of CF in game-like learning environments could benefit from a review of how feedback is conceptualized within the relatively young but emerging discipline of game studies. While this literature deals primarily with COTS games, rather than with educational ones, it might inspire the design of effective educational environments for two reasons. First, good COTS games are considered to be built on sound learning principles (e.g. Aldrich, 2005; Becker, 2007; Gee, 2007). And secondly, COTS games are considered more successful and motivating than educational games (Papert, 1998; Van Eck, 2006), which may be associated with certain design features. An understanding of these features—particularly feedback as an element of game design—and how they sustain motivation, might help to make educational game environments more effective. One caveat must be kept in mind, though. As the literature on feedback in (COTS) games is rather limited and empirical evidence is only just emerging, the following review is largely exploratory.

We already pointed out (see section 2.3.1) that feedback in COTS games rarely gives away correct responses (Becker, 2007), and we discussed this observation in terms of experiential learning, connectionism, and the difference between input-providing and output-prompting CF in SLA. In addition, the literature on game-based learning describes that good games give both positive and negative feedback, and that feedback in games should be abundant and vivid.

In its most simple form, *positive feedback* lets the player know that he or she has reached a certain goal. This has two functions. First, it serves to reinforce learning (Becker, 2007) or—using a connectionist phrase—is assumed to strengthen the connection between nodes in a network of knowledge. Secondly, as will be elaborated further in this section, positive feedback aims to reward players for achievement or for having mastered goals (Becker, 2007). *Negative feedback* is also present in games, and is intended to stimulate “learning from mistakes” (Prensky, 2001, p. 158). Moreover, negative feedback “ties in to the idea of fairness”, namely that the player’s chance of winning the game is equal to the chance of losing it (Becker, 2007, p. 25). So, in its most essential form, negative feedback is about failure. However, as we will see, failure feedback is special, because games “give players the motivation to keep trying” (Prensky, 2001, p. 159).

Further, feedback in games is abundant. Because of the affordances of technology, feedback may be given consistently, and can occur at several levels of the game design. First, “in almost all games [feedback] is immediate” (Prensky, 2001, p. 121). Immediate feedback informs the player of his or her performance on a moment-to-moment basis. Immediate feedback is *granular*, as it maps to an individual action of the player (Rigby & Ryan, 2011, p. 23). In addition, granular feedback can be complemented with *sustained* feedback, such as meters and multipliers which inform the player that he or she is acting consistently within a certain stretch of gameplay, with *cumulative* feedback, which shows more permanent growth (e.g. through total score or increasing levels) (Rigby & Ryan, 2011, pp. 24–25). Such longer-range feedback is sometimes also referred to as an ‘outcome’ (Prensky, 2001, p. 121). As a result of the abundance and high frequency of feedback, digital games can create dense experiences that satisfy competence need (see section 2.3.3.2 for a discussion of this concept related to motivation). Therefore, positive feedback is also known as *competence feedback* or *mastery feedback*: “good games are almost always built around a constant stream of mastery feedback, giving players information about their success and rewarding that success meaningfully by increasing their abilities and strength to conquer the even greater challenges ahead” (Rigby & Ryan, 2011, pp. 11–12).

Finally, in good game designs, feedback is *vivid*. Vividness is a concept that is rooted in research on virtual reality (more particularly in Telepresence Theory), and refers to “the ability of a technology to produce a sensorially rich mediated environment" (Steuer, 1992). This characteristic applies both to positive and negative feedback. Vivid feedback transcends the pure functional level of informing players whether they succeeded or failed, and comes in excessive and visceral forms. Positive feedback that is vivid is often called *excessive* or *juicy feedback*, as it consists of "tons of cascading action and response for minimal user input [... which makes] the player feel powerful and in control of the world, and it coaches them through the rules of the game by constantly letting them know on a per-interaction basis how they are doing" (game designer Kyle Gabler cited in Juul, 2010, p. 45). Thus, positive feedback not only has a cognitive orientation, but is equally intended to deliver a pleasurable experience.

What is particularly interesting about game design is that negative feedback can also be vivid (next to positive feedback). Negative feedback in games occurs in *failure states*. These are phases in the game in which the player fails, followed by some kind of explicit message from the system that indicates the failure (i.e. negative feedback). Game designers recognize that players fail repeatedly in their attempts to master a game and consequently spend a lot of time in failure states. Hence, the challenge of game design is to not just deliver that bit of information (failure at a task), but to make failure states/negative feedback sensational. Designers work hard on failure states, as they want players to seek out failure and to find the negative feedback in those states a) interesting, so that they understand why they failed and can overcome problems, and b) vivid, so that the frustration of failure is lowered (McGonigal, 2011; Prensky, 2001; Purushotma et al., 2008; Swink, 2006). When well designed, failure feedback can make players feel in control of the world represented in the game, and can elicit positive emotional responses (for empirical evidence see Ravaja, Saari, Salminen, Laarni, & Kallinen, 2006). This may motivate players to persevere. Game designer and critic Jane McGonigal (2011) describes *positive failure feedback* as "a *vivid* demonstration of the players' agency in the game … [it] reinforces our sense of control over the game's outcome. And a feeling of control in a goal-oriented environment can create a powerful drive to succeed" (pp. 66-67; emphasis added).

One of the cornerstones of motivational game design, thus, is to make negative feedback vivid. This can be accomplished in a number of ways. According to Telepresence Theory, vividness can vary in terms of breadth, i.e. the number of senses involved simultaneously, and in terms of depth, which constitutes a more qualitative dimension (Steuer, 1992). In games, the breadth of negative feedback can be increased by including visual, auditory, and haptic information simultaneously. Further, the qualitative dimension of depth can be achieved by making (negative) feedback depend on the 'fantasy' or alternative reality depicted in the game (Malone, 1981), or by including narrative elements in negative feedback. Simulation specialist Clark Aldrich (2005) sees feedback in failure states as "an opportunity to wrap a story around the situation", the function of which is to make the experience more immersive (p.25). Finally, variations of particular negative feedback messages can be foreseen, so that each time a player fails, the feedback message is different and interesting: "it's not just one canned animation playing back every time" (Swink, 2006, p. 12). The result of making negative feedback vivid is that it can induce positive emotional responses, even if the feedback was meant to communicate failure. Because the representational context of a game is typically not that of the real world—a salient characteristic of a game is that, in contrast with a simulation, it "creates its own world" (Hubbard, 1991, p. 221)—the negative effects of failure can be minimized. As Laurel (1993) put it: "the distinguishing characteristic of the emotions we feel in a representational context is that there is *no threat of pain or harm in the real world* [emphasis in original]" (p. 114).

Serious game designer Prensky (2001) writes that "designing feedback to be less learninglike and more gamelike is often a big paradigm shift and challenge for Digital Game-Based Learning designers" (p. 159). The question is, however, whether the effort is worthwhile and whether such feedback is actually more effective. From this review of the game studies literature that discusses feedback, it may be hypothesized that the application of abundant positive feedback and vivid negative feedback in educational environments could have positive effects on learners' intrinsic motivation. More specifically, abundant positive feedback may satisfy the need for competence, whereas vivid feedback adapted to the game's theme or fantasy can create a sense of immersion.

2.4 Problem statement, and architecture of the project

In this section, we first state the problem to be addressed in this research project, on the basis of the literature review presented above, and formulate four central research questions. Then, we give an outline of the architecture of the project, comprising the four studies that were conducted in this project. We conclude with an overview of the four studies.

2.4.1 Problem statement and central research questions

Marc Prensky, one of the main players in the resurge of interest in digital game-based learning in the last two decades, states that it is essentially "from the feedback in a game that learning takes place" (2001, p. 121). The current research in SLA and educational psychology, however, suggests that the picture is complex, and that the effectiveness of corrective feedback in digital game-based language learning may depend on the interplay between the following factors: the type of CF, how L2 development is measured, and individual differences related to learners' receptivity to (and actual use of) CF.

First, the literature suggests that explicit prompting after errors, possibly accompanied by extended metalinguistic explanation, is likely to be more effective than more implicit feedback messages and feedback that simply includes the correct answer. Further, the degree to which CF is effective in terms of promoting L2 development may be measured in different ways: empirical studies may measure a) whether it results in explicit L2 knowledge (declarative knowledge about the language, such as knowledge of rules) or also (and ideally) in implicit L2 knowledge; b) whether it helps to proceduralize explicit L2 knowledge; and c) whether knowledge gained on in-game tasks transfers to tasks outside the game. Finally, learners' individual receptivity to CF may determine whether feedback will be effective in game-like learning environments. Learners need to find CF (types) useful for their L2 development, and they need to actually use it. And, in order to lead to sustained and intrinsic motivation to engage in L2 learning in games, CF may not harm learners' perceptions of competence or immersion. Feedback types which are vivid and adapted to the game fantasy may increase intrinsic motivation, which is likely to determine learners' willingness to practise, or their actual time spent in the learning environment. Therefore, we hypothesize that CF will be effective if it results in both cognitive and motivational learning outcomes.

The current research project, then, addresses the following four central research questions:

RQ 1. How useful do learners find CF in digital game-based language learning?

RQ 2. How does the perceived usefulness of metalinguistic CF in digital game-based language learning explain the actual use of such CF?

RQ 3. How does vivid CF affect learners' intrinsic motivation and their willingness to practise in digital game-based language learning?

RQ 4. How does continued practice with CF in digital game-based language learning assist learners in developing L2 grammar knowledge?

Figure II-3 shows the conceptual framework for this research project (i.e. the main research constructs and their hypothesized interrelations), as well as the situation of the four central research questions in the conceptual framework.

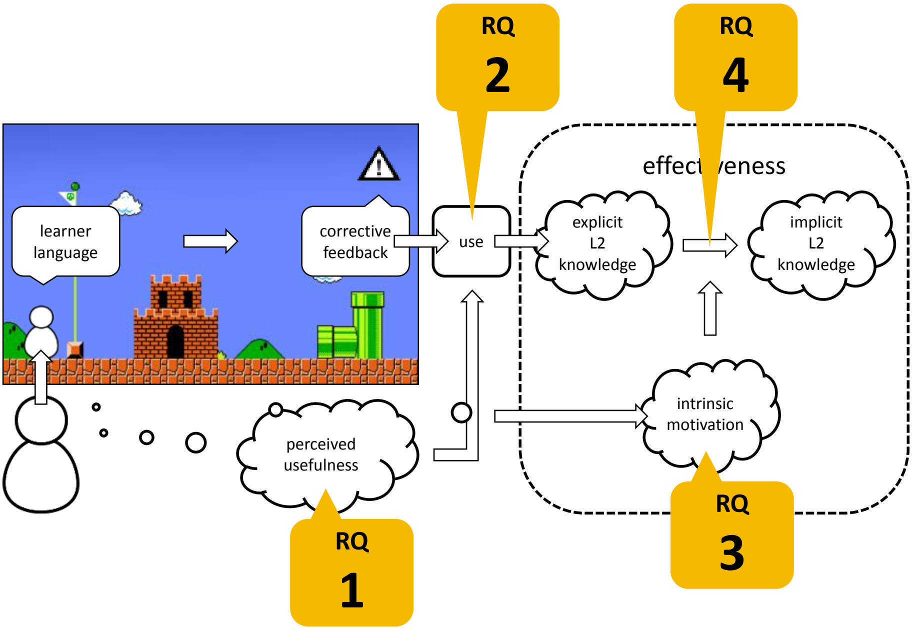


Figure II-3: conceptual framework, with central research questions (RQs)

2.4.2 Architecture of the project

In this research project, each of the four central research questions presented above is addressed in a separate empirical study. The project also comprises three different empirical foci (see Figure II-4), namely learners' perceptions (of CF and of themselves) (investigated in study 1), learners' use of CF (vis-à-vis perceived usefulness) (study 2), and effectiveness of CF on two levels, namely with respect to self-perceptions/intrinsic motivation (study 3) and with respect to L2 grammar learning (study 4).

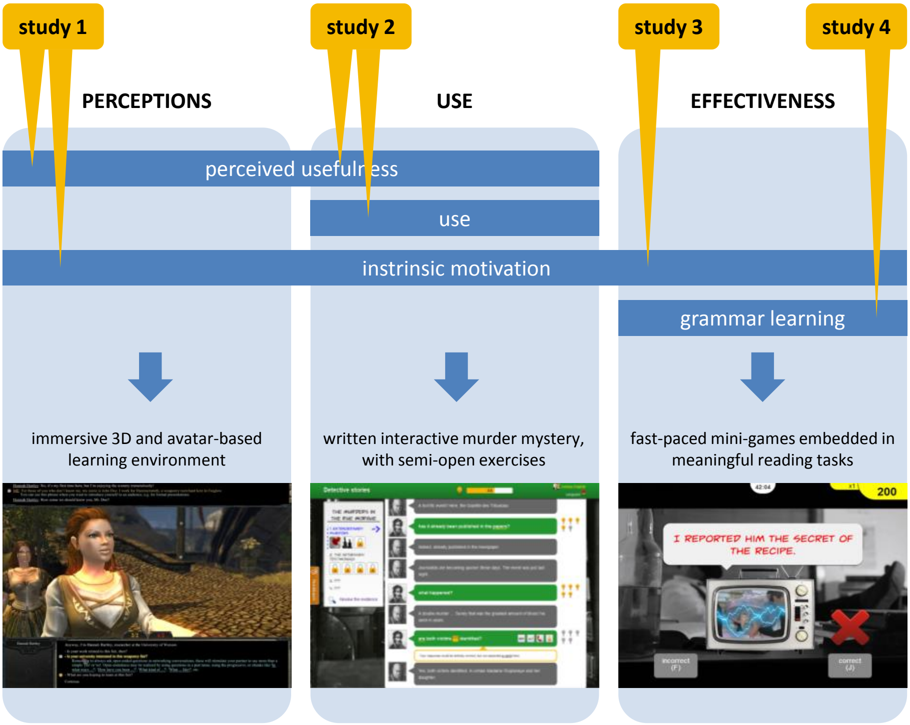

study 1
study 2
study 3
study 4
PERCEPTIONS
USE
EFFECTIVENESS
perceived usefulness
use
instrinsic motivation
grammar learning
immersive 3D and avatar-based learning environment
written interactive murder mystery, with semi-open exercises
fast-paced mini-games embedded in meaningful reading tasks
Detective stories
THE MURDERS IN THE RUE MORGUE
42:04
200
I REPORTED HIM THE SECRET OF THE RECIPE.
incorrect (F)
correct (J)

Figure II-4: architecture of the research project

For each of these three foci, a different prototype of a digital game-based language learning environment was designed and developed. Van Eck (2006) writes that investigations into the effectiveness of digital game-based learning need to recognize that, as a medium, games are highly diverse, and that as a result “not all games will be equally effective at all levels of learning” (p. 22). For the same reason, and because the current research focuses on three different—though related—issues of CF in game-like environments, three different technological environments are used to gather empirical data for the specific issue under investigation. In study 1, we use a fully immersive 3D-based game in order to map learners’ perceptions. In study 2, data were elicited on the use of CF by means of a written interactive murder mystery. Studies 3 and 4 use mini-games in order to investigate the effects of CF on intrinsic motivation and grammar learning.

The general research methodology is cumulative. This project first investigates perceptions; subsequently, it relates perceptions to usage of CF; in the third empirical focus, it relates perceptions to the impact of CF on the development of L2 grammar knowledge. Where possible and relevant, findings of earlier studies are taken into account in later iterations and investigations.

A final general characteristic of this Ph.D. research project is that it relies on various methods. As the project is devoted to a considerable extent to individual differences, which are considered to be non-stable, particular to specific circumstances (i.e. dependent on the environment), and interrelated (Dörnyei, 2009), the research design is based both on quantitative and qualitative methods in order to get a comprehensive picture of the ID variables involved.

2.4.3 Overview of the studies

In this section, we provide an outline of the four studies in function of the three empirical foci identified above. Wherever relevant, we also formulate secondary research questions for each of the four main research questions that were formulated in the previous section.

2.4.3.1 Empirical focus 1: learners' perceptions (study 1)

The first focal issue concerns learners' perceptions, more specifically their perceptions of CF and their self-perceptions, and the relations between self-perceptions and perceptions of CF. These variables are investigated in detail in the first empirical study, which intends to answer the following main research question:

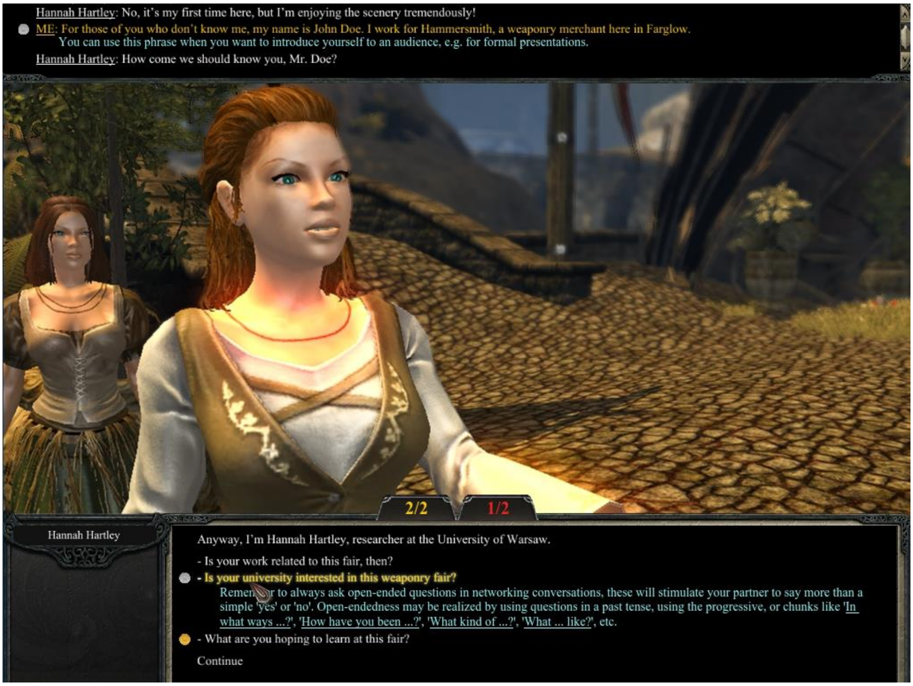
RQ 1. How useful do learners find CF in digital game-based language learning?

In addition, the study investigates the following ancillary research questions:

RQ 1a. Do learners have different perceptions (perceived usefulness and preferences) of 'explicit CF' than of 'implicit CF'?

RQ 1b. How are learners' perceptions of CF related to their perceptions of themselves as receivers of CF?

In order to investigate these constructs in the most 'game-like' educational environment, and to maximize the potential tension between CF/instruction and immersion in play, an immersive 3D environment is used that was developed as a proof-of-concept for English language practice in complex learning tasks (see Figure II-5). The environment is task-based, as learners interact through authentic (written) dialogues with virtual characters, and is high in sensory detail in order to evoke an immersive and game-like experience. Even though the language tasks were intended to create an illusion of authenticity, and hence, openness, they are still focused from a linguistic point of view, and deal with grammatical constructions in use (pragmatics), such as asking open questions in networking dialogues. The environment also comprises different constellations of CF, ranging from 'explicit' to more 'implicit' types. Perceptions are measured through post-experimental questionnaires and semi-structured interviews. In this mixed-method study, quantitative and qualitative analyses are intended to form a comprehensive picture of the relations between these constructs.



Hannah Hartley: No, it's my first time here, but I'm enjoying the scenery tremendously!

● ME: For those of you who don't know me, my name is John Doe. I work for Hammersmith, a weaponry merchant here in Farglow.
You can use this phrase when you want to introduce yourself to an audience, e.g. for formal presentations.

Hannah Hartley: How come we should know you, Mr. Doe?

Anyway, I'm Hannah Hartley, researcher at the University of Warsaw.

- Is your work related to this fair, then?

● - Is your university interested in this weaponry fair?
Remember to always ask open-ended questions in networking conversations, these will stimulate your partner to say more than a simple 'yes' or 'no'. Open-endedness may be realized by using questions in a past tense, using the progressive, or chunks like 'In what ways ...?', 'How have you been ...?', 'What kind of ...?', 'What ... like?', etc.

● - What are you hoping to learn at this fair?

Continue

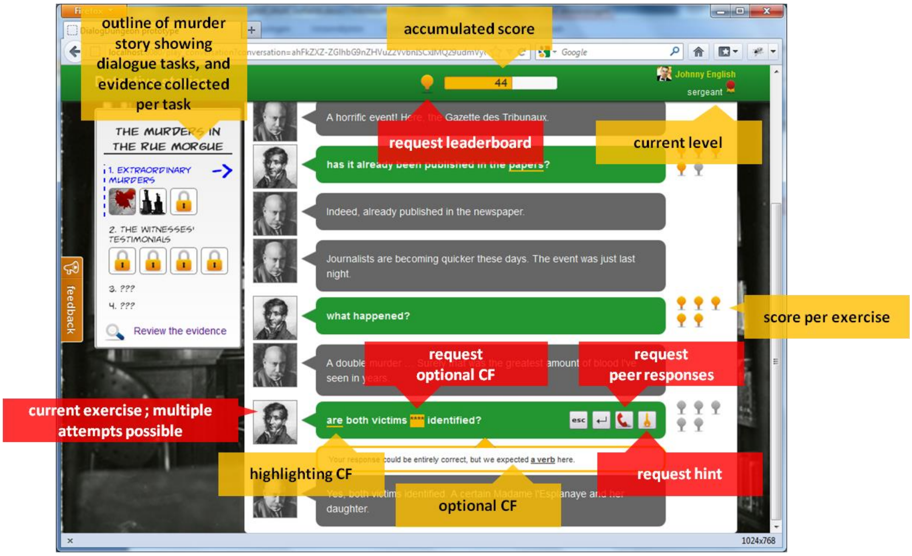
Figure II-5: digital learning environment for study on learners' perceptions

2.4.3.2 *Empirical focus 2: use of CF (study 2)*

The second empirical study zeroes in on learners' use of CF, and investigates links between CF use, perceived usefulness of CF, and L2 knowledge. It does so from a perspective that bridges the SLA and CALL literature on the use of CF with educational technology literature on tool use. The main research question is:

RQ 2. How does the perceived usefulness of metalinguistic CF in digital game-based language learning explain the actual use of such CF?

This study uses a learning environment that was designed to offer learners written L2 practice in scripted dialogue tasks embedded in stories. These tasks feature 'semi-open' exercises (Desmet, 2007), with which in which the learner needs to solve a meaningful problems (such as a murder mystery) by formulating responses that fall within the range of predicted correct utterances (see Figure II-6). Because the scope of these linguistic interactions is quite wide, many alternative responses are possible per interaction, which necessitates the use of CF if learners wish to approximate the predicted correct utterances. Learners, however, need not use this CF in order to advance in the story. In other words, this game-like environment afforded plenty of opportunities for revision of written production, and hence seems best for investigating learners' use of CF. The system uses natural language processing techniques to generate a wide array of feedback options, including which lemmas to use, simple indications of 'error' (i.e. the learner's response deviates from the predicted correct responses), highlighting feedback which shows the location of the deviations in the learner's response, metalinguistic prompts, and 'correct' responses.



A horrific event! Here, the Gazette des Tribunaux.

has it already been published in the papers?

Indeed, already published in the newspaper.

Journalists are becoming quicker these days. The event was just last night.

what happened?

A double murder… Surely that was the greatest amount of blood I've seen in years.

are both victims **** identified?

Your response could be entirely correct, but we expected a verb here.

Yes, both victims identified. A certain Madame l'Espanaye and her daughter.

Figure II-6: digital learning environment for study on use of CF

2.4.3.3 *Empirical focus 3: effectiveness of CF (studies 3-4)*

The third empirical focus of this project concerns learners' development of L2 grammar knowledge. Theoretically, it is guided by Skill Acquisition Theory

(DeKeyser, 2008), which assumes that explicit knowledge of grammar rules will assist the development of implicit L2 knowledge through extended practice in the L2. At the same time, in line with the Cognitive Mediational Paradigm (Winne, 1987), this investigation takes into account that learners' self-perceptions may mediate learning processes, and, hence may also determine the effects of CF on L2 development. More specifically, the effects of vivid CF on perceived immersion and competence (as components of intrinsic motivation) may predict learners' engagement in the tasks and may result in continued L2 practice, which is required for L2 knowledge to proceduralize in the longer term.

This research focus comprises two empirical studies. The first study (labelled 'study 3' in Figure II-4) focuses on the effectiveness of CF for supporting language learners' intrinsic motivation in L2 practice, and intends to answer the following main research question:

RQ 3. How does vivid CF affect learners' intrinsic motivation and their willingness to practise in digital game-based language learning?

The second effectiveness study (labelled 'study 4' in Figure II-4) deals with the effects of continued practice with CF on L2 grammar learning, and asks the following main research question:

RQ 4. How does continued practice with CF in digital game-based language learning assist learners in developing L2 grammar knowledge?

Subsidiary research questions include:

RQ 4a. How does the type of CF (metalinguistic CF vs. 'knowledge of results' CF) that is provided in practice influence the effects of practice on L2 grammar development?

To address these research questions, these two empirical studies make use of *task-oriented mini-games*. We define these as strongly form-focused L2 activities but embedded in meaningful L2 tasks, namely the reading of a mystery text. Therefore, these activities create opportunities for learners to focus on meaning in addition to focusing on form. These activities resemble L2 drills for the following reasons: they deal with well-defined linguistic constructions with a relatively small scope (e.g. the use of quantifiers in English), they revolve around fast-paced interaction, in which the learner needs to respond quickly to stimuli (see Figure II-7). Reaction times and accuracy scores for items are logged by the system, and the linguistic tasks consist of making grammaticality judgments, which together allow to construct a measure of automatized L2 knowledge at runtime (i.e. during the experimental treatment). During the mini-games, immediate 'knowledge of results' CF is given, and metalinguistic CF is available at the end of the mini-game tasks.

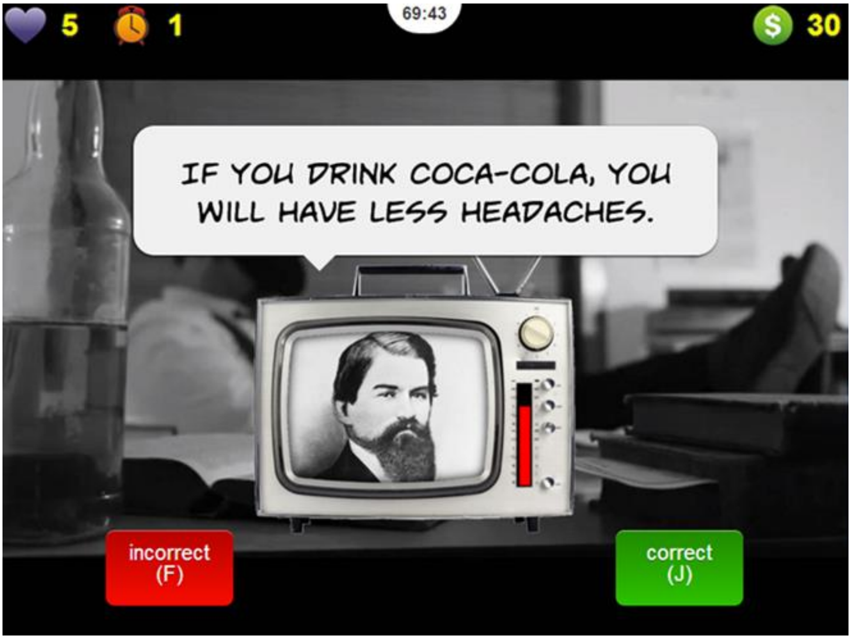


Figure II-7: digital learning environment for studies on effectiveness

2.4.3.4 Summary of the four studies

Table II-2 presents a summary matrix of the four studies. This summary comprises the main features of the learning environments and the research design.

We describe each study in terms of the following features of the learning environment: the 'type' of game, the degree of focus on form, the linguistic problems that formed the subject of language practice, the language skills that were addressed in technology-enhanced practice, the information that was available in CF in terms of the three types of information listed by R. Ellis *et al* (2006), and the gaming features inherent in feedback.

As for the research design, we describe the experimental design and the main variables measured in each study, along with the instruments used to measure those variables. For more detailed information, we refer to the following chapters.

Table II-2: summary of the four studies

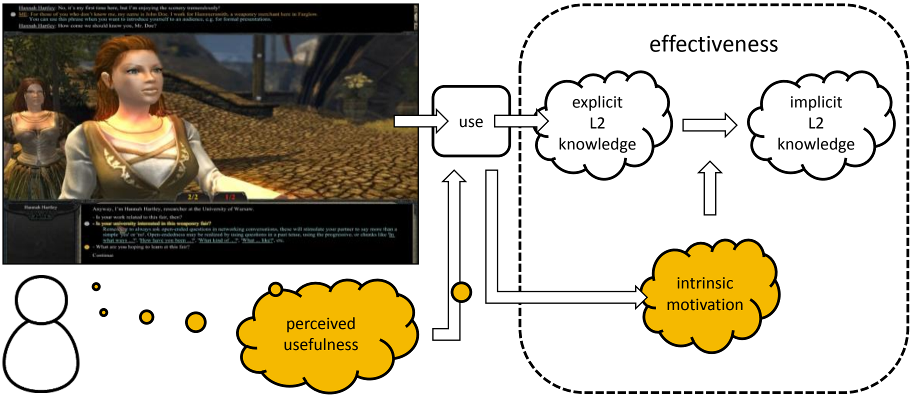
	1. perceptions	**2. usage**	**3. effects on intrinsic motivation**	**3. effects on L2 grammar learning**
		FEATURES OF LEARNING ENVIRONMENT		
'type' of game	immersive (dialogue tasks), 3D	immersive (dialogue tasks in murder mystery), primarily text-based	mini-games (2D)	mini-games (2D) embedded in mystery
degree of focus on form	+	++	+++	+++
linguistic problems	grammatical constructions in use (pragmatics), e.g. asking open questions (ill-defined)	grammar problems in questions, e.g. past tenses, quantifiers, modal verbs, etc. (well-defined)	complex grammatical rules: double object construction (well-defined)	complex grammatical rules: quantifiers, double object construction (well-defined)
skills involved in technology-enhanced practice	reading (listening)	reading, writing	reading	reading
information available in CF	- knowledge of results - metalinguistic feedback - correct response	- knowledge of results - metalinguistic prompts - correct response	- knowledge of results - metalinguistic explanations (delayed)	- knowledge of results - metalinguistic explanations (delayed)
game-like feedback	positive feedback (points) and characters' comments (adapted to the story)	positive feedback (points)	positive failure feedback	positive failure feedback
		RESEARCH DESIGN		
experimental design	correlational research design (*N*=83)	correlational research design (*N*=36)	experimental design, within-subjects (*N*=32), 3 conditions: - CF adapted to the fantasy (i.e. positive failure feedback) ; - CF not adapted to the fantasy ; - CF without fantasy	experimental design, between-subjects (*N*=186), 3 conditions: - practice with metalinguistic CF - practice without metalinguistic CF - no practice
perceived usefulness of CF	self-constructed scale	based on: technology acceptance model (TAM) (Davis, 1989)	-	-

perceived competence	taken from: intrinsic motivation inventory (IMI) (Plant & Ryan, 1985)	-	taken from: IMI	taken from: IMI
perceived immersion	taken from: game experience questionnaire (Ijsselsteijn et al., 2008)	-	taken from player experience of needs satisfaction model (PENS) (Przybylski, Rigby, & Ryan, 2010)	taken from: PENS
use of optional CF	-	tracking & logging data	-	-
explicit L2 knowledge	-	- metalinguistic knowledge test (R. Ellis, 2009b) - grammaticality judgment test (R. Ellis, 2009b)	-	self-constructed written discourse completion test
implicit L2 knowledge	-	-	-	- self-constructed timed grammaticality judgment test - self-constructed oral elicited imitation test - accuracy measures and reaction times from game tracking & logging data

First interlude

In the theoretical framework for the current research project, expounded in the previous chapter, we argued that it is critical that designers and researchers in digital game-based language learning take into account learners' perceptions, namely their perceptions of the learning environment (of corrective feedback in particular) and their perceptions of themselves as engendered by features of the instructional environment.

The following chapter presents the first empirical study of this research project, which explored learners' perceptions of different types of corrective feedback in digital game-based language learning, more particularly how useful learners find different feedback types. The study also charted the relation of perceptions of corrective feedback with respect to learners' self-perceptions, with a view to better understanding how feedback may affect learners' intrinsic motivation in digital game-based practice.



Hannah Hartley: No, it's my first time here, but I'm enjoying the scenery tremendously!
ME: For those of you who don't know me, my name is John Doe. I work for Hammersmith, a weaponry merchant here in Farglow.
You can use this phrase when you want to introduce yourself to an audience, e.g. for formal presentations.
Hannah Hartley: How come we should know you, Mr. Doe?

Hannah Hartley

Anyway, I'm Hannah Hartley, researcher at the University of Warsaw.
- Is your work related to this fair, then?
- Is your university interested in this weaponry fair?
Remember to always ask open-ended questions in networking conversations, these will stimulate your partner to say more than a simple 'yes' or 'no'. Open-endedness may be realized by using questions in a past tense, using the progressive, or chunks like 'In what ways ...?', 'How have you been ...?', 'What kind of ...?', 'What ... like?', etc.
- What are you hoping to learn at this fair?
Continue

Chapter III

Empirical study 1: Learners' perceptions of corrective feedback in an immersive game for English pragmatics

This chapter was reformatted and slightly modified from:

Cornillie, F., Clarebout, G., & Desmet, P. (2012). Between learning and playing? Exploring learners' perceptions of corrective feedback in an immersive game for English pragmatics. *ReCALL, 24*(3), 257–278.

Abstract

This paper aims to provide a rationale for the utility of corrective feedback (CF) in digital games designed for language learning, with specific reference to learners' perceptions. Explicit and elaborate CF has the potential to increase learners' understanding of language, but might not be found useful in a game-based learning environment where the primary focus for the learner is on meaningful interaction and experiential learning. Also, as CF can be perceived as a measure of performance, it could harm learners' perception of competence. 83 learners of English as a foreign language participated in a mixed-method empirical study that aimed to first explore the perceived usefulness of, and preferences for, explicit and implicit CF in an immersive educational game, and to secondly chart the relation between learners' perceptions of CF as they pertain to three individual difference factors related to learners' self-perception, namely intrinsic goal orientation, perceived competence and game experience. Survey and interview data showed that CF was found to be generally useful. A regression model indicated that the three measures of self-perception were positively associated with learners' perceptions of explicit CF; this was not the case for perceptions of implicit CF. Further, learners reported having enjoyed the implicit CF, although they did not find it particularly useful for learning. These findings indicate that the type of CF should be considered in the design of effective and enjoyable educational games.

3.1 Introduction

More than a decade ago, Hubbard (2002), a pioneering theorist, researcher and practitioner in the field of Computer-Assisted Language Learning (CALL), described a gap in the research on language learning games, stating that "the majority of [previous] research focused on demonstrating the validity of the general [game-based] approach rather than specific elements of its implementation." With the exception of a handful of recent studies that modified constituents of game-based learning environments (e.g. deHaan, Reed, & Kuwada, 2010; Ranalli, 2008), the need for careful attention to implementation still exists today. One "specific element of implementation" that deserves greater attention is feedback, which is widely recognized as crucial both for linguistic development and as a core feature of game mechanics.

The game-based learning (GBL) literature identifies feedback as an element that is both central to games and indispensable for learning (e.g. Aldrich, 2005; Becker, 2007; Prensky, 2001). In commercial games, feedback is considered to give players, who engage in a series of goal-directed activities, a measure of how well they are progressing towards goals. Models of GBL propose that it is essentially "from the feedback in a game that learning takes place" (Prensky, 2001, p. 121). While Prensky's statement suggests that feedback applies to both games that are intended to entertain as well as to educate, it is obvious that for the latter, developmentally useful feedback is all the more essential.

This paper will first conceptualize feedback in CALL games by interweaving theory in the second-language acquisition (SLA) and GBL literatures. It will then argue that individual difference factors, in particular learners' self-perceptions, need to be taken into account in the design of CALL gaming environments. Finally, it will present results from an empirical study that aimed to chart the relation between learners' perceptions of feedback and their self-perception in an immersive game for English pragmatics.

3.2 Background research

In this section, we first present a contrastive outline of how feedback is being conceptualized in the separate literatures on SLA and on GBL, and synthesize its hypothesized and observed benefits for learning. Then, we discuss how learners' self-perceptions may mediate the effectiveness CF. We conclude this section with a theme that occurs in the literature on GBL, namely the need for balancing instruction and play, and relate this to how feedback could be implemented in digital game-based language learning.

3.2.1 Conceptualisation of feedback in GBL and SLA

In the literature on SLA, the kind of feedback that is directed towards learning is generally known as *corrective feedback* (CF) or *negative feedback* (Long, 2007). CF refers to all responses to learners' erroneous utterances in a L2, and it may include an indication that an error has been made, can present the correct form or metalinguistic information about the nature of the error, or it may combine these various forms of information (R. Ellis et al., 2006).

In the GBL literature there appears to be almost unanimous agreement on the beneficial role of feedback for learning. However, the field of SLA has long been divided over the topic, and the type and timing of CF remain debated issues (Long, 2007). Depending on the theoretical assumptions concerning how an L2 is acquired, CF is attributed a more or less favourable role, is thought to be more or less effective, or even to have harmful effects, such as to increase anxiety and to foster less favourable attitudes towards learning (Truscott, 1996).

The first assumption concerns the interface between explicit and implicit knowledge (N. C. Ellis, 1997). Theories that are largely constructed on the importance of implicit learning mechanisms for acquisition, such as generativist (nativist) theories of language learning (Schwartz, 1993) and Krashen's Monitor Theory (1981), presume that explicit knowledge is disconnected from implicit knowledge, that CF inculcates explicit knowledge, and that therefore CF can only contribute to learned knowledge and not to acquisition. Conversely, if an interface between explicit and implicit knowledge does exist, CF is seen to foster acquisition in the longer term. For instance, Skill Acquisition Theory (DeKeyser, 2008) emphasizes the possibility of explicit knowledge becoming implicit over time. A related issue is the role of awareness and noticing in SLA (Schmidt, 1990). From this perspective, CF, especially in more explicit forms, is generally considered to stimulate noticing and conscious processing, both of which are presumed to promote SLA. However, some kinds of implicit CF, which signal in less overt ways that an error has been committed (e.g., recasts), have received theoretical attention precisely because they are more implicit, and are considered beneficial for acquisition because they jointly focus on form and meaning, leading to strong form-function mappings in the flow of communicative interaction (Long, 2007).

Thus far, empirical research on learning outcomes suggests that some form of CF is beneficial. The effects of CF on various aspects of L2 development have been demonstrated in a number of studies, both in (quasi-)experimental instructed L2 settings (e.g. Carroll & Swain, 1993; Long, Inagaki, & Ortega, 1998; Takimoto, 2006), in more naturalistic classroom settings (e.g. Havranek, 2002; Loewen & Philp, 2006; Lyster & Ranta, 1997), and in CALL (e.g. Brandl, 1995; Heift, 2004; Nagata, 1993; Pujolà, 2001). Also, reviews tentatively suggest that CF types that include metalinguistic information (such as grammar rules) and/or which function as prompts (signalling the error without providing the correct response), aid language development more than CF types which are generally subsumed under the header 'implicit', such as recasts (R. Ellis et al., 2006; Lyster & Saito, 2010). Although these findings are still tentative due in part to methodological difficulties in CF research (R. Ellis et al., 2006; Long, 2007; Lyster & Saito, 2010; Mackey & Goo, 2007; Russell & Spada, 2006), Norris and Ortega's (2000) quantitative meta-analysis of the effectiveness of explicit vs. implicit instruction showed larger effect sizes for instruction that included rule explanation. In summary, the preponderance of current research suggests that CF which stimulates conscious processing of L2 input through rule explanation or explicit prompting is likely to be more effective.

The GBL literature is less well articulated with respect to what kinds of feedback best support learning in games, and if the purpose is to educate (rather than purely entertain), feedback mechanisms as they relate to learning remain underexplored. A first observation is that, in contrast with non-GBL environments, games seldom give away answers to players (Becker, 2007). Games usually stimulate explorative behaviour (Kiili, 2005), and aim to motivate players to find 'correct' answers through trial and error.

Secondly, it can be observed that games rarely articulate in a direct way the domain knowledge or rules which underlie the in-game content, and which might serve as an immediate support mechanism that leads players to solve problems successfully. Although the primary purpose of commercial games is to entertain rather than to educate, and the learning is only a side-effect of gaming, a comparison may be drawn with experiential learning models: problem-solving in games follows a fixed pattern of being exposed to a concrete experience and to data, making reflective observations, construing mental generalizations and hypotheses about the experience, and testing these hypotheses through active experimentation (Kiili, 2005). The discovery of patterns and construal of generalizations by players themselves is an essential feature of game experience. Koster (2005) points out that part of the attraction of games is the process of pattern seeking, which is a fundamental aspect of human experience and aligns with how our brains work. He argues that our brains seek patterns "so much we don't even realize we're doing it", i.e., without our conscious attention and without explicit teaching, much like the process of first language acquisition (Koster, 2005, p. 16). In this view, gaming seems to cater to implicit learning.

Third, feedback in video games is different because it is connected to the representation of the game's world or theme. For Prensky (2001), feedback comes about "as action" (p. 159), e.g. a player's character may die because enemies are quicker. Or, as in the simulation game *The Sims*, the player's character is sacked for unproductiveness at work as the result of continuous nights without sleep. Feedback in games, in other words, is largely dependent on the content or theme of the game. In SLA terms, this implies a focus on meaning.

Thus, the conceptualisations of feedback in GBL and SLA differ quite clearly, as feedback in games lacks the provision of correct responses and rule explanation, and is adapted to the game's theme. This warrants empirical research on the effectiveness of more 'game-like' types of feedback in comparison with more 'traditional' CF in language learning contexts.

3.2.2 The role of the learners' self-perceptions

It is evident that feedback in games also has purposes other than to deal with mistakes and failure (i.e. negative feedback), and that it can also serve to reinforce, to reward or to maintain motivation (i.e. *positive feedback*) (Becker, 2007, p. 25). Although the latter is somewhat outside of the scope of this paper, it is difficult to separate these two types of feedback in games. Professional game designers spend considerable time on the design of so-called *failure states* (Purushotma et al., 2008). Failure states are phases in the game in which it is made clear to the player that something has gone wrong, or that the player has not adequately performed an activity. For game designers, it is critical that such failure states are interesting and enjoyable, and that the player can repeatedly fail without compromising the motivation necessary for successfully completing an action or task. So, feedback design for GBL may not ignore individual difference factors, more particularly learners' perceptions of themselves, i.e., as successful or failing learners and players.

In SLA research on CF, claims have been made that individual differences have been underestimated, as they may mediate the effectiveness of CF (e.g. DeKeyser, 1993). This reflects the central tenet of the Cognitive Mediational Paradigm (Winne, 1987), which posits that the effects of instruction are mediated by learners' cognitions such as their conceptions and perceptions of the learning environment, their prior knowledge and aptitude, and their attitudes and self-perceptions. As to learners' self-perceptions, some SLA scholars have argued that CF is harmful because it can reduce motivation (e.g. Truscott, 1996) and impede interactional processes. Teachers and pedagogies struggle with a "balancing act of two necessary but seemingly contradictory roles", i.e., to "establish positive affect among students yet also engage in the interactive confrontational activity of error correction" (Magilow, 1999, p. 125). Even in tutorial CALL, where CF may be less face-threatening, learners may experience it as a form of "mild social punishment" (Schulze, 2003) or they may intensify "salient cues [feedback] which are evaluated as negative" (G. L. Robinson, 1991).

The research on individual differences and CF provides ample evidence that students favour CF and that they find it helpful (Cathcart & Olsen, 1976; Chenoweth et al., 1983; Hedgcock & Lefkowitz, 1994; Radecki & Swales, 1988; Saito, 1994; Schulz, 2001). There are also findings suggesting that students prefer detailed metalinguistic feedback more than a 'right/wrong' type of feedback (Nagata, 1993) and more than implicit feedback in the form of recasts (Kim & Mathes, 2001).

Moreover, there is evidence that students' attitudes towards CF (Havranek & Cesnik, 2001), their anxiety (DeKeyser, 1993; Havranek & Cesnik, 2001; Sheen, 2008) and their prior motivation (DeKeyser, 1993) explain differences in learning gains. In these studies, positive attitudes and low anxiety resulted in higher levels of L2 development. In addition, DeKeyser (1993) found that learners with high extrinsic motivation did better without systematic and explicit error correction. Students with low extrinsic motivation, on the other hand, excelled when they did receive systematic CF. So, there is some evidence suggesting that individual differences related to learners' perceptions of themselves, namely anxiety and prior extrinsic motivation, mediate the instructional effectiveness of CF. To our knowledge there is no empirical research on the effects of CF on self-perceptions.

L2 pedagogy and communicative approaches to language teaching, in particular, tend to take into account the learner when advising on when and how to use CF, with the specific recommendation to not correct too frequently during communicative interaction. It is unclear whether the intention is to safeguard learners' self-perceptions, but one purpose certainly is to favour communicative fluency over linguistic accuracy, at least while tasks are being carried out. In a communicative approach, a focus on language as meaningful communication precedes a focus on isolated language forms (R. Ellis, 2003). As a consequence, CF plays a subservient yet crucial role during communicative tasks, as it can draw attention to linguistic form implicitly and/or after the completion of a task. High discrepancy between teachers' beliefs on CF and learners' attitudes towards CF, especially for speaking activities (Magilow, 1999; Schulz, 2001), shows that teachers take into account in daily practice the pedagogical reflex to use CF with care. CF is typically delayed in task-based language teaching (Willis & Willis, 2007) and classroom simulation/gaming (Bullard, 1990), and occurs mainly during the post-task/debriefing phase. In communicative language teaching, teachers make wide use of (more or less) implicit recasts (Lyster & Ranta, 1997), as they are relatively undisruptive to communicative flow. This brings us to the next point.

3.2.3 Need for a balance between instruction and play?

A parallel can be drawn between the role of CF in SLA and how CALL games could embrace instructional feedback. A recurring theme in the game-based learning literature is the opposition between learning and playing. Such a dichotomy is often articulated by the claim that a "subtle balance" needs to be found between learning and gaming (Kickmeier-Rust & Albert, 2010, p. 95; Kiili, 2005). The underlying assumption here seems to be that learning is by default laborious or unpleasant, or that learners are demotivated, and that gaming is the panacea that will raise learners' motivation. The ultimate goal for educational game designers, then, is to protect "flow", that is, the feeling of being fully engaged in an activity (Csikszentmihalyi, 1990), the interaction, and the feeling of immersion in experience. Evidently, debate on the utility of feedback in educational games is influenced by such thinking. Although not necessarily an advocate of the learning vs. playing argument, Prensky (2001) writes that "the art of providing feedback in a game is extremely important and complex because either too little or too much can lead quickly to frustration for the player" (p. 122).

Thus, the GBL literature suggests applying (negative) feedback with moderation, or at least in playful forms, so as to keep the player/learner engaged. In communicative language teaching approaches, with which game-based language learning has been associated because of its emphasis on language as a resource to complete meaningful tasks (Baltra, 1990; Purushotma et al., 2008), the rationale seems different. Here, the primary purpose is to ensure communicative fluency rather than to safeguard motivation or prevent frustration.

Still, the parallel seems worth investigating for two reasons. First, as was indicated above, some SLA scholars have suggested that corrective feedback may reduce the motivation of language learners (Truscott, 1996). To date, however, empirical evidence for this claim is lacking. Secondly, there seems to be no theoretical or empirical reason to posit an opposition between 'learning' and 'playing'. Possibly, the same intrinsic motivational processes are at play in learning as in gaming.

A theory that might help explain this is Self-Determination Theory (SDT) (Ryan & Deci, 2000). SDT, as a comprehensive theory of how human beings are motivated to perform various activities in various contexts, might back up the claim that there is no tension between play and learning, or between intrinsic motivation and CF. According to SDT, people require a certain amount of feedback on their actions in order to build up and experience competence. In games, specifically, a player's need for competence could be satisfied by the provision of "meaningful informational feedback", that is, feedback which is useful, non-judgemental and immediate, and which thus allows a player to improve his or her performance (Rigby & Ryan, 2011, p. 19). Our hypothesis is that in educational games, CF can satisfy learners' need for confirmation of competence if the CF is found useful (i.e., if learners have the impression that they are learning) and if the CF is actually used to complete activities in the game.

3.3 The current study

The convergence between the conceptualization of CF in the SLA literature and language pedagogy on the one hand, and the hypothesized need for a 'subtle balance' between learning and playing in GBL on the other hand, raises issues for the design of effective CALL games. Questions may be asked with respect to the utility and desirability of explicit CF and more implicit, 'game-like' kinds of feedback in relation to the motivation of language learners, their perceptions and use of learning support in educational games, and the effects of different configurations of feedback in regard to L2 development.

The current study aimed to probe learners' perceptions of CF in an immersive game for English pragmatics, and to explore the relation of these perceptions about the learning environment with learners' perceptions of themselves as learners and as players. First, it may be argued that students will not learn from CF if they do not find it useful within the context of the goal-directed actions that comprise the game mechanics. Secondly, there may be differences between learners' perceptions of explicit and implicit CF, as implicit CF may be more aligned with the feel of an immersive game. Third, if learners perceived themselves as competent, possibly as the result of interacting with useful, clear and informational CF in the game, perceived competence might explain perceptions of CF: players will find CF useful and will prefer conditions which provide it. Fourth, learners' intrinsic goal orientation for learning English can predict their perceptions of CF. If they are intrinsically interested in learning English, students will probably find CF useful and will prefer it. Finally, learners' game experience might predict CF perceptions. If they feel immersed in a game, are interested in its story (which implies a focus on meaning), and feel good as a result of playing the game, they might find CF, and especially explicit CF formats, disturbing or less useful.

Therefore, we pose the following research questions:

1. Do learners find CF useful in an immersive game for English language learning?
2. Do learners have different perceptions (perceived usefulness and preferences) of explicit CF than of implicit CF in such a game?
3. Do learners' intrinsic goal orientation, their perception of competence as a result of playing the game, and/or game experience explain their perception of CF?

3.4 Method

3.4.1 Participants

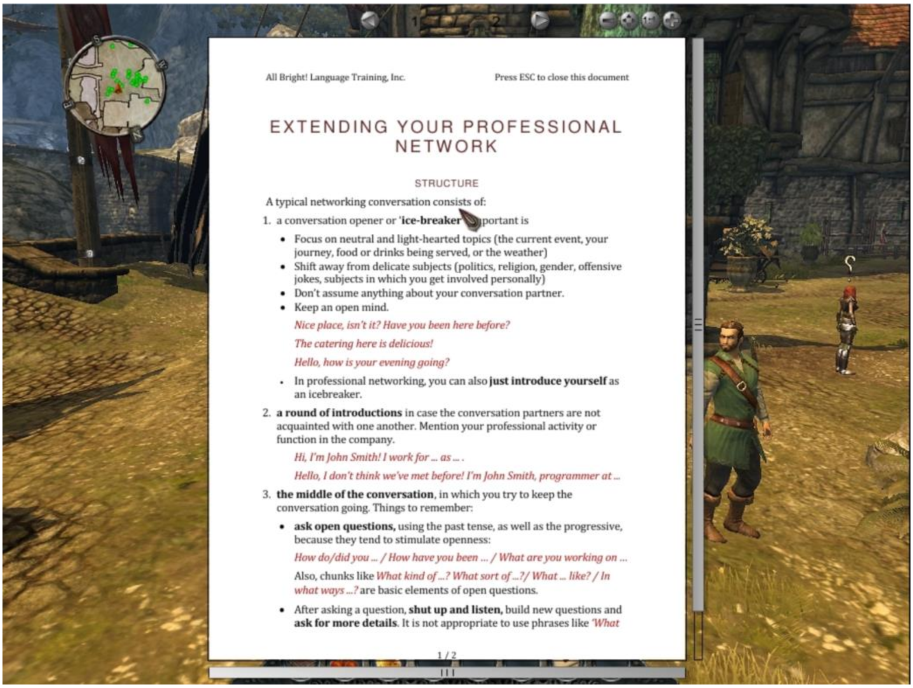
The participants included 83 first-year university students and learners in their final two years of high school in Belgium. The university students, which represented the majority of participants, were enrolled in various programmes, a minority of which were language programmes. The large majority of these students (82 %) did not have English as a compulsory study subject, but did have to read texts in English for other courses. Their level of English was around B1 (intermediate) of the Common European Framework of Reference for Languages (Council of Europe, 2011), the required level at the end of high school education. Sixty-one students were female, twenty-two male. The age range was between 16 and 24 except for two 33-year-olds (*Mdn* = 19).

The game and associated instructional materials were developed through a public-private partnership as a proof-of-concept of a language learning game, and were hence not integrated in the participants' curricula. The university students participated either on a voluntary basis, or as part of a methodology course taught in the educational sciences. The high-school learners were invited to participate in the study through the researchers' personal contacts with teachers. The learners were told that they would take part in an experiment, and played the game only once and in one session. The game sessions lasted between 30 and 60 minutes approximately.

3.4.2 Description of the learning environment

In order to maximize the potential tension between CF and a game environment, we chose to create a fully immersive 3D avatar-based game, using a game development kit which had been provided by a renowned Flemish professional role-playing game (RPG) developer. In the commercial standalone RPG, the player is a dragon slayer who carries out various quests in a medieval-looking fantasy setting. The RPG relies heavily on narrative and point-and-click dialogues. The dialogues contain written transcripts, and feature voice actors and detailed character animations. As is the case with most commercial RPGs, written language is used in object descriptions, the player's inventory, and a logbook of completed and current quests.

On the basis of this RPG engine and the available character and world visuals, the first author co-developed a game customized for the training of English pragmatics. The learning goal was to equip high-intermediate learners with the typical constructions and speech acts necessary for making formal conversation in two domains: social introductions and professional network development. The first author developed the content for the game on the basis of materials provided by a language training company external to the university. Content included dialogues, a document with supportive information (text outlining conversational structures, model utterances of speech acts, and explanations on the situations in which to use these utterances; see Figure III-1), and the elaborate feedback messages that would be shown upon mistakes (see below). Finally, a professional trainer in Business English proofread all instructional materials.



All Bright! Language Training, Inc.

Press ESC to close this document

EXTENDING YOUR PROFESSIONAL NETWORK

STRUCTURE

A typical networking conversation consists of:

1. a conversation opener or '**ice-breaker**' portant is
   - Focus on neutral and light-hearted topics (the current event, your journey, food or drinks being served, or the weather)
   - Shift away from delicate subjects (politics, religion, gender, offensive jokes, subjects in which you get involved personally)
   - Don't assume anything about your conversation partner.
   - Keep an open mind.

     Nice place, isn't it? Have you been here before?

     The catering here is delicious!

     Hello, how is your evening going?
   - In professional networking, you can also **just introduce yourself** as an icebreaker.
2. **a round of introductions** in case the conversation partners are not acquainted with one another. Mention your professional activity or function in the company.

   Hi, I'm John Smith! I work for ... as ... .

   Hello, I don't think we've met before! I'm John Smith, programmer at ...
3. **the middle of the conversation**, in which you try to keep the conversation going. Things to remember:
   - **ask open questions,** using the past tense, as well as the progressive, because they tend to stimulate openness:

     How do/did you ... / How have you been ... / What are you working on ...

     Also, chunks like *What kind of ...? What sort of ...?/ What ... like? / In what ways ...?* are basic elements of open questions.
   - After asking a question, **shut up and listen,** build new questions and **ask for more details**. It is not appropriate to use phrases like *'What*

1 / 2

Figure III-1: in-game document with supportive information

The learners' overall objective in the customized game was to advance from quest to quest by choosing in the conversations with *non-player characters* (NPCs) the most pragmatically appropriate response from a list of options that was presented on the screen. The voices of the NPCs had been pre-recorded with a team of English language teachers; the recordings were post-produced in order to make them sound 'real' and to improve the audio quality. In order to make the NPCs further 'come alive', professional game development artists animated them by applying gestures and lip-synchronization. Although the setting was medieval, the projected world was highly detailed in visuals (3D, fine-grained textures) as well as in aural detail (human voices and sound effects). So, it provided to learners a simulation environment that was high in perceptual fidelity—it looked and sounded 'real'—and moderate in functional fidelity (de Jong, 2005); the quests were real-world tasks but the point-and-click dialogues were impoverished in comparison with natural conversation. As a consequence of high fidelity, learners could practice real-world tasks in a safe environment, and were immersed in the 3D world, which resulted in a mean score on the intrinsic motivation inventory (Plant & Ryan, 1985) (see below) of 66 percent (*SD* = .12).

In order to maximize learning opportunities, instruction was designed according to the Four Component Instructional Design (4C/ID) model (van Merriënboer & Kirschner, 2007). 4C/ID is suitable for complex learning tasks and is composed of [1] tasks (grouped in 'task classes' or sets of tasks that are similar in content and complexity), [2] supportive information (general models and schemas that help to teach nonrecurrent aspects of tasks), [3] just-in-time information (aimed at encoding recurrent aspects of tasks into rules), and [4] part-task exercises (designed to help learners gain automaticity of recurrent aspects of tasks). The latter were not included because of the brief play period associated with this research. In the game, the tasks coincided with interactive dialogues. Learners had to complete two task classes (one for introducing people and one for networking), each of which contained three tasks and supportive information including a 'theory' document (see above and Figure III-1) and a dialogue model (a non-interactive dialogue between NPCs which demonstrated the use of speech acts). In accordance with the chosen instructional design model, the dialogues within one task class were very similar in complexity and in content, but support within the dialogues of a task class, more specifically just-in-time information, was gradually withdrawn. Just-in-time information was operationalized as CF, and was different in each of the three tasks/dialogues.

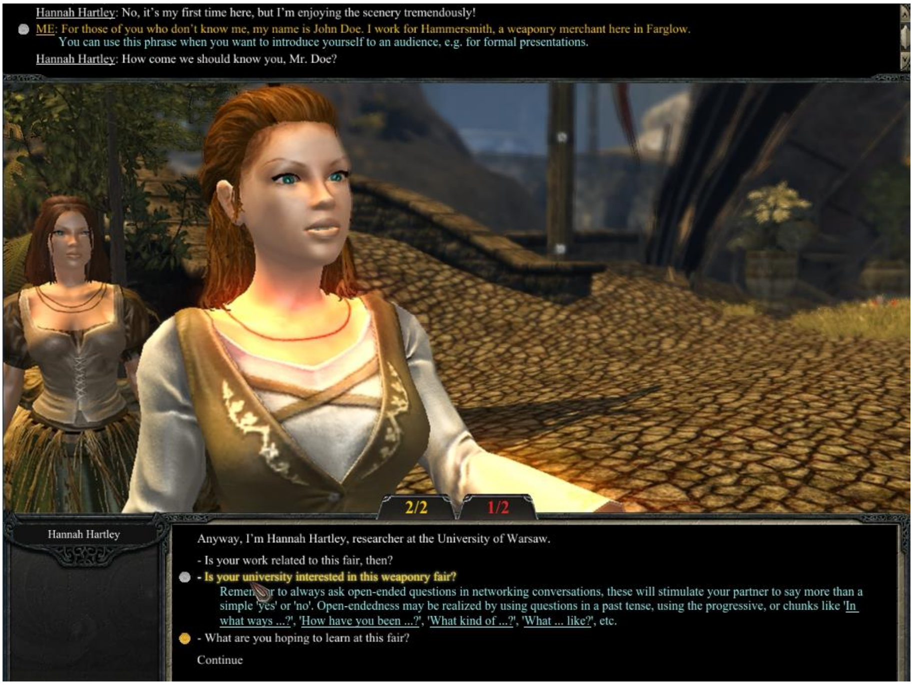
In the first task, there was a high level of CF (type A): when a response was clicked, the dialogue paused, and learners were shown visual feedback as to whether their selection was correct or incorrect. When the response was incorrect, they were shown the written correct response and a written metapragmatic explanation. They also had access to explanations for alternative responses. When the learners had finished reading the feedback, they clicked a button, upon which the dialogue continued. In this phase, if an incorrect response is selected, the NPC replies with an implicit spoken comment which expresses surprise or a lack of understanding. This response is accompanied by the NPC's gestures (e.g., frowning, showing surprise, waving arms). In the second task (CF type B), metapragmatic explanation was no longer shown by the system, but learners could still request it for each possible answer by clicking on the answers. The third task (CF type C) contained the least support: the system no longer paused, but immediately moved on to the character's response, which is the default interaction in commercial RPGs, and the correct response was hidden. The system did show, as in the previous tasks/support levels, whether the chosen response was correct or incorrect, and the NPC responded accordingly. A written metapragmatic explanation could be requested for the response if it was inappropriate.

Table III-1: CF types

	CF type A (task 1)	**CF type B (task 2)**	**CF type C (task 3)**
dialogue state	pause	pause	continue
visual positive/negative feedback	yes	yes	yes
written correct response	yes	yes	no
written metapragmatic explanation	yes	on learner's request	on learner's request
character's spoken response	delayed	delayed	immediate

As this study was part of another study (Vandercruysse, Vandewaetere, Cornillie, & Clarebout, 2013) which intended to investigate the effects of the game element 'competition', the positive/negative feedback was visualized to half of the students as a green checkmark and a red exclamation mark, respectively, and to the other half of the students as a golden coin and a silver coin, respectively. Students were in each case told before the experiment what the meanings of the icons were.

In total, three types of CF were included in the game ranging from explicit to implicit. Type A was most elaborate and showed a metapragmatic explanation immediately upon a mistake (see Figure III-2); type B was thought to stimulate self-discovery of rules as learners could compare appropriate with inappropriate responses and could see the metapragmatic explanation on request; type C was more aligned to formats common to recreational game environments as it relied principally on the reactions of NPCs. Students would typically have been familiar with CF types A and B on the basis of their experience in a 'traditional' learning environment, but familiar with CF type C only if they had experience playing commercial video games.



Hannah Hartley: No, it's my first time here, but I'm enjoying the scenery tremendously!

● ME: For those of you who don't know me, my name is John Doe. I work for Hammersmith, a weaponry merchant here in Farglow.
You can use this phrase when you want to introduce yourself to an audience, e.g. for formal presentations.

Hannah Hartley: How come we should know you, Mr. Doe?

Anyway, I'm Hannah Hartley, researcher at the University of Warsaw.

- Is your work related to this fair, then?

● - Is your university interested in this weaponry fair?
Remember to always ask open-ended questions in networking conversations, these will stimulate your partner to say more than a simple 'yes' or 'no'. Open-endedness may be realized by using questions in a past tense, using the progressive, or chunks like 'In what ways ...?', 'How have you been ...?', 'What kind of ...?', 'What ... like?', etc.

● - What are you hoping to learn at this fair?

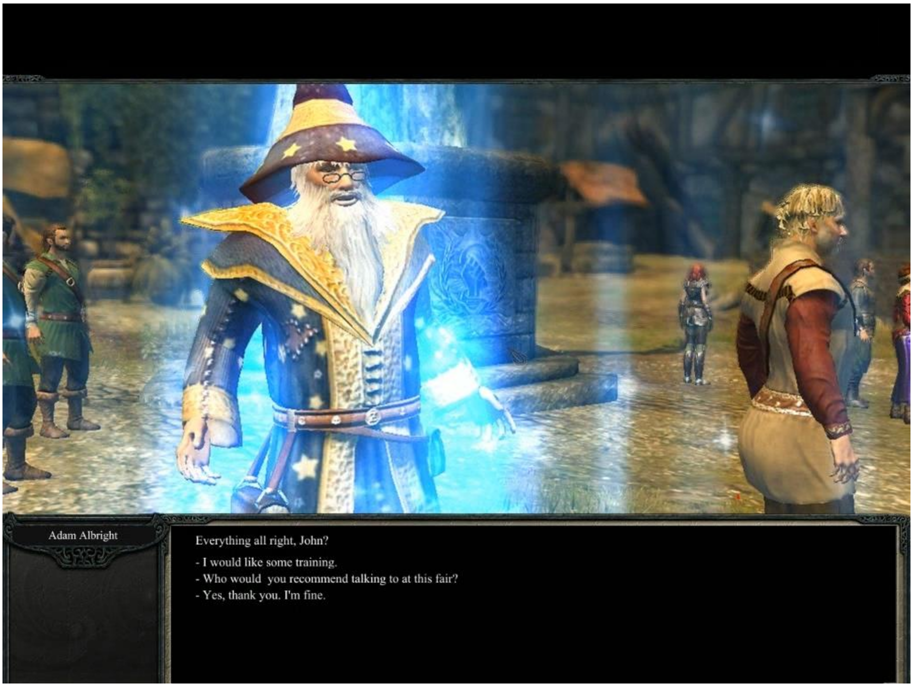
Continue

Figure III-2: elaborate CF (type A)

For the metapragmatic explanations, we avoided terminology with which the learners would not normally be familiar at the targeted proficiency level. The metapragmatic explanations were between two and four sentences in length. Below is a sample of the metapragmatic CF:

> Remember to always ask open-ended questions in networking conversations, these will stimulate your partner to say more than a simple "yes" or "no". Open-endedness may be realized by using questions in a past tense, using the progressive, or chunks like 'In

what ways ...?’, ‘How have you been ...?’, ‘What kind of ...?’, ‘What ... like?’, etc.



Adam Albright

Everything all right, John?

- I would like some training.
- Who would you recommend talking to at this fair?
- Yes, thank you. I'm fine.

Figure III-3: language trainer character offering help

Apart from the CF immediately following a response, students could rely on other support mechanisms. First, at the top of the screen a student could see the transcript of the currently active dialogue as he or she had played it, and in which the CF was available so that past errors could be reviewed. Secondly, a pedagogical agent in the form of a language trainer appeared after each dialogue in which the student made mistakes. This character offered help, which learners could reject or accept (see Figure III-3). In the latter case, learners could access the supportive information (see above): a document containing speech acts related to the past dialogue (see Figure III-1) and the dialogue model. Students could also access the supportive information in the absence of the trainer by pressing a key, and they were continually reminded of these materials through the game interface.

3.4.3 Data collection

The data were collected in December 2010 and January 2011 through questionnaires, interviews and game logs. Before the game, students filled in the motivated strategies for learning questionnaire (Pintrich, Smith, Garcia, & Mckeachie, 1993), of which we retained the intrinsic goal orientation subscale. This scale consisted of four items (Cronbach's α = 0.69), and focused on mastery, learning and challenge (e.g., "I prefer tasks which I can learn from, even when this does not result in good grades"). After the experiment, we measured perceived competence using the intrinsic motivation inventory (Plant & Ryan, 1985). The subscale of perceived competence contained six items (α = 0.89) (e.g., "I think I am pretty good at this activity"). Students also filled in a game experience questionnaire (De Grove, Van Looy, & Courtois, 2010), of which we retained the scales targeting immersion (e.g., "I felt totally absorbed"), vividness (e.g., "I was captivated by the story of the game") and positive affect (e.g., "I felt satisfied") (6 items; α = 0.63), as these dimensions seemed most crucial to determine a positive experience of playing such a 3D RPG. Finally, students also filled in a 7-point Likert questionnaire on CF (see Appendix 1), which we developed for this study. All 83 students filled in these questionnaires before and after playing the game.

Additionally, the first author conducted semi-structured interviews with twelve students on the basis of a convenience sample. The students were interviewed in their mother tongue (Dutch). Nine of these students worked in the 'competition' condition, which implies that they also saw their in-game score compared with the score of a virtual opponent, which was dynamically adapted to the student's score to create a feeling of competition. The other three students did not see any scores. During the interviews, the researcher first probed students' general conceptions about CF, and then asked what they felt about the in-game CF in terms of usefulness and preferences. In the fashion of a *stimulated recall* measure (Loewen & Reinders, 2011), students were shown screenshots of different types of CF in the game. Interview data were coded and analysed in two cycles, first deductively on the basis of an initial set of constructs (e.g., a number of general feedback characteristics, perceived usefulness, perceived competence, preference, immersion), followed by more fine-grained inductive coding on the basis of a second reading.

The game logs measured learners' behaviour in the game, such as the chosen responses in the dialogues, how long they worked on the tasks, and whether they made use of the supportive materials.

3.5 Results

3.5.1 Preliminary analyses

Using the game logs, we first determined whether students were actually 'exposed' to the CF during the game. By taking the ratio of correct responses as a proportion to the total number of responses (three students responded a few times less than the expected 38 times), we computed the students' performance in the game, which ranged between accuracy scores percentages of 31 and 77 (M = .58, SD = .1). This implies that all students had been presented with CF. This was confirmed in the interviews, as none of the students showed surprise when seeing the screenshots containing CF, and because they could clearly explain the purpose of the CF.

In what follows, findings will be presented and discussed with respect to the perceived usefulness of CF in the game, the perceptions of explicit and implicit CF, and the relation of these perceptions with variables associated with self-perception.

3.5.2 Usefulness of CF

Quantitative analysis showed that, generally, students found the CF quite useful, with average scores for items 2-5 of over 5 on a 7-point scale with 7 being perceived of as most useful (see Appendix 1). The first item, which targeted inductive/discovery learning had a lower mean and a higher standard deviation.

The relatively high score for perceived usefulness was confirmed in the interviews. None of the students responded that the CF was not useful generally. When questioned about the reasons, students said they found CF useful because it helped them to learn, and to remember the content. Additionally, one particular student seemed to claim that CF helped to realize transfer to contexts outside of the game.

> Interviewer: Let's summarize. How did you experience correction on mistakes in general?
> Student 2: Err, well, I just remembered it. Also in order to use it further on in the game, *but also for now*. Well yeah, I think it's good that it [the game] contained feedback, because, as I said, you learn from your mistakes. [emphasis added]

3.5.3 Perceptions of explicit and implicit CF

We defined explicit CF as CF which contains a rule (according to the definition of explicit instruction, Hulstijn, 2005) and explicit information about the correctness of learners' responses (positive/negative feedback), which is immediately given when mistakes are made, and which offers students the opportunity to reconsider the options after they have responded (7 items, α = .69). Implicit CF was defined as CF that is adapted to the game environment (i.e., the characters' reactions), which stimulates autonomous inquiry by learners, and in which errors and rules are only shown after the task (6 items, α = .59). In order to create the subscales, we took the means of the corresponding items. 'Perception of CF' was defined as the combination of perceived usefulness of and preference for a particular type of CF. The data from the Likert responses (see Appendix 1) show a higher mean score for perception of explicit CF (*M* = 5.2, *SD* = .8) than for implicit CF (*M* = 3.8, *SD* = .9).

Analysis of the interviews revealed similar results. Students specifically found the most elaborate feedback (type A) most useful. They also preferred it to appear throughout all the dialogues, instead of just in the first one.

> Here the answers were judged on what was best in the situation, and also the second best … that I found good. Also the blue text that appeared, explaining the situation in which the chosen response was better, I found really useful. I had preferred it to appear all of the time, but sometimes it didn't. And also that the answers stayed. Because sometimes I still had to look in the history of the text, whether I had given the correct response or not, and there the correct response wasn't shown. (Student 4)

Additionally, students indicated that they found the rule-based feedback most useful because they could easily memorize it and apply it to similar situations in the game.

> Interviewer: Had you preferred the blue text to appear all of the time?
> Student 4: Yes.
> Interviewer: Why?
> Student 4: To be able to learn from your mistakes, to know why you made a mistake, and as such to be able to apply it to future problems.
> Interviewer. So you did not find this [the CF without rule] sufficiently informative?
> Student 4: No, not really. Well, of course it's good that you can still, well, see what the correct answers are, but … not really why. In this way you cannot extrapolate it to other problems.

> Student 9: Yes, well that the response was incorrect. And the feedback you get. [Explicit explanations of] Why the response was incorrect, I found really good. Because then you are immediately taken on your point, and then you know that you will maybe not make the same mistake in the future. Or hopefully not. And then you also know what the correct response was. So that was positive, yes.

Implicit feedback was found less useful for learning:

> Student 2: In fact he is responding to the incorrect response that was uttered before.
> Interviewer: And is that useful? Is that an example of feedback for you?
> Student 2: No. Not that.
> Interviewer: Why not?
> Student 2: (laughs) Well in fact what that man is doing is to answer beside the point, so in fact you don't learn anything from that. In fact he says: 'what are you saying?', for example. So that is not really feedback to me.

> Student 7: But wasn't that so with these characters? If you indicated the incorrect response, then it was like 'hmm yes, ok, well'. Incorrect.
> Interviewer: And what does that do to you?
> Student 7: I actually preferred the feedback so that I knew I was wrong.
> Interviewer: Why?
> Student 7: Because then I could learn something. But now I just thought 'they are reacting so foolishly, come on!'.

Implicit CF was preferred by two students, one of whom thought she was very competent in English, while the other was reading for a degree in languages. Interestingly, they argued that they preferred it because it was most fun, not because it helped them to learn.

> Student 6: The reactions. Yes, really.
> Interviewer: Why?
> Student 6: Ah, because it was more human. In real life that is the most evident reaction in a conversation. And that is what you'll have to do with. With the reactions, in real life. So you need to attend to that, in reality.
> Interviewer: So that was sufficient for you?
> Student 6: Yes, I also found that most fun. Most challenging.
> Interviewer: OK. And if you had to choose one of these three, for yourself? In your case it would be to learn, right?
> Student 8: Well in that case the implicit one. It was fun, when they responded so weirdly.

Further, some students preferred a combination of explicit and implicit CF, as they considered the rule-based CF most useful for learning, and the implicit CF to be most fun, challenging and attention-grabbing.

> Student 9: I think a combination of both would be best. If you didn't give the blue feedback [explicit explanations], then the person who played the game, or used the educational program, would just start the conversation over. He would just pick something else until it was correct, but he wouldn't know why. So the blue feedback is certainly necessary. On the other hand, when the NPC responds 'incorrectly', as here, that shakes you up a bit. When he is agitated.
> Interviewer: The fact that you have an impact on the world?
> Student 9: Yes, yes.

Further, others claimed that the CF type is best adapted to prior knowledge, and that attention and noticing played a significant role.

> It depends. It depends on the age. If it's more for secondary education, then I think it's better to have some explanation at the bottom. And the more you advance, the less explanation you need each time. Or maybe work with levels. When you reach a certain level, it's going better for you, and you don't need the explanation each time. (Student 5)
>
> Then I think that, if you make a full game, and it's all implicit, that it's a bit dangerous, whether it would always be understood. Or maybe in a first phase explicit, and when you have a better command, then implicit. Because then most of the times you know that it's a comment, because you've been pointed at it a few times. (Student 8)

3.5.4 Relation with intrinsic goal orientation, perceived competence and game experience

Pre-experimental intrinsic goal orientation, and perceived competence and game experience (measured afterwards) can be considered possible predictors of how students perceive feedback. As the three predictors had low inter-correlations (see Table III-2), they could be jointly used in the regression model to explain different aspects of the data.

Table III-2: Pearson's correlation coefficients between different self-perceptions

	1.	2.	3.
1. intrinsic goal orientation	1.00	.03	.11
2. perceived competence		1.00	.23
3. game experience			1.00

The three predictors jointly explained more than 20 percent of the variance of how students perceive explicit CF (R^2 = .21, F(3,79) = 6.97, p < .001). Perception of explicit CF was positively related to perceived competence (β = .04, p < .01) and game experience (β = .09, p < .05). The difference in perception was not significant for intrinsic goal orientation (β = .05, p < .1).

Upon removal of two outliers (Cook's distance higher than 2.5), the three predictors jointly explained 25 percent of the variance of how students perceive explicit CF (R^2 = .25, F(3,77) = 8.50, p < .001). Perception of explicit CF was positively and significantly related to all three predictors: perceived competence (β = .03, p < .05), intrinsic goal orientation (β = .06, p < .01), and game experience (β = .25, p < .05).

Implicit CF was regressed onto the same three predictors but none of the effects described above for explicit CF were found to be significant.

3.6 Discussion

For research questions 1 and 2, respectively, we found that students found the CF useful in general, and that they found immediate and explicit CF (containing metalinguistic explanation) more useful than and preferable to implicit CF (delivered through the characters' responses and designed to stimulate autonomous inquiry). However, these findings do not imply that implicit and more playful feedback types are irrelevant. In the interviews, several respondents replied that the implicit CF was fun and made them feel immersed. What seemed optimal for them was a combination of elaborate and immediate CF (type A) with feedback that is adapted to the game (type C). Such feedback can elevate learners' sense of immersion, which might increase their commitment to work through the CF in order to advance in the game. Thus, playful and creative feedback loops can complement explicit feedback mechanisms deemed effective for learning.

For research question 3, it was found that positive perceptions of explicit CF could be partly explained by three factors related to self-perception: intrinsic goal orientation, perceived competence, and game experience. This indicates that learners who consider themselves intrinsically interested in learning English as a foreign language, who felt competent while playing the game, and who had a positive experience playing this particular game had more positive perceptions of explicit CF. This finding is in line with our hypotheses that intrinsic goal orientation and perceived competence would explain positive perceptions of CF, but runs counter to the idea that learners who had a more positive experience of the game would have less positive perceptions of CF, especially of explicit CF. This tentatively suggests that in educational games, explicit CF could be most helpful for learners who are *a priori* intrinsically motivated, and that it might also contribute to the motivation of individual learners as the result of playing.

In the warm-up phase of the interviews, we probed students' general conceptions of feedback. They associated it with testing, exams, assignments, and scores (i.e., summative evaluation), rather than with meaningful information that would support their learning (i.e., formative evaluation). The finding that students generally did not conceptualize feedback as formative could partly explain their positive perception of explicit CF in the game, as it was elaborate, immediate, non-judgemental and helped them to do better on subsequent tasks. It is thus likely that, as a result of getting such CF and by using this information in the similar tasks that followed, students felt supported and competent in the game, and formed positive perceptions of such CF after playing. However, this finding is somewhat surprising; since students' average performance was quite low, they would have received a large volume of CF that might have them feeling less competent. So, either students' perceived competence was unaffected by the CF, or they improved as a result of such feedback. The relation between learner's self-perceptions and learners' in-game behaviour requires deeper investigation which is outside of the scope of this paper, but an additional regression analysis revealed that higher performance partially predicted positive perception of explicit CF (β = 1.8, p < .05) (R^2 = .05, F(1,78) = .075, p < .05).

3.7 Conclusion

Our findings can be summarized as follows. First, language learners generally found CF useful in an immersive educational game, and found implicit CF that lacks correct responses or metalinguistic explanation too weak for L2 learning. Secondly, individual difference factors related to learners' self-perception determined the perceived usefulness of and preferences for explicit CF in the immersive game (not so for implicit CF): learners who were intrinsically interested in learning English, who perceived themselves as competent during the game, and who had an enjoyable game experience had more positive perceptions of explicit CF (i.e., they found it useful and preferred it). Third, learners reported 'fun' and a sense of immersion when being confronted with CF that was implicit and adapted to the game (the characters' comments).

These findings have two implications. First, if we define 'intrinsic motivation in games' as an individual's subjective experience that is the combined result of enjoyable immersive gameplay with his or her positive perception of competence as the result of play, then instruction, including non-judgemental CF, need not necessarily get in the way of intrinsic motivation. This study thus provides evidence that the "dichotomy between overt instruction/guidance, on the one hand, and agentful immersion in experience is a false one" (Gee, 2007, p. 156). This has positive consequences for educational game design: instruction in games does not necessarily sacrifice 'fun', and designers should not shy away from including CF and other forms of instructional support as "overt verbal information [...] 'just in time' (when it is needed and can be used) or 'on demand' (when the player is ready for it and knows why it is needed)" (Gee, 2007, p. 156).

A second and potentially more crucial implication of our findings is that the effectiveness of feedback in game-based language learning might depend on how useful learners think it is, and on whether it stimulates intrinsic motivation. This is in line with the Cognitive Mediational Paradigm (Winne, 1987), which posits that the effectiveness of instruction is determined by a host of individual differences such as learners' intrinsic motivation and their perceptions of the (instructional) environment and its constituents. In this study, learners found elaborate and explicit CF with explanations most useful, especially if they were highly motivated, and reported to have learnt from it most. Further research should thus also study the impact of feedback on learning outcomes.

This study was limited in a number of respects. First, the target audience included mainly highly educated learners, whose intrinsic interest in and prior knowledge of English was relatively high. The motivation, learning strategies, and actual usage of CF can be quite different for less advanced learners (Brandl, 1995; Heift, 2002).

Furthermore, research on (corrective) feedback in CALL games is quite novel, so a more explorative method seemed appropriate. Consequently, a second limitation of the study is that all learners received the same kinds of feedback, which makes it difficult to say anything about whether discrete elements of feedback (such as error indication, metalinguistic information, or feedback that is adapted to the theme of the game) could actually affect learners' perceptions of instruction, their intrinsic motivation or their sense of immersion.

Therefore, future studies should first of all recognize that feedback in CALL games is a multidimensional construct, which needs to be taken apart in order to experimentally examine the effects of its constituents on learners' perceptions, motivation and learning outcomes. We propose that further research should distinguish by and large between, on the one hand, corrective feedback (and its different subcomponents) aimed at increasing a learner's understanding and, on the other hand, more 'game-like' feedback elements that can contribute to intrinsic motivation, namely positive feedback (designed to increase a learner's sense of competence) and situational feedback adapted to the game's theme (which can increase a sense of immersion). Various configurations of the constituents of feedback in a game-based language learning environment need to be implemented in different experimental conditions, so that the effects of feedback can be investigated directly. A key aspect that seems worthy of future research concerns the composite question (a) whether learners do actually process metalinguistic CF in games, which requires temporary time-outs from the flow of play; (b) whether this processing leads to the acquisition of explicit and/or (automated) implicit knowledge (through continued practice); and (c) what the complementary motivational role of positive and situational feedback might be in this respect.

Acknowledgements

This study is based on research funded by iMinds (Flanders' research centre that stimulates strategic research on digital innovation), and was conducted within the LLINGO project "Language Learning in an Interactive Game Environment" (iMinds-ICON 2009-2010).

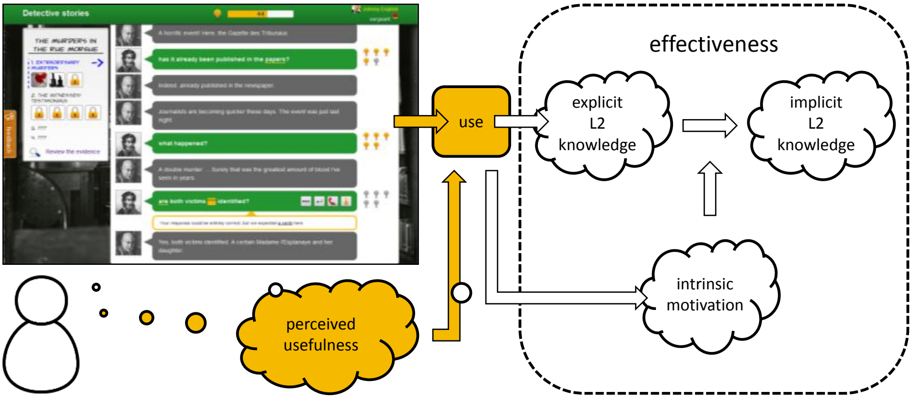
The PhD candidate designed the immersive 3D game as a proof-of-concept for goal-oriented language instruction and practice in immersive environments, together with colleagues from his research group. He developed the immersive educational space, a mod of the commercially available game *Divinity II: Ego Draconis*, with kind support from the Belgian game development company *Larian Studios*. The practice materials were developed in interaction with the language training company BLCC. The candidate wrote the article on which this chapter is based, receiving feedback from both his co-authors.

We are grateful to Dr. Wilfried Cools, who gladly offered his expertise in statistics, and prepared the game logs for subsequent analysis.

Second interlude

The previous chapter concluded with the finding that feedback in digital game-based language learning is complex, and that research needs to carefully unravel its different constituents in order to investigate the effects of feedback on cognitive and motivational learning outcomes. The third and fourth empirical study that compose this PhD project address this issue of effectiveness, and investigate the potential benefits of, respectively, vivid corrective feedback on learners' intrinsic motivation, and of metalinguistic corrective feedback on learners' development of explicit and implicit linguistic knowledge.

Before we turn to the effectiveness studies, however, we present the results of a study which investigated to what extent learners actually used corrective feedback in digital game-based language learning (effectively timing out from the flow of meaning-focused language play), and whether perceived usefulness of feedback could explain actual use of feedback.



Chapter IV

Empirical study 2: Learners’ use of corrective feedback in a written interactive murder mystery

This chapter was reformatted and slightly modified from:

Cornillie, F., Lagatie, R., Vandewaetere, M., Clarebout, G., & Desmet, P. (2013). Tools that detectives use: in search of learner-related determinants for usage of optional feedback in a written murder mystery. In P. Hubbard, M. Schulze, & B. Smith (Eds.), *Learner-Computer Interaction in Language Education: A Festschrift in Honor of Robert Fischer* (pp. 22–45). San Marcos, TX: Computer Assisted Language Instruction Consortium (CALICO).

Abstract

This paper seeks to identify individual difference factors as determinants of usage of optional metalinguistic corrective feedback (CF) in a written and task-based tutorial CALL environment for English grammar practice that contained gaming features. Previous research in CALL has highlighted the importance of prior knowledge for explaining learners' usage of CF options (Brandl, 1995; Heift, 2002), but the contribution of metacognitive and motivational variables to usage of CF remains unexplored. Based on insights from the literature on tool use (e.g. Clarebout & Elen, 2009), this study (N = 36) considered that learners' usage of optional CF in CALL might, in addition to prior knowledge, be determined by the perceived usefulness of CF and by learners' achievement goal orientation. Quantitative analysis of tracking and logging data in combination with questionnaire and language test data showed that usage of optional metalinguistic CF was associated with prior explicit L2 knowledge, but no relation was found with perceived usefulness and achievement goal orientation. Future research could benefit from fine-tuning the questionnaires used in this study, as well as from more qualitative in-depth analyses of learners' perceptions and motives. Also, in future studies game-like features could be implemented in different experimental conditions in order to investigate effects on learner behaviour.

4.1 Introduction

Across a wide range of theories in the field of Second Language Acquisition (SLA) and in second language (L2) pedagogy, feedback is increasingly being considered a developmentally useful feature of instructed L2 environments (for reviews see Lyster & Saito, 2010; Russell & Spada, 2006). Specifically, the notion of *corrective feedback* (CF), which may be succinctly defined as any utterance that is intended to correct a learner's erroneous response (for more comprehensive definitions and theoretical discussion see Carroll, 2001; Ellis, Loewen, & Erlam, 2006), has received substantive and intensified attention in SLA research, especially for the learning of grammar-related features. Although SLA theories are rather divided on the question of whether and how CF facilitates learning, research has altogether been guided by Steven Pinker's (1989) argument that CF can in principle support language development if the following conditions are met: 1) feedback needs to be available in the learner's environment, 2) it needs to be useful (i.e. psycholinguistically relevant), 3) it has to be actually used by language learners and 4) it must be "necessary" (i.e. the only feature that explains a specific change in L2 development) (pp. 9–14).

Pinker's first condition can be satisfied particularly in tutorial computer-assisted language learning (CALL) environments (Hubbard & Bradin Siskin, 2004), which can—notwithstanding the technological and pedagogical challenges involved—provide the learner with immediate, consistent, and error-specific feedback, possibly accompanied by additional help such as extended explanations on the nature of errors. However, the mere availability of such features in CALL programs does not imply that learners will actually use it (Fischer, 2007). Markedly, this is the case for support devices that are non-embedded and hence optional—i.e. the learner needs to click a button to get access to these devices—which are also known as *tools* in the more general research on computer-assisted learning (Clarebout & Elen, 2006) [1]. This research recognizes that learners do not always make the right choices for their learning, and argues for the investigation of the complex interplay of factors that might determine tool usage, viz. the nature of the tool, task characteristics, and learner-related factors, such as the motivation to work with particular tools and the functionality that learners attribute to these tools (Clarebout & Elen, 2009). Outside educational settings, use of technologies can be explained to a significant extent by how these technologies are perceived in terms of usefulness (Davis, 1989), and this line of reasoning seems well applicable to educational research which presupposes that learners' perceptions of instructional features mediate learner behaviour and learning processes (Winne, 2004).

This paper reports on a pilot study in a task-based tutorial CALL environment that includes gaming features and in which CF was available for responses that deviated from the predicted correct responses. The practical aim of the study was to prepare the learning environment and other instruments for a longitudinal experiment. In addition and more importantly, the study was intended to explore whether learners did actually make use of optional CF, and whether this usage was related to the perceived usefulness of the CF, to learners' explicit L2 knowledge, or—taking into account the achievement-oriented nature of gaming ecologies—to their achievement goal orientation. The instruments used in this exploratory study include, first, log files of the learners' interactions in the software, such as their usage of optional CF, secondly, tests to assess explicit L2 knowledge prior to practice and third, questionnaires to measure perceived usefulness of CF and achievement goal orientation. Before turning to the empirical study, we review the relevant literature, and present the conceptual framework that forms the backdrop of the study.

[1] Note that the notion of *tool* in this literature is different from Levy's (1997) conceptualization of the term in the field of CALL, and is more closely related to the *tool* and *monitor* functionality types of CALL applications described in Colpaert (2004).

4.2 Background research

In this section, we review two areas of research that are pertinent to this study: first, the research on the perceived usefulness and usage of CF in instructed L2 environments (including CALL research), and secondly, the literature on tool use in computer-assisted learning environments. We conclude the section with a summary of the findings of CALL research and with outstanding questions for research.

4.2.1 Perceived usefulness and use of CF in instructed L2 settings

In instructed L2 settings, there is ample evidence that language learners find CF generally helpful in a wide range of tasks. This finding applies both to feedback given directly by teachers and native speakers (e.g. Chenoweth, Day, Chun, & Luppescu, 1983; Radecki & Swales, 1988; Schulz, 2001) and to feedback generated by or mediated through technology (Cornillie, Clarebout, & Desmet, 2012; Nagata, 1993). Learners have also been found to prefer feedback that comprises metalinguistic explanations rather than less informative 'correct/incorrect' feedback (Nagata, 1993) or recasts (correct reformulations of erroneous utterances) (Kim & Mathes, 2001). In addition, research on meaning-focused L2 instructional settings indicates that learners would like to be corrected more than their teachers think is good for them (Magilow, 1999; Schulz, 2001), which reveals a discrepancy between students' and teachers' beliefs about the instructional goals of CF. Such discrepancies may be detrimental to the effectiveness of instructional designs.

Use of CF, which we will define here broadly as what learners do with or in response to CF, comprises diverse constructs, namely [1] noticing of CF, [2] uptake and [3], in CALL settings, use of optional CF. These constructs have been measured either through self-report instruments (such as stimulated recall and think-aloud protocols) or on the basis of behavioural data (including log files and eye-tracking data).

First, as for noticing, the research that taps into correction episodes in communicative interactions has gathered consistent evidence that feedback needs to be sufficiently explicit in order to be noticed. For instance, although the research on corrective recasts in L1 development has produced promising results and has spurred continued research in communicative L2 settings over the last decades (for a contrastive review see Nicholas, Lightbown, & Spada, 2001), L2 learners typically do not notice recasts if these lack perceptual salience (e.g. Lai & Zhao, 2006), if they are unsystematic (Nicholas et al., 2001) or non-contingent (i.e. if they do not immediately follow the erroneous response) (Lai, Fei, & Roots, 2008) or if they are long and involve many changes to the original utterance (Philp, 2003). Research on synchronous computer-mediated communication (CMC) has found that recasts which are textually enhanced are also associated with higher levels of awareness at the level of understanding than non-enhanced recasts (Sachs & Suh, 2007). In addition, it has been reported that learners have difficulty in identifying the linguistic focus of implicit CF types, especially for morphosyntactic features of the target language (Mackey, Gass, & McDonough, 2000), and that even in the case of explicit recasts learners notice semantic and syntactic problems more easily than morphological ones (Smith, 2012).

A second construct that can be put under the umbrella of CF usage is *uptake*, defined as learner utterances in response to CF (Sheen, 2011, pp. 7–8). It is important to note that uptake is considered evidence of whether CF has been noticed, not whether it has facilitated development (Mackey & Philp, 1998). Uptake is known to be facilitated especially by CF types that are explicit and/or contain detailed information, which comprise techniques such as elicitation, corrective repetitions, metalinguistic feedback (Heift, 2004; Lyster & Ranta, 1997) and explicit recasts (Sheen, 2006). Next, recasts in general have proven more successful for eliciting uptake of lexical features than for grammatical features of the L2, although teachers use them widely for correcting grammatical errors (Mackey et al., 2000; Sheen, 2006). Heift (2001) describes metalinguistic feedback strategies in tutorial CALL activities for L2 grammar development, and concludes from the preponderance of learners' *repair* movements (i.e. successful uptake) that learners attended to this feedback. In a similar research setting, Heift (2004) considered that uptake may be determined by two learner characteristics, viz. gender and language proficiency, but found no relation.

Third, CALL settings afford to investigate learners' use of CF options, i.e. non-embedded support that comes with CF. Feedback sessions in CALL environments (particularly in tutorial CALL) may provide the learner with options such as the possibility to see the location of the error, metalinguistic prompts, more extended grammar explanations and correct responses, which are all typically seen as making part of CF (R. Ellis et al., 2006). CALL research in this area has focused on the relation between usage of CF options and individual differences—this research focus is motivated by the desire to come to an understanding of what makes learners seek additional feedback. Heift (2006) showed that students' usage of context-sensitive grammar help following CF is contingent upon the level of detail in CF and upon proficiency level: learners that were confronted with less detailed immediate feedback and beginning learners tended to make more use of the error-specific help pages. Heift (2002) found that when introductory level university students of German were shown metalinguistic CF in tutorial grammar activities, the majority of students sought to correct errors mainly without relying on the use of correct answers. In addition, there appeared to be a relation between the learners' strategies and their performance as measured by the system: learners that peeked at correct answers frequently, either in response to CF or without submitting in the first place, were low- to mid-performers; students that generally attempted to correct errors themselves and sometimes requested correct responses were mid-performers; and the students that virtually never relied on the usage of correct responses ranged from mid to high performance. Heift's (2002) study corroborates previous findings by Brandl (1995), who concluded that students' previous performance in class determined their usage of feedback options in tutorial grammar activities: low achievers looked up correct answers more often, whereas high-achieving students showed more willingness to engage in the correction process. Brandl also hypothesized that the low-achieving learners lacked adequate cognitive and motivational processing, and consequently he made a plea for more research into the relation between learners' usage of CF and individual differences, in particular motivational variables. In the following section, we will review some of the more general literature on tool use in educational technology that might inspire CALL research in this area.

4.2.2 Tool use in computer-assisted learning

As indicated in the introduction, the recent research on tool use in computer-assisted learning has, in an attempt to come at a detailed understanding of tool use, begun to map the complex relation between characteristics of the learning task, tool features, and learner characteristics. Learner characteristics that are thought to determine the usage of tools include prior knowledge, metacognitive skills and knowledge (including conceptions about the usefulness of instructional interventions), the functionality which learners attribute to specific tools in specific learning environments and 'motivation' (Clarebout & Elen, 2006, 2009). The latter two constructs may both be considered *perceptions*, i.e. perceptions about the usefulness of tools (perceived usefulness), and perceptions about the learner's self, respectively. Perceptions are thought to emerge in the dynamic interaction between the learner and his or her environment, and are typically measured by means of self-report (questionnaires and/or interview data).

The construct of *perceived usefulness* originates in expectancy theory and is central in Davis' technology acceptance model (TAM) (1989), which posits that users' behaviour (i.e. their use of technology) can be predicted by how useful they find the system and how easy they find it in actual usage, with perceived ease of use functioning as a causal antecedent of perceived usefulness. In the educational technology field, Lust et al. (2011) found evidence for the explanatory power of perceived usefulness with respect to students' actual use of webcasts in a blended learning course.

Next to perceptions about usefulness and ease of use of tools, learners' perceptions of themselves might hint at their motivation for using (or not using) specific tools. This reasoning is reflected in current research on help-seeking (for a review see Aleven, Stahl, Schworm, Fischer, & Wallace, 2003), a line of educational research related to tool use which has started to investigate the relation of help-seeking strategies with learners' *achievement goal orientation*. Achievement goals are typically bifurcated into *mastery goals* (also known as *learning goals*), which comprise 'intrinsic' goals focused on the development of competence or task mastery, and *performance goals*, which constitute a more extrinsic goal orientation, viz. demonstrating competence (e.g. relative to peers) rather than developing it (Elliot, 1999). Achievement goal theory assumes that mastery goals are associated with positive learning processes and outcomes (such as persisting through failure), whereas a performance orientation would lead to less favourable behaviour and outcomes (e.g. lower effort in the face of failure, or surface processing of useful pedagogical materials). Along these lines, research on help-seeking has gathered evidence that mastery goals are typically associated with *instrumental help-seeking* (intended to promote learning, such as making use of hints), whereas performance orientation is more likely to be linked to *executive help-seeking* (intended to avoid work, such as peeking at correct responses) (Aleven et al., 2003). In addition, design features of learning environments may change learners' achievement goal orientation, and subsequently such features could influence how individual learners seek help, e.g. by emphasizing performance or interpersonal performance comparisons (Karabenick, 2011). If learners' use of optional CF can indeed be seen as related to instrumental help-seeking, then the literature on help-seeking may provide fertile theoretical models for individual difference research on the usage of optional CF in CALL. Specifically, it may help to explain how use of optional non-embedded feedback is driven by achievement goal orientation.

4.2.3 Summary and outstanding questions

To summarize, previous studies in the SLA literature have found that learners' usage of CF is determined both by characteristics of CF and by individual differences. First, explicit and detailed feedback is more likely to facilitate noticing and uptake. Second, CALL research shows that students' prior knowledge plays a role in how they engage with optional detailed feedback: beginning learners seem to request optional context-sensitive feedback more frequently than do advanced learners. And third, also learners' usage of correct response feedback has been shown to depend on their prior knowledge, or on their performance: low-achieving students look up correct responses more often than they work through feedback that does not give away the correct answer (i.e. output-prompting feedback). Thus, these learners may be considered to engage in executive help-seeking rather than in more independent problem-solving.

Hence, the outstanding question is why learners do or do not make use of certain feedback options. Of particular relevance is the case of weaker learners that resort to looking up correct responses, and hence make less frequent use of output-prompting feedback options in order to complete tasks. Potentially, these learners do not find output-prompting feedback useful, as they might lack the knowledge to cognitively process (meta-)linguistic explanations. Or, low-achieving students may lack "motivational processing" (Brandl, 1995, p. 207) to deal with such detailed feedback. Thus, these three variables, i.e. prior knowledge, perceived usefulness of CF, and the broader construct of 'motivation', may be seen as determinants of learners' use of optional non-embedded CF.

4.3 The current study

The data for the current study were collected as part of a pilot study, which had a few practical aims. First, we wanted to evaluate whether students found the metalinguistic prompts at all usable. Secondly, we wanted to check whether the self-report instruments were reliable for use in future experiments. A third aim was to collect typical responses from learners in order to expand the domain model of the tutoring system, i.e. to populate the content database on the basis of learner language with additional and more evidence-based instances of grammatical and ungrammatical responses, and to evaluate the accuracy of the string matching algorithm (an investigation which is beyond the scope of this paper). Another objective was to evaluate the gaming features (positive feedback specifically), in consideration of providing learners with different positive feedback types (or none at all) in future experiments. As a final practical aim, we wanted to evaluate the technology in the setting of a typical secondary school classroom in order to detect potential performance problems with the software.

Additionally, and more importantly, the study aimed to explore the question of why learners use optional non-embedded feedback in CALL materials or refrain from using it. In addition to prior knowledge, which had already been shown influential in learners' use of optional feedback in CALL settings (see section 4.2.1), we considered perceived usefulness (of optional metalinguistic CF) and achievement goal orientation—as an operationalization of 'motivation'—as potential determinants of learners' usage of optional metalinguistic CF. The construct of perceived usefulness was chosen since it had been previously identified as a significant predictor of tool usage (e.g. Davis, 1989; Lust et al., 2011). We defined 'motivation' as achievement goal orientation (Elliot, 1999), because the learning environment which we wanted to evaluate comprised features that stress achievement (see section 4.4.1) and may thus intensify differences between mastery-oriented learners and those that are primarily performance-oriented. Learners with the latter orientation are known to more frequently show executive help-seeking behavior (such as peeking at correct responses), whereas the former learners (oriented towards learning) would seem to take an interest in solving problems independently by attending to and working through detailed linguistic feedback.

In addition to considering that the three key individual difference factors described above may explain use of optional metalinguistic CF, namely prior knowledge, perceived usefulness of CF and achievement goal orientation, we also need to recognize that the perceived ease of use of this CF might be determined by prior knowledge, more particularly explicit L2 knowledge. Carroll (1995) notes that CF is language about language, and is thus “quintessentially metalinguistic in nature” (p. 76)—this is irrespective of whether the CF includes metalinguistic information. This implies that learners need to be equipped with explicit (metalinguistic) L2 knowledge in order to decode feedback instances, and that, hence, their explicit L2 knowledge might determine partly how easy it is for them to learn from CF, in addition to e.g. usability issues in interface design. Figure IV-1 summarizes the conceptual framework that forms the backbone of the current study, and shows the main targeted variables (highlighted in grey) with their hypothesized interrelations. For the sake of completeness, we also include variables related to the broader notion of ‘use of CF’ (as defined in the literature review above), namely noticing of CF and uptake, which will however not be investigated in this study.

Further, we hypothesized that the perceived usefulness of the CF may be determined by the perceived difficulty of the task, taking into account the challenge for learners to construct responses that fell in the scope of predicted utterances in these semi-open activities; we predicted that the CF would not be found useful if the task was too difficult. Next, we also considered that frequent executive help-seeking strategies (requesting hints and peer responses) could be associated with a performance goal orientation. A final aim was to empirically explore the relation between usage of optional CF, use of hints, use of peer responses and attempts per exercise.

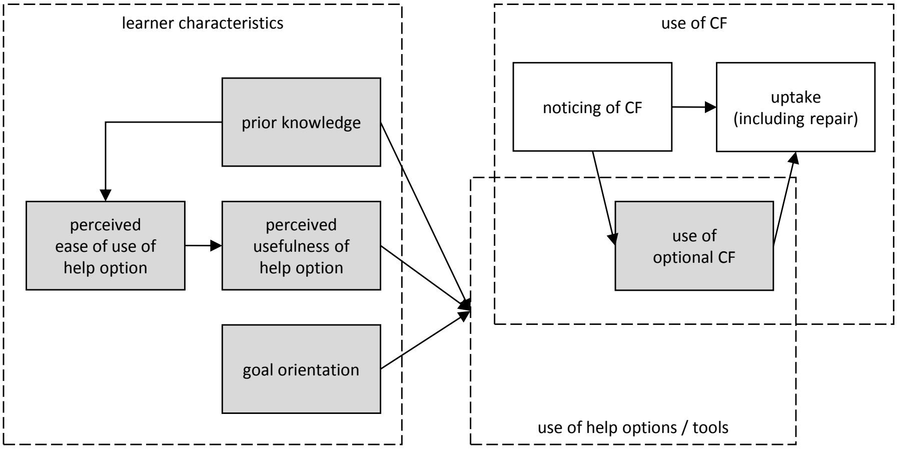

learner characteristics
prior knowledge
perceived ease of use of help option
perceived usefulness of help option
goal orientation
use of CF
noticing of CF
uptake (including repair)
use of optional CF
use of help options / tools

Figure IV-1: conceptual framework

The research questions were as follows:

1. How useful do learners find the CF? How is the perceived usefulness of CF related to its perceived ease of use, and to the perceived difficulty of the task?
2. How is perceived ease of use of CF related to prior explicit L2 knowledge?
3. How frequently do learners use the optional CF? How is this usage related to the perceived usefulness of CF, to prior explicit L2 knowledge and to achievement goal orientation?
4. How is the usage of hints and peer responses related to achievement goal orientation?
5. What is the relation between usage of the optional CF, use of hints, use of peer responses and attempts per exercise?

4.4 Method

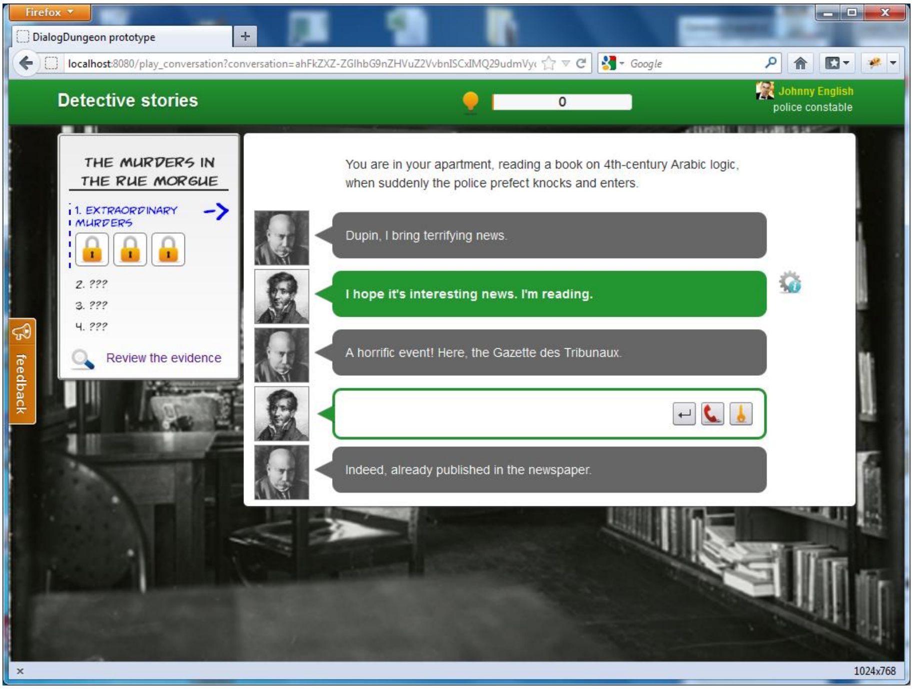
4.4.1 Description of the learning environment

The learning environment used for the study was a prototype of an online task-based tutorial CALL system for grammar practice in which learners played the role of a detective in 'semi-open' (Desmet, 2007) written activities, and had to solve a murder mystery by formulating responses that fell within the range of predicted correct utterances. It utilized natural language processing (NLP) and crowdsourcing techniques to generate explicit embedded CF and non-embedded options that were deemed necessary to perform the tasks. The learning environment also contained features associated with gaming such as positive feedback. In this section, these features will be described in detail.

The learning environment was task-based (R. Ellis, 2003), as it was intended to capture learners' interest by confronting them with a meaning-focused problem which required them to work towards a non-linguistic outcome (i.e. solving the murder mystery through dialogue tasks), but on a lower level the activities involved writing responses to grammatical exercises integrated in the dialogues (hence also implying a strong focus on form). Although the unit of response was at the level of the utterance, which implies that many alternatives are possible, the range of appropriate utterances for particular exercises was constrained: first, by the immediately preceding and following utterances in the linear dialogues, which were provided by so-called *non-player characters* in the story (see Figure IV-2), and secondly by four grammatical topics in English, which are notoriously difficult for Dutch-speaking learners and for which errors are known to persist even in the speech of fairly advanced learners (Tops, Dekeyser, Devriendt, & Geukens, 2001). The grammatical topics and distribution of these topics in the exercises are shown in Table IV-1.

Table IV-1: grammatical topics of the exercises

grammatical topics	**number of exercises**
past time reference: simple past vs. present perfect	20
the quantifiers *some* and *any*	5
modal verbs for ability, possibility, deduction	5
future tenses: *will* and *going to*	1
(liaising interactions)	2
TOTAL	33



You are in your apartment, reading a book on 4th-century Arabic logic, when suddenly the police prefect knocks and enters.

Dupin, I bring terrifying news.

I hope it's interesting news. I'm reading.

A horrific event! Here, the Gazette des Tribunaux.

Indeed, already published in the newspaper.

Figure IV-2: ‘semi-open’ written activities in murder mystery dialogues (learner’s role in the even speech bubbles)

A string matching algorithm matched learners’ responses to a number of predicted responses for the grammatical exercises and for other ‘liaising interactions’ (i.e. sentences that did not target specific linguistic problems but that simply moved the dialogue forward). For the grammatical topics, both correct and incorrect utterances were predicted; the number of predicted responses for these interactions ranged between 1 and 19 (M = 5.5), depending on the scope of the exercises.

The string matching algorithm utilized relatively simple string matching techniques (Levenshtein distance; see e.g. Lagatie & De Causmaecker, 2010) in combination with NLP (part-of-speech tagging and lemmatization) to compute the similarity between the learner’s response and each of the predicted responses. The outcome of the analysis included the closest match and the closest correct match, which were used to calculate a score and to generate immediate and explicit utterance-specific CF, as well as linguistic annotations at the level of the individual tokens in the learner’s utterance.

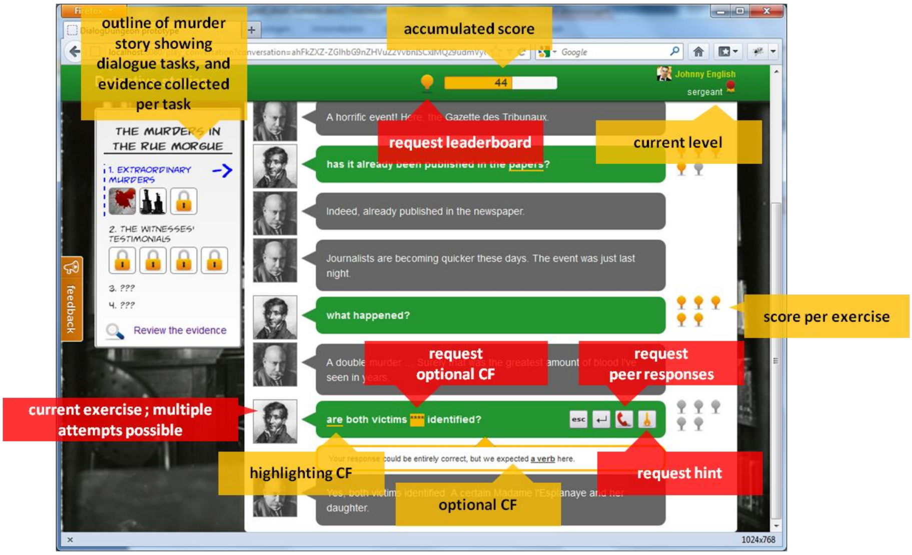
The utterance-specific CF comprised a visualization of the breakdown of the string matching procedure (see Figure IV-3), which was tested and improved in several iterations on the basis of experts' comments. For each attempt the learner's response was compared with the 'closest match': a predicted response that matched the learner's response the most closely. If the similarity was below a threshold of .5 (i.e. the learner's response was less than 50 per cent similar to the closest predicted utterance), the system showed an icon with a tooltip which explained that the response had not been recognized and that the learner could try again (see first interaction in Figure IV-2). If the similarity was above .5 then the CF visualization routine underlined each of the tokens in the learner's response that deviated from the corresponding token in the aligned version of the closest *correct* match. If the string-matching algorithm had not found a corresponding token in the aligned version of the closest correct match, then the CF visualization routine would simply show asterisks in that position. In a nutshell, and put more simply, the string matching algorithm detected words that were different from predicted words, words that were superfluous and words that were missing (Desmet, 2007); the CF visualization routine showed the location of the (potential) error in the sentence by means of underlining, a technique highly similar to highlighting CF in CALL and elicitation/repetition in classroom settings (Heift, 2004).

On top of the highlighting CF that was shown immediately after the learner's response, non-embedded prompts were available for the dissimilar tokens. These prompts were based on the part-of-speech and lemmatization analyses and included mainly metalinguistic terminology which we thought could increase the chance that learners would be able to correct responses (equivalent to "You might need a modal verb in this position."), as well as encouraging and (arguably) humorous statements in line with the detective metaphor (such as "Oops, we expected nothing here. Try not to waste any words; detectives use as few words as possible.") [2]. Taking into account a

[2] For this feedback message, we acknowledge our indebtedness to a similar design feature in Sanders' and Sanders' (1995) classic tutorial CALL system *SPION*.

pedagogical framework for the design of game-like activities for language learning (Purushotma et al., 2008), it was decided that this optional feedback would only be shown on the learner's request and with varying degrees of specificity, dependent on the dissimilarity between the tokens (due to combinatorial differences with respect to lemma and part-of-speech). So, while feedback is generally embedded and hence not considered a tool (i.e. a non-embedded support device) (Clarebout & Elen, 2006), this additional feedback was available as one: learners had to click on the highlighted tokens in order to get access to context-specific detailed feedback.

In addition to the optional CF, other non-embedded help options were available. Based on usability testing, we assumed that the exercises would be quite challenging, especially the task of finding which words to use. The learners would be told they could find many words and chunks to be used in their responses 'hidden' in the utterances of the non-player characters, but still we decided to include two additional help options: access to the responses of others (represented as a red phone to call other detectives)—the learners seemed to like this idea—and an option to request a hint (represented as a key). The first help option is a *crowdsourcing* feature that showed the responses of other learners ranked by frequency (computed on the basis of the number of times the utterance had been submitted or chosen). When a peer response was chosen, it was evaluated in the same way as a free response (see above), so learners would also get CF if this response deviated from the predicted grammatical responses. Each time the second option was clicked, learners would receive a target word they could use in their response, and the option could be used until all words for a sentence had been disclosed. Learners would not lose any points for using any of these two options, and they were available both before and after each attempt.



A horrific event! Here, the Gazette des Tribunaux.

has it already been published in the papers?

Indeed, already published in the newspaper.

Journalists are becoming quicker these days. The event was just last night.

what happened?

A double murder ... Surely that was the greatest amount of blood I've seen in years.

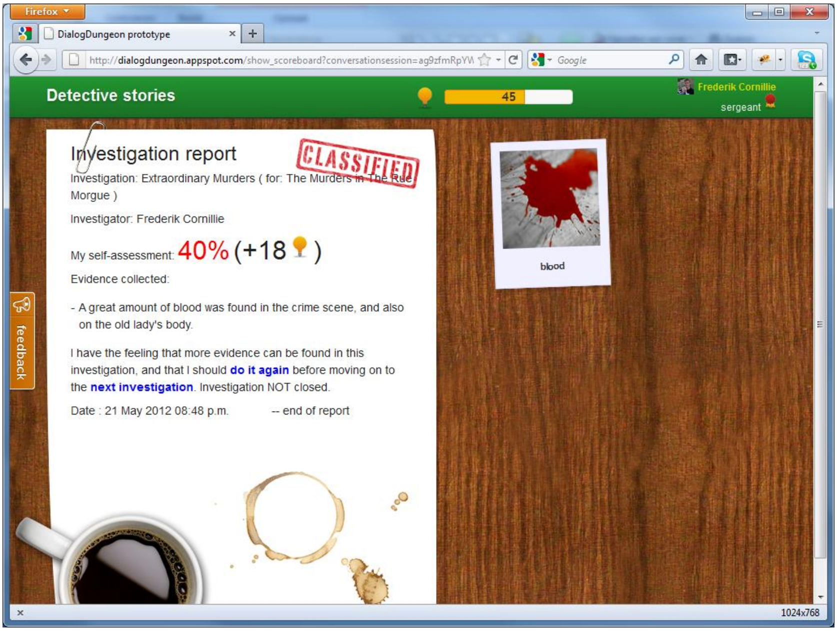
are both victims **** identified?

Your response could be entirely correct, but we expected a verb here.

Yes, both victims identified. A certain Madame l'Esplanaye and her daughter.

Figure IV-3: task screen with main design features (features in red, e.g. “request optional CF”, are non-embedded and afford measurement of learner behaviour)

Finally, the learning environment also contained features associated with gaming. First, the format of the learning environment is related to that of the *interactive participatory drama*, a genre which has a long history in CALL and which has been discussed in game-based learning venues (Hubbard, 2002). In acknowledgement of the skills and effort required to write an engaging drama that would withstand the inclusion of grammar practice activities, the first author wrote a story on the basis of Edgar Allan Poe’s *The Murders in the Rue Morgue*. Secondly, the system computed scores for the learners’ individual responses, and represented these scores as “ideas” (light bulbs), adapted to the detective metaphor (see Figure IV-3). Next, after each task/dialogue, learners would be shown a debriefing screen with meaningful task outcomes in the form of pieces of evidence on the murder mystery; the number of depended on how many “ideas” they had gathered during the task (see Figure IV-4). Moreover, accumulating points over time would result in increases in the learner’s level in the game: they would start out as constable, then become inspector, etc., and eventually end up as superintendent. The level increase thresholds were balanced and play-tested in order to make sure that learners in our target audience would actually experience these *level-ups* while working through the tasks. Additionally, there was a *leaderboard* feature: learners could request the names and scores of the 10 highest-ranked peers by clicking on their own accumulated score. Finally, as is customary in game design, all of these instances of "positive feedback" were visualized rather excessively (Juul, 2010, p. 45)—episodes containing perfect responses and level-ups would include messages of verbal praise such as "Bravo!", and were animated using the jQuery library for HTML5-compliant websites.



Investigation report

Investigation: Extraordinary Murders (for: The Murders in The Rue Morgue)

Investigator: Frederik Cornillie

My self-assessment: 40% (+18)

Evidence collected:

- A great amount of blood was found in the crime scene, and also on the old lady's body.

I have the feeling that more evidence can be found in this investigation, and that I should **do it again** before moving on to the **next investigation**. Investigation NOT closed.

Date : 21 May 2012 08:48 p.m. -- end of report

Figure IV-4: debriefing screen with meaningful task outcomes

4.4.2 Participants, procedure, and instruments

The study was carried out in May 2012 in two classes in a secondary school in Kortrijk, Belgium. 36 Dutch-speaking high-intermediate learners of English in the 5th and 6th form of the Modern Languages programme were invited to the study through their teachers. There were 29 girls and 7 boys, and although we did not have their exact ages, these would typically be in the 16-18 year range.

One week before the learners worked with the online learning environment, they filled out an English grammar test that was intended to measure their prior explicit L2 knowledge. The test included adapted versions of the metalinguistic knowledge test (MKT) and grammaticality judgment test (GJT) published in Ellis (2009b). These tests cover a wide range of grammatical structures (17 in total) that are known to be universally problematic to learners of English as L2 at various stages in their development, and may hence provide a representative performance measure of linguistic knowledge and ability. In the MKT, participants are presented with ungrammatical utterances for these structures, and are required to select for each utterance the rule that best explains the error out of a list of four options. As the MKT draws heavily on learners' knowledge of metalanguage, involves a high degree of awareness, and focuses attention on form, it is considered to be a measurement of learners' explicit L2 knowledge. The grammatical structures that form part of the MKT are also included in the GJT. In this task, participants decide whether utterances are well-formed or deviant. The GJT may measure implicit or explicit L2 knowledge, depending on the conditions of the task, more particularly the time learners are given to make the judgments, and the nature of the stimuli. If participants are given only limited time to respond, they may need to rely primarily on their implicit (intuitive) L2 knowledge. Conversely, if they have unlimited time to judge utterances, especially ungrammatical ones, it is more likely that they are relying on explicit L2 knowledge (Loewen, 2009). Ellis (2009b) found very strong significant correlations between the MKT and the ungrammatical items on the GJT if the latter was untimed, which suggests that learners' performance on the ungrammatical items of the untimed GJT also provides a measure of their explicit L2 knowledge.

For this study, we selected the MKT and the untimed version of the GJT , as we presumed that learners' explicit L2 knowledge might affect how they would use optional metalinguistic CF. Also, these tests were chosen, rather than other L2 knowledge tests, as they had been found reliable in a previous study (R. Ellis, 2009b). For time constraints, the GJT was reduced to 34 utterances (half the size of the original test) but covered the 17 grammatical topics of the original test, and we also left out items that measured participants' self-reported use of rules and how certain they felt about their responses. For the MKT, we used only the first part, which consists of 17 multiple-choice items that target knowledge of rules and which requires understanding of metalinguistic terminology. For both tests, the items and terminology were slightly altered to more accurately reflect how the grammar had been taught in class; these adjustments were done on the basis of comments from one of the participating teachers.

In the next session, one week later, the learners worked in the online environment in a computer room at school. The first author guided learners through the environment by means of a slideshow with screenshots of its key features, emphasizing the feedback and other help options, and briefly reviewed the grammar rules of the four topics that the learners were about to practise. The final slide contained a summary of this walkthrough including the suggestion to make use of the feedback and other help options, which remained visible while learners were working through the tasks. All instructions were given in English. The researcher and the teachers helped the learners if they experienced technical difficulties, but did not intervene for problems with grammar.

The participants were not familiar with the murder mystery, and it was the first time they worked in the environment. Four dialogue tasks were available in the story, through which the learners would navigate in a fixed order. Each subsequent task would only become available after the learners had finished the previous one, and learners were allowed to repeat tasks if they wanted to. Each dialogue task contained between 7 and 9 exercises. The number of attempts for each exercise was not restricted; the system would only move on to the next exercise automatically if the score was perfect. The metalinguistic terminology in the non-embedded CF prompts (e.g. "infinitival *to*", "linking word", "modal verb") had been adapted to the terminology used in the learners' grammatical compendium in their course book.

In line with recommendations from Fischer (2007) to put the study of learners' reliance on help features in CALL on more solid empirical footing, all learners' actions (except keystrokes) were logged. Most learners used a personal online account (Facebook or Gmail) to log into the system; three learners who did not have an online account or who had difficulties logging in used one of the spare accounts (Gmail) provided by the researcher. The first class practised the grammar problems in the murder mystery for about 40 minutes; due to time constraints the second class only practised 25 minutes. Because not all learners completed the four tasks, large differences were expected between learners for the number of completed tasks, completed turns and number of attempts. This needed to be taken into account for the analyses (see the end of section 4.5.1).

After having worked in the online environment, the learners filled in a questionnaire (in their mother tongue) which contained four sections with 7-point Likert-scale items. The first section targeted learners' achievement goal orientation while working in the environment, and was based on and translated into Dutch from the 3x2 achievement goal model (Elliot, Murayama, & Pekrun, 2011). It consisted of items on 6 goal constructs: task-approach and task-avoidance (i.e. goals focused on learning), self-approach and self-avoidance (i.e. performance goals focused on improving previous performance) and other-approach and other-avoidance (i.e. performance goals focused on outperforming others). The second section of the questionnaire was based on the TAM (1989), and comprised items on the perceived ease of use of the optional metalinguistic CF as well as items on its perceived usefulness; a screenshot was included to remind learners of the immediate and optional feedback shown in the exercises (see Appendix 2). For each construct we included multiple items, in order to provide reliable and valid measures for each of the constructs involved (Dörnyei, 2003b). To further improve reliability, some items in the TAM scale were negatively worded in order to be able to detect participants that would consistently pick e.g. a 6 or 2 without reading the items. Table IV-2 shows an overview of the constructs, and for each construct the number of items included in the questionnaire as well as an example item. In the final two sections, learners filled in additional single 7-point Likert-scale items on how motivating they had found particular features in the environment (the murder mystery, the collection of evidence, the accumulated score, leveling up, the leaderboard, collaborating with others, the corrective feedback), and how difficult they had found the task (in general, with respect to finding the right words, with respect to the grammatical problems; one item for each construct).

Table IV-2: overview of the constructs targeted in the questionnaire

scale	construct	number of items	example item
3*2 achievement goal model	task approach	3	I found it important to get a lot of questions right.
	task avoidance	3	I tried to avoid getting a lot of questions wrong.
	self approach	3	I tried to do well relative to how well I did on previous attempts and exercises.
	self avoidance	3	I tried to avoid performing worse in comparison with previous attempts and exercises.
	other approach	3	I tried to do better than my peers.
	other avoidance	3	I tried to avoid doing worse than my peers.
technology acceptance model	perceived ease of use	6	I found it difficult to understand the grammatical feedback.
	perceived usefulness	6	The grammatical feedback helped me to learn why I had been wrong.

4.5 Results

4.5.1 Data preparation and preliminary analyses

We tested the reliability of the grammar tests using the Kuder Richardson's Formula 20 (KR-20), which calculates the homogeneity of a test with dichotomous measures (Kuder & Richardson, 1937). Tests with a KR-20 coefficient over .90 are considered homogeneous and hence reliable. Analysis revealed that the grammar tests did not reliably measure learners' explicit L2 knowledge (KR-20 α = .02 for the ungrammatical items in the GJT; KR-20 α = .04 for the MKT). Inspection of the sum scores for the tests indicated that overall, the learners' scores on the tests were rather high (for the ungrammatical items in the GJT: M = .91, SD = .06 ; for the MKT: M = .74, SD = .10). Possibly the low reliability of the grammar tests was due to the fact that there was little variation between the learners in terms of their explicit L2 knowledge. Yet, we decided to keep the tests in our analyses for two reasons: they had been carefully constructed and found reliable in previous studies (R. Ellis, 2009b), and they had been adapted to the learners' curriculum with the help of one of the teachers, so they would reflect how grammar was typically taught and assessed in class.

The self-report measures (questionnaires), on the other hand, were found to be reliable. For the questionnaires, we computed Cronbach's α, which is a measure of the internal consistency of a scale or any of its subscales/constructs (Cronbach, 1951). Scales with α coefficients above .70 are considered reliable, whereas scales with α coefficients below .60 should be treated with caution (Dörnyei, 2003b). For achievement goal orientation Cronbach's α was between .78 and .96 for the items on the subscales; for the TAM Cronbach's α was .95 for the items on perceived usefulness and .84 for the items on perceived ease of use. Hence, we considered the subscales for the constructs related to learners' perceptions reliable. The subscales were created by taking the means of the corresponding items.

Inspection of the correlation matrix for the subscales of the achievement goal orientation questionnaire (see Table IV-3) revealed significant positive intercorrelations between many of the subscales, particularly between task- and self-focused goals, which suggest that the latter goals "emerge from very similar dispositions in general" (Elliot et al., 2011, p. 641). Other-approach goals were unrelated to the former goals; this may have emerged from the fact that the learning environment stressed interpersonal comparisons through the leaderboard. Hence, for further analysis, we chose two approach goals that were unrelated to each other: task-approach, focused on learning, and other-approach, focused on outperforming others.

Table IV-3: Pearson’s correlation coefficients for achievement goal motivation, adjusted for multiple comparisons using Holm’s method (** p < .01 ; * p < .05)

Variable	1.	2.	3.	4.	5.	6.
1. task approach	—	.84**	.76**	.74**	.28	.46*
2. task avoidance		—	.74**	.80**	.32	.57**
3. self approach			—	.88**	.25	.46*
4. self avoidance				—	.32	.48*
5. other approach					—	.84**
6. other avoidance						—

As for the log data, as was expected, there were large differences between learners with respect to the amount of completed tasks, completed turns and attempts, which reflects differences in the amount of time which the learners had spent in the online environment. So, after extraction, the log data were normalized in order to even out these differences. All learners consulted the optional CF at least once. Usage of optional CF was calculated by dividing the number of times learners had clicked a token to see metalinguistic prompts by the amount of highlighted tokens for which a metalinguistic prompt had been available. Next, learners’ usage of hints and their usage of others’ responses were computed by dividing the number of times they had requested these help options by their individual total number of attempts on the level of the exercises.

4.5.2 Findings

As for research question one, the results indicate that the majority of learners found the CF quite useful, with a median of 4.75 (M = 4.51 ; SD = 1.12), scores ranging between 2 and 6.67, and 69% of the participants scoring 4 or higher on the perceived usefulness scale. Further, correlation analyses (see matrix in Table IV-4) showed a significant positive correlation between perceived usefulness of CF and perceived ease of use of CF, and a significant negative correlation between perceived lexical difficulty and perceived ease of use of CF (r = -.42 ; p = .05). The latter finding indicates that learners who experienced less difficulty in getting the words right in the dialogue exercises found the CF more easy to use, and vice versa.

Table IV-4: Pearson's correlation coefficients for research question 1, adjusted for multiple comparisons using Holm's method (** p ≤ .01 ; * p ≤ .05)

Variable	1.	2.	3.	4.
1. perceived usefulness	—	.62**	-.40	-.11
2. perceived ease of use		—	-.42*	-.31
3. perceived lexical difficulty			—	.45*
4. perceived grammatical difficulty				—

The second research question focused on the relation between perceived ease of use of CF and prior explicit L2 knowledge as measured by the MKT and by the ungrammatical items on the GJT. No significant correlations were found between these variables (see Table IV-5).

Table IV-5: Pearson's correlation coefficients for research question 2, adjusted for multiple comparisons using Holm's method (** p ≤ .01 ; * p ≤ .05)

Variable	1.	2.	3.
1. perceived ease of use	—	.03	-.10
2. GJT (ungrammatical items)		—	.07
3. MKT			—

For the third research question on usage of optional CF, descriptive statistics show high variability between learners in terms of their usage of CF (M = .40 ; SD = .26 ; range between .03 and 1.02). Correlation analyses for the third research question (see Table IV-6) shows that the only variable significantly related to use of optional CF is explicit L2 knowledge as measured by the MKT—the direction is positive, and the association is strong, with a correlation size of .59. This signifies that learners with higher explicit L2 knowledge used the optional metalinguistic CF more often.

Table IV-6: Pearson's correlation coefficients for research question 3, adjusted for multiple comparisons using Holm's method (** p ≤ .01 ; * p ≤ .05)

Variable	1.	2.	3.	4.	5.	6.
1. use of optional CF	—	-.08	.59**	.03	-.02	.03
2. perceived usefulness		—	-.05	-.23	.44	-.16
3. MKT			—	.09	.15	.26
4. GJT (ungrammatical items)				—	-.13	.02
5. task approach					—	.27
6. other approach						—

As for research question 4, no significant relations were found between use of hints and peer responses on the one hand, and achievement motivation on the other hand (see Table IV-7).

Table IV-7: Pearson's correlation coefficients for research question 4, adjusted for multiple comparisons using Holm's method (** p ≤ .01 ; * p ≤ .05)

Variable	1.	2.	3.	4.
1. use of hints	—	.16	-.12	-.21
2. use of responses from peers		—	.00	.26
3. task approach			—	.28
4. other approach				—

The results of the correlation analyses for research question 5 reveal significant and medium-sized negative correlations between use of CF and use of hints (r = -.38 ; p = .05) and between use of CF and use of responses from peers (r = -.39 ; p = .05) (see Table IV-8). This suggests that students who requested the additional CF more often used both hints and responses of others less often, and vice versa. In addition, there are significant correlations between the number of attempts per exercise and use of help options: the relation goes in the negative direction for usage of hints and use of responses from peers; the relation is positive for use of CF. Thus, we may derive that learners who made more frequent use of the hint and peer response options were engaged to a lesser extent in the correction process, whereas learners who used the optional metalinguistic CF more often showed more willingness to correct their responses autonomously.

Table IV-8: Pearson's correlation coefficients for research question 5, adjusted for multiple comparisons using Holm's method (** p ≤ .01 ; * p ≤ .05)

Variable	1.	2.	3.	4.
1. use of CF	—	-.38*	-.39*	.49**
2. use of hints		—	.14	-.46*
3. use of responses from peers			—	-.58**
4. attempts per exercise				—

4.6 Discussion and conclusion

This study was aimed at identifying individual difference factors that might explain learners' use of optional metalinguistic CF. First, we found that use of CF was related to prior explicit L2 knowledge as measured by the metalinguistic knowledge test described in Ellis (2009b): more 'advanced' learners made more frequent use of the optional metalinguistic CF. Given the low reliability of the grammar test in this particular context, this finding needs to be taken with care. Nonetheless, it seems to be consistent with previous research in CALL which has found that 'high-achieving' learners make more use of output-prompting CF (Brandl, 1995; Heift, 2002), but contradicts Heift's (2006) findings: there, language proficiency was inversely related to usage of grammar help following feedback. Theoretically, one might expect the latter, namely that more advanced learners would need less feedback than less advanced learners in order to successfully complete a specific task—this prediction has been articulated most clearly by sociocultural approaches to feedback use (e.g. Aljaafreh & Lantolf, 1994). Hence, our finding is somewhat hard to interpret in the face of current research and theory, but it may be explained by our specific operationalization of language proficiency as explicit (metalinguistic) L2 knowledge: learners equipped with explicit (metalinguistic) L2 knowledge might have been more able to decode the optional metalinguistic feedback, and hence used it more. Future research could, however, include other language tests as measures of prior knowledge besides tests of metalinguistic knowledge and grammaticality judgment tests in order to yield a more complete picture of the learners' proficiency levels.

On the basis of our findings, we can neither confirm nor deny that use of optional CF depends on its perceived usefulness, a prediction made by the TAM. The lack of relation between perceived usefulness and actual use is in line with previous findings in CALL, and can be explained by the hypothesis that learners' perceptions may be inaccurate with respect to actual learning processes and outcomes, and that "researchers should be ever mindful of the discrepancy between statements of learners' perceptions/beliefs and their actual behaviors" (Fischer, 2007, p. 427).

This explanation, however, makes it difficult to put the widely supported TAM to the empirical test in educational settings. In this particular case, another explanation might be in order: learners might have found the embedded CF (highlighting) sufficiently useful, so that they did not need to use the non-embedded CF. This explanation is supported by the observation that in general, the participants were quite advanced learners of English. For these learners, possibly, the optional CF did not add much information to the embedded CF, save for a hint on which part of speech was expected for a particular position in the utterance. The optional CF did not include extended grammar rules—in fact, grammar rules were orally presented to the learners before practice. So, the embedded CF might have served as a proxy for the grammar rules (as we had intended), which could have constituted sufficient learning support for this particular task and for these particular learners. Subsequently, when responding to the items on perceived usefulness in the post-questionnaire, the participants might have restricted their judgments to the highlighting CF, even though the questionnaire items were introduced by a screenshot showing a highlighted token with the optional CF (see Appendix 2). Hence, this would have yielded an inaccurate measurement of perceived usefulness of optional metalinguistic CF. This does not mean that future research should abandon investigating learners' perceptions, but that improved versions of the questionnaire should more clearly distinguish between the embedded and non-embedded CF. Also, additional instruments could be used, such as post-experimental interviews to gain a more detailed understanding of learners' perceptions, or eye-tracking to measure another aspect of CF usage, namely noticing.

Similarly to perceived usefulness, learners' achievement goal orientation could not explain their use of optional CF, nor could it explain use of hints and peer responses. In other words, there was no evidence that task-approach was associated with instrumental help-seeking behaviour (intended to promote learning), which we hypothesized would be reflected in the use of optional CF, and no evidence that executive help-seeking (in this case frequent use of hints and peer responses) was associated with a performance orientation (other-approach) (see also Baker et al., 2008 for the lack of association between these variables). We did, however, find that both use of hints and use of peer responses were inversely related to use of optional CF and to the number of attempts per exercise, which could support the distinction between executive and instrumental help-seeking behaviour.

On a more general plane, the lack of relations between learners' perceptions (perceived usefulness and achievement goal orientation) and use of optional CF might be attributed to the limited sample of participants and short period of practice characteristic of this pilot study. Learners may not yet have been sufficiently familiar with the features of the learning environment, and their perceptions might not have stabilized and may hence have been inaccurate with respect to their actual interaction in the learning environment. The following observation illustrates this hypothesis. During the practice sessions one learner remarked that he refrained from using the hint option because he thought he would cheat and hence lose points. This was not the case—which implies that this learner's perceptions were not 'calibrated' (Winne, 2004) to actual design of the environment, nor were they in tune with the intentions of the instructional designer—but it constitutes a plausible belief in the context of gaming ecologies. Thus, in future studies, participants might need to be given more technical guidance or additional time to practise in the environment, before their perceptions are measured.

Lastly, future studies could consider exploiting the gaming features inherent in the environment, in an attempt to incentivize learners' use of optional (metalinguistic) CF. Performance-oriented learners might use optional help features more often if they get rewarded for correct responses and when the help offered is actually useful for solving problems. In order to empirically investigate the effects of game-like features on learner behaviour, positive feedback and competition mechanisms could be implemented in separate experimental conditions.

Acknowledgements

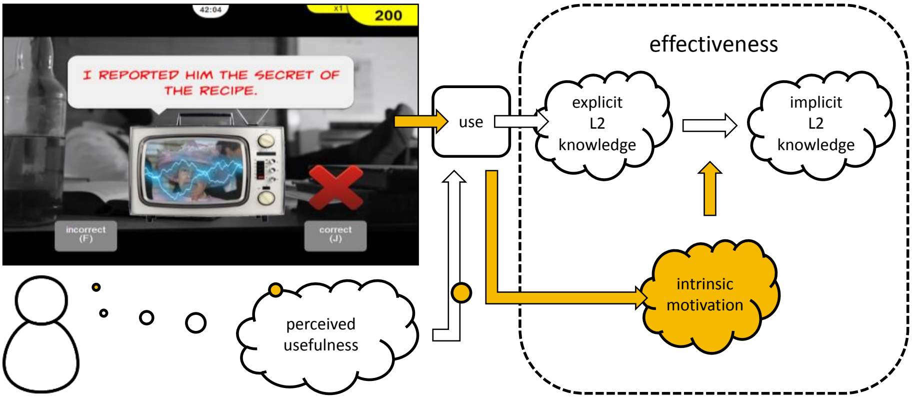
The development of the learning environment used in this study was partly funded by the LLINGO project ("Language Learning in an Interactive Game Environment"), conducted within the ICON programme that is co-funded by iMinds (Flanders' research centre that stimulates strategic research on digital innovation).

The design and parts of the source code of the technology-based learning environment were based on work conducted within the LLINGO project, to which the iMinds media innovation center MIX contributed significantly. The PhD candidate further developed the technology for the application, and implemented natural language processing techniques with great help from Ruben Lagatie. The technology was developed with and deployed through the web application framework Google App Engine. Dr. Mieke Vandewaetere offered methodological support. Prof. dr. Geraldine Clarebout and Prof. dr. Piet Desmet contributed to the revision of this manuscript.

Third interlude

The results of the first empirical study in this PhD project (reported in chapter 3) suggest that design features of corrective feedback may have a strong impact on learner motivation and development. The next two chapters elaborate this theme, reporting on two effectiveness studies that address the impact of corrective feedback on, respectively, intrinsic motivation and second language grammar learning. To investigate the effectiveness of corrective feedback, we use *mini-games*, i.e. short, focused games that lend themselves well to controlled practice of particular features of a target language, as well as to the experimental manipulation of instructional design features.

The first of these effectiveness studies addresses the effectiveness of elements related to vivid corrective feedback on intrinsic motivation.



Chapter V

Empirical study 3: Effectiveness of vivid corrective feedback for supporting learner motivation in grammar practice with mini-games

This chapter is an expanded[3] version of a manuscript that was published as:

Cornillie, F., & Desmet, P. (2013). Seeking out fun failure: how positive failure feedback could enhance the instructional effectiveness of CALL mini-games. In *Global perspectives on Computer-Assisted Language Learning. Proceedings of WorldCALL 2013* (pp. 64–68). University of Ulster.

[3] Compared to the manuscript published in the conference proceedings of *WorldCALL 2013*, all sections (introduction, method, discussion, and conclusion) have been significantly expanded, and additional analyses have been carried out.

5.1 Introduction

Controlled practice, which we define here as learning aimed at the improvement of performance of specific routines as a part of the acquisition of complex skills, is considered a necessary step towards the achievement of skilful behaviour in many areas of human development (Anderson et al., 2004), including the learning of a second language (L2) (DeKeyser, 2008). Tutorial CALL activities (Hubbard & Bradin Siskin, 2004) epitomize controlled practice, and, moreover, afford opportunities to combine potentially useful language practice with meticulous experimental research on learning processes and learner characteristics. By means of such technology-enhanced activities, researchers can elicit and direct learners' use of specific linguistic phenomena; they can deliver features of instructional design that are inherent in practice (such as corrective feedback) in consistent ways, and they can manipulate such features carefully; and tracking systems built into tutorial CALL are capable of closely monitoring learners' actual behaviour (such as their performance or use of supportive information). So, CALL offers unique methodological benefits for research on controlled practice.

Conversely, it may be argued that drawing the card of computerized tutorial practice—often referred to as *drilling*—flies in the face of current thinking on L2 pedagogy and L2 acquisition. In this perspective, tutorial CALL practice presents at least three challenges. First, explicit focus-on-form practice needs to engage learners foremost in meaningful L2 processing. In current views on language acquisition, surface processing of L2 'input' without reading for its meaning—as is typically the case in mechanical drilling—is not considered input, and even has adverse effects (Wong & VanPatten, 2003). Secondly, Dörnyei (2009) argues that "the key to the effectiveness [of controlled practice] is to design interesting drills that are not demotivating" (p. 289), and sums up a range of techniques to accomplish this, such as creating variation in repeated utterances, making drills personally relevant to learners, or using CALL or games. Ideally, L2 practice environments catalyse self-sustained and potentially *intrinsically motivated* behaviour—behaviour that is performed because it is inherently interesting or enjoyable (Ryan & Deci, 2000)—so that learners are willing to practise their language skills and remediate problems in self-directed contexts, for the inherent sake of practice, and without external regulation (without the teacher present). A third and related challenge for tutorial practice concerns the provision of corrective feedback (CF; also known under the more general term 'negative feedback'), as CF may be downright detrimental to learners' motivation. More specifically, the immediacy, salience, and consistency of computer-generated negative feedback may cause errors to be more prominent than in face-to-face practice contexts. In keeping with the view that failure states in learning are likely to be "blown out of proportion" (G. L. Robinson, 1991, p. 193) and to last longer in memory than positive learning experiences—Hattie & Yates formulate this as the principle that "bad is stronger than good" (2014, p. 65)—this has led to the claim that CF may be damaging to learners' self-concept (G. L. Robinson, 1991; Schulze, 2003).

The current study mainly addresses the latter two challenges, and attempts to empirically evaluate the usefulness of game design mechanics with respect to learners' intrinsic motivation in tutorial CALL practice. In current-day language classrooms, there is little scope for controlled practice, as the time spent on such activities in class reduces the already limited opportunities for communicative practice. A possible approach for teachers is to rely on *mini-games*, i.e. games that can be played in brief sessions, are constrained in scope, provide consistent feedback, and thus lend themselves easily to focused practice (Cornillie & Desmet, n.d.). Such games might be particularly powerful in terms of motivating learners to practise the L2 outside the classroom, without much external regulation from teachers or parents.

Our attempt of using game mechanics in order to support learner motivation in tutorial CALL practice may be seen as related to what is currently known in the field of human-computer interaction as *gamification*, i.e. the application of elements of game design to non-gaming contexts, with the aim of engendering user engagement (Deterding, Dixon, Khaled, & Nacke, 2011). However, whereas gamification approaches typically seem to rely on extrinsically motivating strategies, such as positive feedback in the form of points and reward systems, this study is guided by a notion in game design that may have a more intrinsic appeal, namely the notion of *positive failure feedback*. The next section defines such feedback, and presents a theoretical exploration of the links between its design attributes and intrinsic motivation.

5.2 Background research

In a discussion of an empirical study that revealed associations between negative feedback sessions in gameplay and subsequent positive affect (Ravaja et al., 2006), game designer and researcher Jane McGonigal (2011) describes *positive failure feedback* as “a vivid demonstration of the players’ agency in the game” (p. 66). She argues that, while the essence of such feedback is to signal failure, it does so in ways that the player is more likely to persevere than “in real life [where] we experience diminished interest and motivation” (p. 66). Game designer Steve Swink (2006) writes that positive failure feedback engenders “a very visceral ‘oooh daaaamn!’ kind of reaction, one that has a hugely positive effect both on learning and capture”, and that “because the failure state is so much fun, learning is much easier and frustration mitigated”. In other words, positive failure feedback in games serves two functions: to communicate failure to the player—which puts it on a par with negative ‘knowledge of results’ feedback in learning settings—and to simultaneously support motivation.

So, whereas negative feedback in L2 teaching settings is often considered to impair learner interest and bring about negative affect (Magilow, 1999; Truscott, 1996), negative feedback in games has the potential to enhance motivation. Hence, Purushotma, Thorne, & Wheatley (2008) list the provision of positive failure feedback as the first of their “10 key principles for designing video games for foreign language learning”, arguing that, by relying on such feedback, “video games offer an opportunity to lower some of the frustration and anxiety students often feel while learning a second language”.

In what follows, we will first present a recent model of how gaming environments are thought to intrinsically motivate players. Subsequently, we zero in on two aspects of this model, and discuss each of these in light of a dedicated theory, namely attribution theory and telepresence theory. We will argue on the basis of these theories that design attributes of positive failure feedback, namely signals of failure, and vividness, have the potential to affect intrinsic motivation in tutorial CALL practice.

5.2.1 Intrinsic motivation in digital play: Player Experience of Needs Satisfaction

Recently, a model has been proposed to explain human engagement in video games, known as the Player Experience of Needs Satisfaction (PENS) model (Ryan et al., 2006). The merit of the model is that it builds on Self-Determination Theory (SDT) (Ryan & Deci, 2000), a macro-theory of motivation that has been validated in many areas of human development. SDT distinguishes various types of motivation which differ in their degree of self-regulation, and considers intrinsic (self-regulated) motivation as the ideal regulatory style: "a natural wellspring of learning and achievement that can be systematically catalyzed or undermined by parent and teacher practices" (Ryan & Deci, 2000, p. 55). It posits that human beings have three universal needs that contribute to intrinsically motivated behaviour and general well-being, namely the need to feel *competent*, the need for *autonomy* (i.e. being able to make choices), and the need for feeling *related* to other people.

PENS is an extension to SDT, adding two factors in order to account for people's engagement in playing digital games. These two factors are *intuitiveness of controls* (i.e. the ease with which players interact with the game, as mediated by input devices), and *perceived immersion*. Immersion has three dimensions: physical immersion (i.e. the sense of being physically present in a game), emotional immersion (i.e. the sense of experiencing feelings in a game as in 'real life'), and narrative immersion (the sense of being part of a story).

In the next sections, we will discuss how positive failure feedback has the potential to support perceived competence and perceived immersion. This discussion will be guided by, respectively, attribution theory (and the phenomenon of learned helplessness), and telepresence theory.

5.2.2 Attribution theory: the communication of failure and perceived competence

In the field of psychology, attribution theory sets out to explain why human beings seek causes for particular behaviour and events. This theory has been used to investigate the phenomenon of *learned helplessness* (Abramson, Seligman, & Teasdale, 1978). Learned helplessness refers to a state of amotivation that is most likely to occur when human beings are confronted repeatedly and *ad nauseam* with failure, and when three conditions are satisfied: 1) the individual believes that failure is due to a lack of his or her own competence rather than to task factors, such as complexity (i.e. attribution to internal rather than external causes); 2) the individual attributes failure to general incompetence rather than to a lack of ability at a specific task (i.e. attribution to a general rather than a specific cause); 3) the individual thinks that failure cannot be overcome (i.e. attribution to stable rather than unstable causes). This ties in directly with negative feedback, which communicates failure at a task to an individual. Research shows that negative feedback can decrease perceived competence and intrinsic motivation (Vallerand & Reid, 1984), and that strong and repeated negative feedback given in tasks concerning unsolvable problems can lead to a state of learned helplessness (e.g. Mikulincer, Kedem, & Zilkha-Segal, 1989; see also discussions in Kluger & Denisi, 1996).

From an applied perspective, learned helplessness theory offers sensible principles for the design of negative feedback in instruction. In line with the three dimensions discussed above, it predicts that learning environments are more likely to support learners' beliefs in their competences when substandard performance is blamed on specific characteristics of learning tasks rather than on the self (external cause), when it is emphasized that failure is not general but particular (specific cause), and when information is offered that can help learners to improve their performance (unstable cause). The latter can be done by including hints or extended explanations in addition to negative 'knowledge of results' feedback.

Next to its practical implications for the design of learning environments, learned helplessness theory is a useful tool to evaluate the design of negative feedback in digital games. According to game researcher Juul (2013), a great deal of commercial off-the-shelf games violates the third implication derived from learned helplessness theory, by not providing any direct information (such as hints) that may help players to improve their performance. This is in strong contrast with instructional environments, in which both teachers and learners strongly value hints and explanations concerning the underlying causes of undesirable performance. Moreover, a small group of games even violates the first two principles and "flaunts good manners" (Juul, 2013, p. 54), for instance by verbally insulting players when they fail. This would seem to affect learners' perceived competence in negative ways.

Yet, as concerns the first two design principles derived from learned helplessness theory, games have one defining characteristic which, in contrast with 'real-life' activities, explains why players tolerate such transgressions of real-world social conventions: games are essentially *not* real life. Games explicitly evoke representational contexts that are different from players' usual frame of reference. This brings us to the game design element of *fantasy*, a rather broad and somewhat elusive concept that combines elements of narrative, role-play, and the uncanny. We define fantasy as:

> Make-believe environment, scenarios, or characters. It involves the player in mental imagery and imagination for unusual locations, social situations, and analogies for real-world processes. The player is also

> required to take on various roles in which they are expected to identify. (Bedwell et al., 2012, p. 4)

The key point is that, once players step into the magic circle, there are no real-world consequences, and the negative effects of failure and negative feedback on self-perceptions caused can be minimized. As drama expert and human-centred interaction designer Brenda Laurel puts it: "The distinguishing feature of the emotions we feel in a representational context is that there is *no threat of pain or harm in the real world*" (1993, p. 114; emphasis in original). The next section discusses the ways in which games conjure up fantastic representational contexts.

5.2.3 Telepresence theory: vividness of feedback and perceived immersion

The concept of *telepresence*, in game studies used interchangeably with the notion of *immersion*, originates in research on virtual reality, and refers to human beings' experience of being in an environment as mediated by technology (Steuer, 1992). Telepresence theory tries to explain how this experience is rendered and mediated by technological artefacts. It distinguishes between two general dimensions that contribute to telepresence: interactivity and vividness.

Interactivity refers to how users can influence the form or content of the mediated environment. Feedback plays an essential role here, for it is by way of feedback that interactive environments communicate the results of users' actions back to them. The second dimension, *vividness*, signifies the ability of a technology to produce a sensorially rich mediated environment. More particularly, it concerns "the representational richness of a mediated environment as defined by its formal features, that is, the way in which an environment presents information to the senses" (Steuer, 1992, p. 81). Steuer discusses two generally recognized factors that contribute to vividness, namely *sensory breadth* and *sensory depth*. The former notion refers to the number of senses that are simultaneously involved in the experience. For instance, in virtual reality systems feedback can be delivered aurally, visually, or in tactile form. The notion of depth can be best illustrated by means of visual or auditory resolution, for instance the number of polygons involved in rendering an object on a screen, or the number of frames per second for animations.

Remember that McGonigal (2011) defines positive failure feedback as "a *vivid* demonstration of the players' agency in the game" (p. 66; emphasis added). Further, she formulates the following design principle: "The trick is simple, but the effect is powerful: you have to show players their own power in the game world, and if possible elicit a smile or a laugh" (p. 67). How exactly feedback can be made vivid depends on the representational context of a specific game. Serious games designer Prensky (2001) writes that "feedback [in games] comes via action" (p. 159), for instance as big bangs and dead bodies. It may be derived that feedback in games is typically contingent upon the theme of the game—when the game is about conflict in war, feedback will take the shape of blood or dead bodies. Further, instructional designer Clark Aldrich (2005) considers feedback in (instructional) games as "an opportunity to wrap a story around the situation" (p. 25), the function of which is to make the experience more immersive. So, feedback in games often comprises elements of a game's fantasy, which is intended to make the experience more vivid, and to increase players' sense of immersion.

5.2.4 Summary: how positive failure feedback may enhance intrinsic motivation in tutorial CALL practice

As noted in the introduction, a major issue in the design of tutorial CALL activities concerns the provision of CF. Learned helplessness theory learns us that controlled practice with consistent CF may undermine learners' perceptions of competence when they are working on a difficult task, receive CF on the same problem repeatedly, and if CF does not help them to "close the gap"

(Hattie & Timperley, 2007) between current and desired knowledge states—for instance if no useful supportive information is provided or if learners cannot deal with such information. Yet, perceived competence may be harmed less if learners feel immersed in an experience that allows them to blame failure mainly on the task rather than on the self. Games are the medium *par excellence* that allow learners to do this, because they create their own representational context via fantasy. What is more, motivation in tutorial CALL practice can be increased if CF functions like positive failure feedback in games—that is to say, if it is vivid, emphasizing the learner's agency in the world represented in the game.

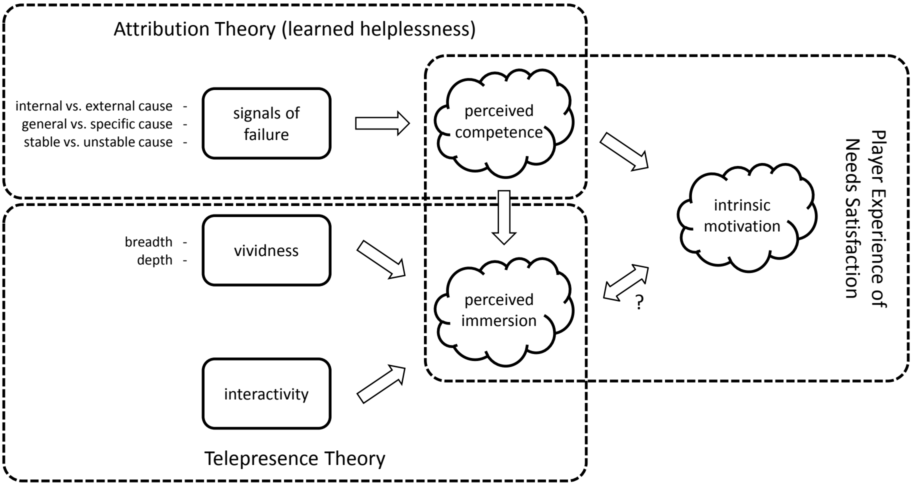

Attribution Theory (learned helplessness)
internal vs. external cause -
general vs. specific cause -
stable vs. unstable cause -
signals of failure
perceived competence
breadth -
depth -
vividness
perceived immersion
interactivity
Telepresence Theory
intrinsic motivation
?
Player Experience of Needs Satisfaction

Figure V-1: theoretical framework for the study

Further, PENS may be used as a composite theory to account for the potential relationship between vivid CF in gamified tutorial CALL and intrinsic motivation (see Figure V-1). Note that in PENS (in accordance with SDT), needs satisfaction (including competence need satisfaction) is clearly seen as an antecedent to intrinsic motivation, but that the relation between immersion, intrinsic motivation, and perceived competence is less transparent. In contrast with telepresence theory, which claims that immersion is created by features of the environment, PENS predicts that needs satisfaction is the major predictor of immersion (Przybylski et al., 2010)—it is not entirely clear why. Further, the relation between immersion and intrinsic motivation is not articulated in the PENS model.

5.3 The current study

In the previous section, we argued that signals of failure, fantasy, and vividness of CF in tutorial CALL practice may affect learners' perceived competence and immersion in positive ways. Further, in line with insights from game designers, we discussed that the design of vivid CF (i.e. positive failure feedback) likely depends on the fantasy represented in a particular game context. This yields three possible combinations of (the presence or absence of) fantasy and CF type (see Table V-1).

Table V-1: configuration of vivid CF

	CF	**Vivid CF**
No fantasy	A	?
Fantasy	B	C

The current study, then, explores the utility of fantasy and vividness of CF for sustaining learners' intrinsic motivation in controlled L2 practice. We adopt the view that game attributes are likely to influence perceived competence and immersion, which function as antecedents of intrinsic motivation, which in turn may predict learners' willingness to practice in the future. We address the following research questions:

1. How do fantasy and vividness of CF affect learners' perceived competence and immersion in a difficult grammar practice task?
2. How are perceived competence and immersion related to learners' intrinsic motivation, and to their willingness for future practice?

5.4 Method

5.4.1 Research design and conditions

The study used a within-subjects (repeated measures) experimental design, which involved all participants in practising a complex syntactic rule in English using three versions of a computerized grammaticality judgment task that differed in terms of vividness: one baseline version without fantasy, and two versions that included a fantasy, one of which contained vivid CF. To control for order effects, the design was counterbalanced: upon logging in, the system assigned participants at random to one of six groups which differed only in terms of the order in which the three conditions were presented.

In all three practice conditions, participants judged the grammaticality of up to 36 different sentences that were presented on the computer screen, one at a time and in a random order (see section 5.4.3 for details on the content of these sentences). With the aim of adding challenge to this already difficult task, two design features were included that are commonly associated with mini-games, namely time pressure and a points system (Cornillie & Desmet, n.d.).

As for time pressure, participants were asked to judge as many sentences as possible within 60 seconds—the software was programmed to loop through the sentences in case all 36 had been presented—and there was also a 10-second time limit per sentence. Admittedly, the 10-second time limit was somewhat arbitrary. This corresponds to the upper limit of the time range reported by Loewen (2009) for studies that intended to measure implicit (quickly retrievable) knowledge in grammaticality judgment tests. Considering that the practice sentences used in the current study are quite short, a time limit of 10 seconds may have allowed learners to draw both on implicit and explicit knowledge to make their judgments. However, considering the complex nature of the practised grammar rules (see section 5.4.3), it was considered that this lime limit would create a great degree of challenge for the learners. In addition, the time pressure may have encouraged learners to develop automatized knowledge of patterns which they did not fully master.

Participants used the computer keyboard to indicate their judgments: if they thought the sentence was incorrect, they had to press the “F” key; for correct sentences, the “J” key had to be used. So, participants used their right hand to indicate correctness, and their left hand to indicate incorrectness. As a support mechanism designed to help learners operate the keyboard, the two keys were visualized as buttons on the screen, in a bottom left and right position, and in red and green, respectively, and visual feedback was displayed directly above the buttons. The operational instructions were also included in a briefing screen shown prior to each practice task (see below), and were given orally in the introduction to the experiment.

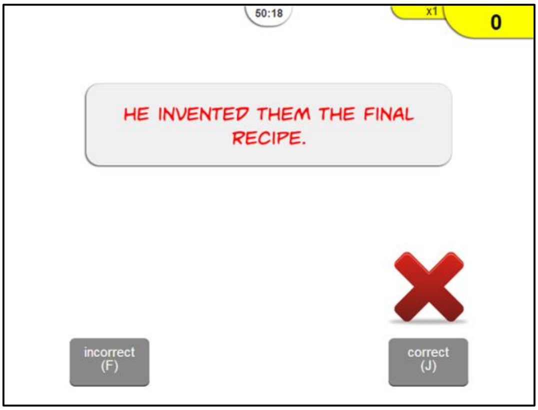
After each response, feedback was given. For correct responses, a green checkmark was shown above the button that corresponded to the participants’ judgment, and positive feedback was given in the form of points (100 per correct response), with sound support. A bonus system was also implemented: for each set of four successive correct responses, a score multiplier M was increased, so that subsequent sets of four successive correct responses would receive M times 100 points, up to a maximum of 500 points per response. The multiplier M was reset to 1 upon the first next incorrect response. The computer also kept track of the participants’ personal best scores, which were shown on the debriefing screen (see further).

The condition determined how immediate CF was implemented for incorrect responses. Two elements of CF were common to all conditions: a red cross was displayed above the button corresponding to the learner’s judgment, and ungrammatical sentences which were judged as correct were highlighted in red. To give learners the opportunity to process this CF, the possibility to interact was paused for 2 seconds (while the 60-second timer continued). In conditions A (see Figure V-2) and B (see Figure V-3), plain CF was given. Plain CF comprised the common elements of CF (the red cross, and red highlighting for ungrammatical sentences), complemented with a sound effect which may be best described as an ‘incorrect’ sound typical of quiz shows. Conditions B and C contained a fantasy, viz. a detective (the learner) that questions witnesses of a theft (in this case, celebrities), using some kind of videophone coined a *Tele-Interrogator*. Whereas fantasy condition B included plain CF, fantasy condition C contained vivid CF (see Figure V-4), differing from plain CF in terms of sensory depth and quality (i.e. it was adapted to the fantasy). Vivid CF consisted of the common CF features, but now complemented with animations: the facial expression of the current witness changed to horrified or angry, and any of three animation effects was shown (an electric shock, water filling the screen, and an alien space ship flying over (see Figure V-7). These animations were capped at the length of the feedback pause (2 seconds), and for each animation, a specific sound effect was played back instead of the 'incorrect' sound. In combination with 7 interviewed celebrities, the 3 animation effects resulted in 21 possible forms of vivid CF, which were randomly selected at run-time. Usability tests (see section 5.4.4) and best practices in game design (Swink, 2006; Wright, 2003) indicated that such variation in vivid CF was desirable. So, in a nutshell, the three conditions differed in terms of vividness (with and without fantasy, with and without vivid CF) (see Table V-2).

Table V-2: differences in vividness between the three conditions

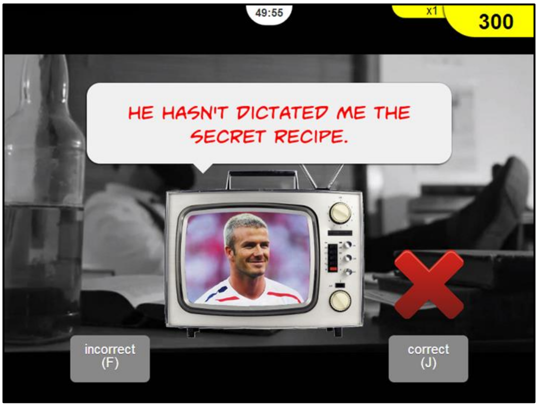
task feature	condition A	condition B	condition C
fantasy	no	yes	yes
visual CF	red cross	red cross	red cross + animation
auditory CF	'incorrect' sound	'incorrect' sound	animation sound

Further, in all three conditions a jazzy tune was played during the practice task. While this was done primarily to support the fantasy in conditions B and C, it could be argued that the gamified nature of all three conditions demanded that music be played. More importantly, from a methodological point of view, we wanted to exclude the possibility that the lack of music in condition A would influence the responses on the questionnaires.



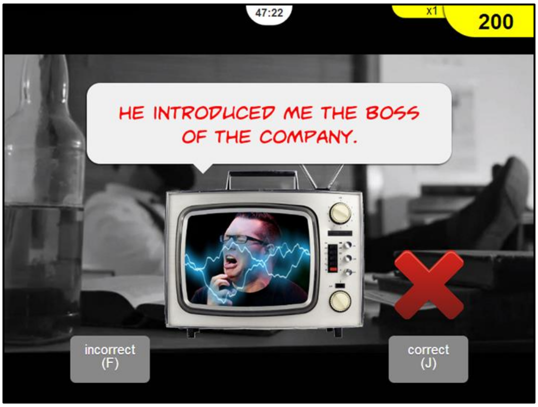
HE INVENTED THEM THE FINAL RECIPE.

Figure V-2: condition A (no fantasy, plain CF)



HE HASN'T DICTATED ME THE SECRET RECIPE.

Figure V-3: condition B (fantasy, plain CF)

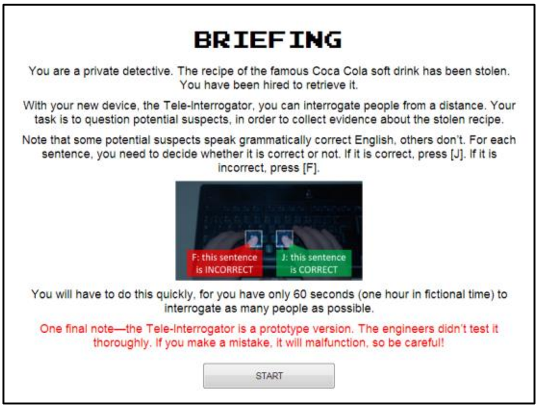


HE INTRODUCED ME THE BOSS OF THE COMPANY.

Figure V-4: condition C (fantasy, vivid CF)

Each practice task was preceded by a briefing screen, and ended with a debriefing screen. The briefing explained the purpose of the practice task, and also comprised instructions on how to operate the keyboard. In conditions B and C, some additional information was given to introduce the fantasy (see Figure V-5). Moreover, condition C contained one statement related to vivid CF. This particular instruction was rendered in a red font, in order to increase the chances that learners who had seen an otherwise identical briefing screen in condition B would notice and read it.

After each practice task, a debriefing screen was presented (see Figure V-6), displaying the learner's current score, his or her personal best score, and an overview of the incorrect responses, with for each incorrect response some brief declarative (metalinguistic) information on the nature of the mistake, which referred back to the instructions given prior to practice (see section 5.4.5). Finally, the wording of the introductory statement on this screen was adapted to the availability of vivid CF. In the conditions with plain CF (A and B), this statement read (in case the learner had made a couple of mistakes): "You made some mistakes. Here is some information that may help you to improve". In condition C, which included vivid CF, the corresponding statement was: "Woops, it looks like your machine made some persons unhappy. Don't worry – the Department of Complaints is dealing with it ... Meanwhile, here is some information that may help you to prevent this from happening in the future!". Admittedly, while this adaptation may have been all too subtle, it was done in order to safeguard learners' perception of competence, and was designed according to instructional principles derived from learned helplessness theory (see section 5.2.2). In the former case, the learner was blamed for failure; in the latter case, failure was attributed to the machine in the fantasy context of the game.



BRIEFING

You are a private detective. The recipe of the famous Coca Cola soft drink has been stolen. You have been hired to retrieve it.

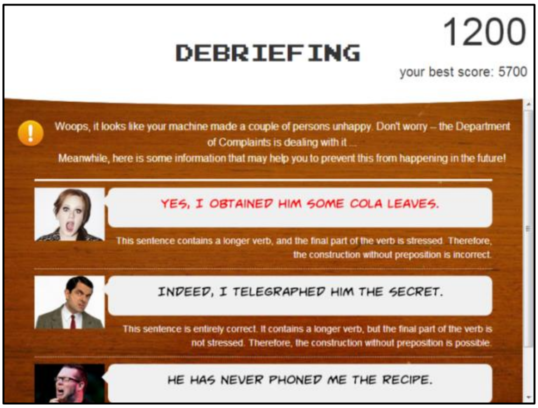
With your new device, the Tele-Interrogator, you can interrogate people from a distance. Your task is to question potential suspects, in order to collect evidence about the stolen recipe.

Note that some potential suspects speak grammatically correct English, others don't. For each sentence, you need to decide whether it is correct or not. If it is correct, press [J]. If it is incorrect, press [F].

You will have to do this quickly, for you have only 60 seconds (one hour in fictional time) to interrogate as many people as possible.

One final note—the Tele-Interrogator is a prototype version. The engineers didn't test it thoroughly. If you make a mistake, it will malfunction, so be careful!

Figure V-5: briefing screen for condition C



DEBRIEFING

1200

your best score: 5700

Woops, it looks like your machine made a couple of persons unhappy. Don't worry – the Department of Complaints is dealing with it ...
Meanwhile, here is some information that may help you to prevent this from happening in the future!

YES, I OBTAINED HIM SOME COLA LEAVES.

This sentence contains a longer verb, and the final part of the verb is stressed. Therefore, the construction without preposition is incorrect.

INDEED, I TELEGRAPHED HIM THE SECRET.

This sentence is entirely correct. It contains a longer verb, but the final part of the verb is not stressed. Therefore, the construction without preposition is possible.

HE HAS NEVER PHONED ME THE RECIPE.

Figure V-6: debriefing screen for condition C

5.4.2 Questionnaire design

The questionnaire was drawn up in the participants' native language (Dutch), and consisted of two sections. The first section comprised four self-constructed 7-point Likert scale items that were connected to how participants had experienced the practice task cognitively. These self-report items addressed learners' reliance on implicit knowledge ("When judging the sentences, I responded on the basis of my intuition"), reliance on explicit knowledge ("When judging the sentences, I was thinking about the grammar rules”), the degree to which they had been able to focus on meaning (“I feel capable of telling someone else what the sentences were about”), and perceived difficulty (“I found the task difficult”).

The main section of the questionnaire was a combination of subscales related to motivation in a 7-point Likert format, selected from the Intrinsic Motivation Inventory (IMI) (Plant & Ryan, 1985) and from the Player Experience of Need Satisfaction (PENS) instrument (Ryan et al., 2006). Both scales draw on Self-Determination Theory, but the latter was customized for research on games. The subscale ‘interest/enjoyment’ (i.e. the scale that was intended to measure intrinsic motivation; 7 items, e.g. “This task was fun to do”) and the subscale ‘perceived competence’ (6 items, e.g. “I think I was pretty good at this task”) were taken from IMI, as we already had Dutch translations that had proved reliable in earlier studies. The subscale ‘perceived immersion/presence’ (9 items, e.g. “Doing this task felt like taking a trip to another place”) was based on PENS, and focused on physical, emotional, and narrative aspects of presence. The three subscales formed one part of the questionnaire’s main section, and the order of the items was randomized for each condition. The final part of the questionnaire concerned participants’ willingness for self-directed future practice (1 item; “In the future, I would like to use this electronic environment for practising grammar on my own initiative”).

5.4.3 Target structure

The target structure that was practised was dative alternation in English. This structure was chosen because it is known to be highly complex, and—for L2 learners specifically—is required to demand explicit instruction and practice with CF (Carroll & Swain, 1993). Further, we selected the structure because L2 learners are thought to acquire it late, and because it is typically not instructed in L2 English curricula (R. Ellis, 2009b). Equally, the participants of this study had not received any overt instruction on dative alternation prior to the experiment. Hence, it was anticipated that learners would repeatedly make errors during practice, and would hence receive a great deal of CF.

For the practice tasks, we constructed 36 sentences for five syntactic patterns of dative alternation (see Table V-3). These five patterns were largely based on Carroll & Swain's (1993) description of dative alternation for L2 learners (see also Mazurkewich & White, 1984). All sentences were declarative and active, and the indirect object was always a human beneficiary, in order to avoid that learners would be confused by rather absurd sentences such as 'She sent that address the letters'. Sentence length ranged between 5 and 10 words (M = 7.4). Further, all sentences developed for use in the practice tasks were related to the theme 'the recipe of Coca-Cola'. This was done for three reasons. First, the use of a fantasy in conditions B and C (see above) demanded that the sentences were somehow grouped meaningfully. Secondly, we wanted thematic coherence because this could increase the chances that learners also processed the sentences for their meaning—although admittedly, learners could perfectly tune out of meaning if they wanted to. And third, contextualizing the linguistic patterns was done with a view to embedding the controlled practice tasks in more authentic and communicative tasks in future studies.

Table V-3: distribution of the patterns for the practice tasks, with sample sentences (* denotes ungrammaticality)

syntactic pattern	**number of sentences**	**sample sentence**
verb + NP + PP	4	*Yes, I obtained some cola leaves for him.*
* verb + NP + NP (no transfer)	8	** Indeed, he opened me the recipe book.*
monosyllabic verb + NP + NP	8	*No. I made him that first bottle of cola.*
polysyllabic verb with initial stress + NP + NP	8	*I have never offered him any money.*
* polysyllabic verb with final stress + NP + NP	8	** He introduced me the boss of the company.*

5.4.4 Evaluation of usability and instruments

Prior to the experiment, the interface of the system and the experimental procedure were piloted with three experts: a computational linguist and former English language teacher with a strong interest in CALL, one of the participating teachers who had over 30 years of teaching experience, and an educational scientist who described herself as having difficulties with foreign languages. All three experts practised once in each of the conditions A, B, and C (in this order), and were briefly interviewed after each condition.

Observation of and interviews with the experts indicated that the interface, instructions, input, and CF mechanisms were clear. The experts' responses to the vivid CF were more interesting. In the software used for this usability study, only one animation effect was included, namely the electric shock, supported by human screams that were so overdone that they were intended to be humorous. However, the first expert had experienced this effect as "somewhat fun, but frustrating"—his behaviour during the practice task testified especially to the latter. The third expert found the animation repetitive, said that it drew too much attention to her own failure, and expressed her preference for vivid positive feedback, rather than vivid negative/corrective feedback. Moreover, none of the participants preferred condition C. On the basis of these reactions, and insights in the game design literature (Swink, 2006; Wright, 2003), two animation effects were added to the vivid CF (see Figure V-7), which created more variation, and the human screams were removed. Positive feedback was not changed, as our interest was in CF.

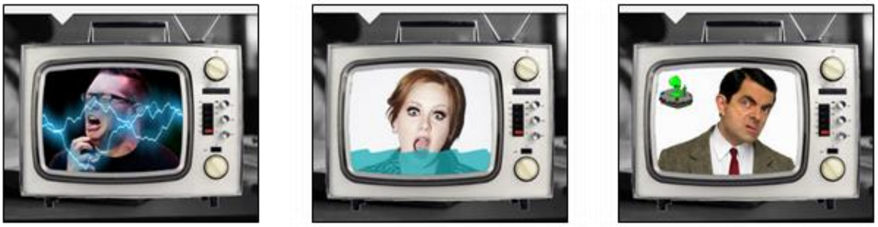


Figure V-7: three animation effects for vivid CF

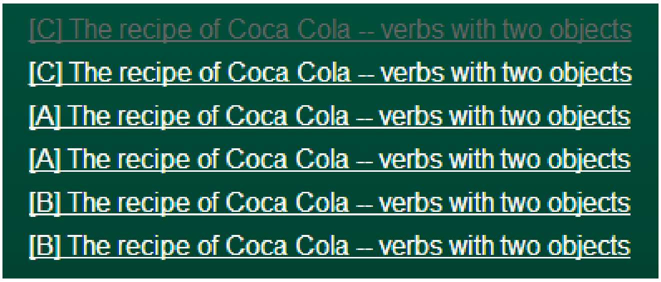
Further, after practising in condition B, the second expert indicated that he had not noticed anything different about the task format, because he had been too much focused on judging the sentences. This alerted us to the possibility that learners might have the same experience, and, hence, that their responses on the questionnaire might be of no interest. As a result, we included an additional open question in the second and third questionnaires: "Did you notice anything in this practice task in comparison with the previous one? What?". This allowed us to control for the possibility that learners had not experienced the second and third condition differently than the first one. Also, we considered that this open question could provide a measure of the degree to which they had taken the questionnaire seriously.

5.4.5 Participants and procedure

The study took place in May 2013. The participants were 32 intermediate-level Dutch-speaking learners of English in the 5th (N = 20) and 3rd (N = 12) grades of secondary education in Flanders, Belgium. In the Flemish curriculum, these grades correspond to the 4th and 2nd year of formal English language teaching, respectively. The within-subjects experimental design justifies the small number of participants methodologically, because all participants serve as their own control.

Prior to practice, learners received explicit instruction on dative alternation in English, provided by the researcher. The metalinguistic terminology used in the explicit instruction (e.g. 'indirect object') was adapted to the terminology used in their training, and the researcher also repeated the terminology in Dutch. Then, learners received some guidance on how to operate the computerized grammaticality judgment task, and practised dative alternation in class, with 4 sample sentences that were projected in a format equal to condition A.

Next, learners were asked to practise individually, twice in each condition, with headphones. During practice, the system logged learners' reaction times and accuracy rates. The colour of the hyperlinks to the practice tasks indicated the learners' progression (see Figure V-8). After each condition, they filled out a questionnaire on paper (see section 5.4.2). Some of the participants in the first group had filled out their first questionnaire after practising in two or three conditions. The researcher took note of this during the experiment. Further, lack of time prevented the first group from completing the last questionnaire.



[C] The recipe of Coca Cola -- verbs with two objects

[C] The recipe of Coca Cola – verbs with two objects

[A] The recipe of Coca Cola – verbs with two objects

[A] The recipe of Coca Cola – verbs with two objects

[B] The recipe of Coca Cola – verbs with two objects

[B] The recipe of Coca Cola – verbs with two objects

Figure V-8: example of a learner's progression through the practice tasks

Before and after the treatment, learners completed a paper-based grammaticality judgment test on dative alternation. Analysis of the pre- and post-test data is outside the scope of this chapter.

A couple of days after the experiment, follow-up interviews were held with six volunteers (in two groups of three). These learners first practised again, once in each condition (in alphabetic order), and then participated in a semi-structured group discussion moderated by the researcher.

5.5 Results

5.5.1 Data preparation

The logged data of learners' behaviour during practice were processed in order to construct measures of the mean accuracy rate and mean reaction time per practice session of 60 seconds. Further, the experimental condition of each session was extracted and added to the data set for analysis. The logging data of two participants were discarded, as they had accidentally used the same login code. We also discarded logging data from participants who, contrary to instructions, had practised more than twice in each condition, so as to have an equal amount of practice for all participants in subsequent analyses. We ended up with behavioural data from 23 participants. Table V-4 shows that the distribution of the experimental conditions over the session numbers is more or less counterbalanced.

Table V-4: distribution of the experimental conditions over time

condition	session 1	session 2	session 3	session 4	session 5	session 6
A	9	9	8	8	6	6
B	6	6	8	8	9	9
C	8	8	7	7	8	8

As for the questionnaire data, a couple of qualitative checks were performed after entering the raw data in spreadsheets. First, since the orders of the experimental conditions were automatically determined at run-time, and were hence unknown to the researcher beforehand, the log data were inspected in order to add the references of the experimental conditions to each of the observations for the questionnaire data. Next, a number of observations for the questionnaire data were discarded for any of the following reasons. First, some of the participants in the first group had only completed their first questionnaire after practising in the second or third condition. This may have influenced their responses, so these data were removed from the data set. Secondly, as to the motivation scale, one participant had responded consistently in the middle, and the practice session data indicated that he had taken very little time to complete the questionnaire; another participant had used a zigzagging pattern. This raised suspicion about how seriously these two participants had taken this part of the questionnaire, so these observations were deleted. Finally, on the second or third questionnaire, some participants had not reported noticing any differences in comparison with the previous condition(s). Because we reasoned that this would result in responses on the questionnaire that would differ from the responses of other participants, we also discarded these observations (see also section 5.4.4).

Reliability analyses of the motivation scales showed that all subscales had Cronbach's α scores higher than .7 (see Table V-5), and that the scale as a whole had a reliability score of .89. Therefore, we considered all scales reliable. Subsequently, for each of the subscales, an overall score was computed on the basis of the mean of all corresponding items.

Table V-5: reliability analysis of the motivation scales

Subscale	number of items	reliability
interest/enjoyment (i.e. intrinsic motivation)	7	.83
perceived competence	6	.8
perceived immersion	9	.88

Table V-6 shows the number of observations for the motivation parameters per condition and per questionnaire. The low numbers for questionnaire 3 are due to the fact that the first group did not complete the last questionnaire, and to the fact that some of the observations for questionnaire 2 and 3 were discarded because the participants had not reported seeing any differences compared to the previous practice task. The lower numbers for questionnaire 2 are also due to the latter. The table shows that there are missing observations, and that the data are unbalanced.

Table V-6: number of observations for the motivation parameters per condition and questionnaire

condition	questionnaire 1	questionnaire 2	questionnaire 3	TOTAL
A	9	6	3	18
B	9	6	3	18
C	7	6	2	15

5.5.2 Data analysis

Descriptive statistics of the logging data show that the mean accuracy rate is .62 for all conditions (see Table V-7). The results also show that the mean reaction time was slightly higher for the practice sessions in condition C than the other two conditions, and that the standard deviation of the reaction time was highest in condition C. As a result, the learners made fewer judgments in condition C.

Table V-7: descriptive statistics of logging data, per experimental condition

condition	number of sessions	mean accuracy rate	mean reaction time	mean number of judgments per session
A	46	.62 (*SD* = .13)	2.52 (*SD* = 1.05)	21.35 (*SD* = 9.90)
B	46	.62 (*SD* = .14)	2.51 (*SD* = .93)	21.00 (*SD* = 9.17)
C	46	.62 (*SD* = .13)	2.67 (*SD* = 1.12)	20.07 (*SD* = 7.88)

Looking at the reaction times and accuracy rates over the 6 practice sessions, the data show that the mean reaction time drops significantly (in the final session to about half of the mean reaction time in the first session), while the accuracy rates remain relatively stable over time (see Table V-8).

Table V-8: descriptive statistics of logging data, per session number

	session 1	session 2	session 3	session 4	session 5	session 6
mean accuracy rate	.62 (*SD* = .14)	.59 (*SD* = .17)	.63 (*SD* = .14)	.66 (*SD* = .10)	.59 (*SD* = .11)	.64 (*SD* = .11)
mean reaction time	3.58 (*SD* = .81)	3.08 (*SD* = .83)	2.74 (*SD* = .92)	2.26 (*SD* = .79)	2.01 (*SD* = .77)	1.73 (*SD* = .84)

To confirm these trends, two-way repeated measures ANOVAs were carried out on the accuracy rates and response times, with session number and condition as independent factors. The ANOVA for accuracy rate showed that neither session number ($F(5, 98) = 1.41$, $p = .23$) nor condition ($F(2, 98) = 0$, $p = 1$) significantly affected the accuracy rates. The second ANOVA showed that reaction time was affected significantly by session number (F(5, 98) = 49.84, p < .001) and condition (F(2, 98) = 3.23, p < .05). The interaction between session number and condition was insignificant for both accuracy rate (F(10, 98) = 0.86, p = .57) and reaction time (F(10, 98) = 1.68, p = .096).

The descriptive statistics of the motivation parameters are displayed in Table V-9. To analyse the effect of the experimental conditions on the motivation parameters, linear mixed effects models were used so as to be able to deal with the unbalanced (incomplete) data.

Perceived competence was not affected by the experimental condition when considered by itself (F(2, 24) = 2, p = .11) (51 observations for 25 subjects). This was also the case when observed difficulty (i.e. mean accuracy) was added to the model (F(2, 14) = 1, p = .41) (34 observations for 17 subjects). However, when perceived difficulty was added to the model, the condition did impact perceived competence differently (F(2, 23) = 4, p = .03) (51 observations for 25 subjects). Post-hoc contrasts of the third model (i.e. with perceived difficulty as an extraneous variable) showed that there was only a significant difference between conditions A and B (p < .01).

As for perceived immersion, there was a significant effect of the experimental condition (F(2, 24) = 5, p = .01). Here, the post-hoc contrasts revealed a significant difference only between conditions A and C (p < .01).

Table V-9: descriptive statistics of motivation data

	condition A	**condition B**	**condition C**
perceived competence	4.0 (*SD* = .98)	4.6 (*SD* = .95)	4.3 (*SD* = .9)
perceived immersion	2.9 (*SD* = .89)	3.4 (*SD* = 1.14)	3.8 (*SD* = 1.42)

Further, a correlation analysis was carried out in order to investigate the relationships between perceived competence and immersion on the one hand, and interest/enjoyment (i.e. intrinsic motivation) and willingness for future practice on the other hand (see Table V-10). A medium-sized correlation was found between perceived competence and interest/enjoyment (r = .41, p = .01), but there was no relation between perceived competence and willingness for future practice (r = .31, p = .06). Immersion was strongly correlated to interest/enjoyment (r = .68, p < .01) and to willingness for future practice (r = .57, p < .01). Finally, interest/enjoyment correlated strongly with willingness for future practice (r = .68, p < .01).

Table V-10: Pearson's *r* correlation coefficients for the motivation parameters, adjusted for multiple comparisons using Holm's method (** *p* < .01 ; * *p* < .05)

variable	1.	2.	3.	4.
1. perceived competence	—	.02	.41*	.31
2. perceived immersion		—	.68**	.57**
3. interest/enjoyment			—	.68**
4. willingness for future practice				—

5.6 Discussion

The logging data show that learners' average performance was slightly above the chance level (62%). This implies that learners were, on average, presented with CF for 38% of their judgments. Together with the high number of judgments per practice session (around 21 judgments per 60 seconds on average), this entails that learners received relatively large amounts of CF. Further, the mean accuracy rates did not increase with practice, which suggests that the task indeed was difficult for the learners.

Moreover, while the results show that, on average, learners' performance remained stable in terms of accuracy, they also show that learners responded twice as quickly at the end of practice, suggesting that learners may have engaged in guessing behaviour, and that they may have felt rather helpless. An alternative explanation for the learners' guessing behaviour concerns the reward mechanics and time pressure. One participant was observed repeatedly pressing the F and J keys. When the researcher confronted him with this after the practice sessions, the participant replied he had found out that this gained him more points than applying the grammar rules.

As for research question 1, the results of the questionnaire data suggest that fantasy affected learners' perceived competence in positive ways, and that fantasy with vivid CF increased learners' sense of immersion. However, this study produces no evidence that vivid CF by itself (i.e. the difference between conditions B and C) had a positive impact on learners' sense of competence or immersion.

For the second research question, the results imply that learners were more intrinsically motivated when they experienced higher levels of competence and immersion. Moreover, learners who felt more immersed were also more willing to practise in the future. No relation was found between perceived competence and immersion. This disconfirms the PENS model, and also does not back the claim immersion in a 'safe environment' may help to cancel the potentially negative effects of CF on perceived competence.

The follow-up interviews shed some more light on the questionnaire results. None of the six interviewees preferred the condition with fantasy and vivid CF (i.e. condition C): five students preferred the condition with fantasy and plain CF (B), and one student preferred the condition without fantasy (A). Learners reported that vivid CF made them feel they lost more time in the speeded practice task—this wasn't the case, in objective terms—and that the animations and sound effects were distracting and even frustrating. The logging data confirm that learners may have been distracted in condition C: the mean and standard deviation of the reaction times were highest in this condition.

However, one student reported that he had engaged in 'gaming behaviour': he had intentionally sought out the vivid CF, just to see what would happen if he made a mistake, and said this was fun because the vivid CF was varied.

> Learner: I thought the second version was best for mistakes, because in the third version, it was like … let's make a mistake, so that it shows a funny effect. (Laughs)
> Interviewer: Why do you do that?

> Learner: I just thought, sometimes it was fun … the same effect, or another one.
> Interviewer: And if it wasn't the same effect?
> Learner: Then I wouldn't do that anymore.

While this is consistent with reports of gamers that actively seek out failure during play (McGonigal, 2011), designers of game-like learning environments may need to be wary of using vivid CF, as it may result in cognitive load that hinders learning (see also deHaan, Reed, & Kuwada, 2010).

5.7 Conclusion

The findings of this study produce tentative support for fantasy and vivid CF to support learners' motivation in technology-enhanced controlled practice, which may stimulate practice in self-directed contexts. This is important in a Skill Acquisition Perspective on L2 learning (DeKeyser, 2008), which posits that extensive periods of practice are needed for the development of automatized knowledge of specific linguistic constructions as a part of the development of holistic L2 skills. This implies that the game attributes of fantasy and vividness need to be considered in the design of tutorial CALL feedback.

The main limitation of this study is its reliance on self-report. Behavioural measures of engagement, such as observations of learners in play, actual continued practice, or perhaps psychophysiological measures such as skin conductance or heart rates in reaction to exposure to feedback, may provide a less biased picture. Another limitation is the small sample size. Studies with more participants could apply path analysis to more clearly determine the relations between perceptions of competence, immersion, intrinsic motivation and willingness for future practice.

Lastly, the animations and sound effects of CF in this study were irrelevant with respect to the learning content, presenting no information that could inherently support learning. Future studies could explore ways in which animations and sound effects in CF can be 'intrinsically integrated' with the instructional content (Habgood & Ainsworth, 2011). For instance, in a mini-game on two-way prepositions in German, CF could visually render all the different ways in which incorrect use of a two-way preposition results in undesirable but arguably funny events in the game world (cfr. mini-game described in Wylin & Desmet, 2005).

Acknowledgements

The conceptual design of the technology-based practice environment used in this study, but not its development, was partly realized through interaction with the Games Online for Basic Language learning (GOBL) project (519136-LLP-2011-NL-KA2-KA2MP), funded with support from the European Commission. This publication reflects the views only of the author, and the Commission cannot be held responsible for any use which may be made of the information contained therein.

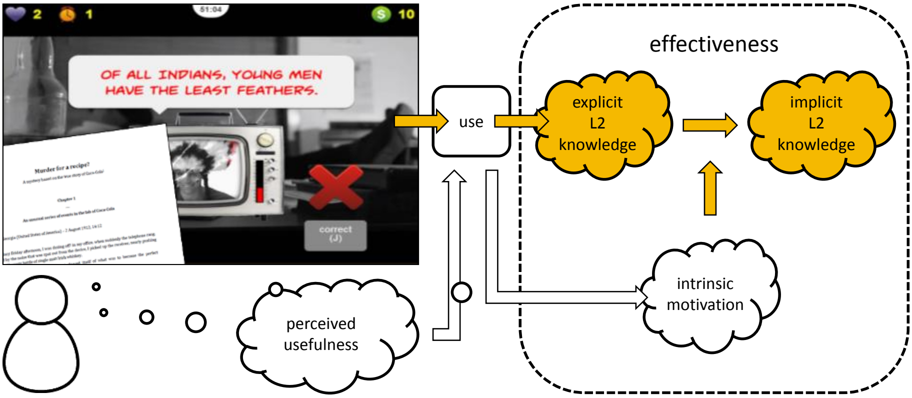
The PhD candidate developed and deployed the technology through the web application framework Google App Engine. The client-side application was built with HTML5, and relied on techniques and some source code from the *HTML5 Game Development* course from Udacity (CS255). The lighting animation is from IndieGameDesigner.com; the alien animation is from netanimations.net. Icons are from the Coquette icon set. The two versions that comprised the detective fantasy contained a still from the film noir project *The Big Smoke* (copyright Kenneth Gawne). The omission of any other due credit is entirely unintentional.

We wish to thank dr. Mieke Vandewaetere for her kind methodological advice.

Fourth interlude

The previous chapter addressed two challenges inherent in designing effective technology-enhanced activities for controlled practice, namely the challenge of engendering intrinsic motivation, so that learners are willing to practise language and remediate specific problems outside of classrooms and without much regulation by teachers, and the related challenge of rendering corrective feedback in controlled practice in such ways that feedback supports, rather than undermines, learners' intrinsic motivation. We investigated these challenges by means of an effectiveness study with a mini-game, and found that fantasy and vividness of corrective feedback have some potential—if well designed—to afford intrinsically motivated behaviour.

The next chapter continues on this thread, and reports on the final empirical study of this PhD project, which was intended to investigate whether extensive periods of controlled practice resulted in automatized knowledge that is considered useful for performance on more holistic language tasks.



From the point of view of design, the study tackled the third challenge related to controlled practice activities that was introduced in the previous chapter but only minimally addressed in the design of the practice environment, namely to involve learners in meaningful language processing. We did this by embedding the practice tasks in an authentic text that was read and discussed in class. Further, the vivid animations and sound effects in response to incorrect responses, as given form in the previous study with mini-games, were replaced by less salient animations and sound effects, so that they would not interfere so much with learning. Finally, in order to prevent deliberate guessing behaviour, as observed in the third study, the number of incorrect responses allowed during one practice session of 60 seconds was limited to five.

Chapter VI

Empirical study 4: Effectiveness of metalinguistic corrective feedback for second language grammar learning supported by mini-games

6.1 Introduction

In current-day second language (L2) learning and teaching, the power of some kind of focus on form is undisputed, preferably in complex and communicative tasks in the L2, embedded within meaning-oriented language use, disrupting the communicative flow to a minimal degree, and focusing on forms that are psycholinguistically relevant and necessary for the communication to succeed (Doughty & Williams, 1998). Teachers trained in communicative L2 pedagogy make wide use of such implicit focus-on-form techniques in an attempt to react to grammatical errors, most notably of implicit corrective feedback (CF) (e.g. Lyster & Ranta, 1997; Sheen, 2006). In contrast with implicit focus on form embedded within meaning-oriented L2 use, there is little scope in current L2 education programmes for controlled practice of specific linguistic constructions (ranging from concrete lexical items to more complex grammatical schemata) accompanied by explicit CF, and equally embedded within meaning-oriented L2 use. Yet, such practice may help to automatize knowledge about lexical, morphosyntactic, and phonological, and pragmatic aspects of the L2 in implicit memory, which could result in effortless and target-like performance in the L2 while freeing up attentional resources for higher-order skills during complex tasks (Segalowitz & Hulstijn, 2005). Hence, research on the effectiveness of meaningful controlled practice for L2 learning is relevant both from a theoretical and pedagogical perspective.

There are a number of methodological challenges, though, for experimental research into the effects of controlled practice on the development of automaticity in a L2. First, in order to carefully manipulate features of the learning environment and measure learner performance accurately, the available studies have typically taken place in laboratory settings, sometimes with artificial languages (e.g. DeKeyser, 1997; Robinson, 1997), thereby compromising the ecological validity of such research. A second requirement is to provide learners with consistent CF throughout practice, which is possible in laboratory research but impractical if not infeasible in classroom settings. A third challenge—for pedagogy and ecologically valid research alike—is to move beyond mechanical L2 practice, by coupling automatization through repetition with meaningful information processing in highly contextualized L2 use (DeKeyser, 1998; Segalowitz & Hulstijn, 2005). In essence, the challenge for research on L2 practice is to bridge controlled experimental trialling and L2 learning in real classrooms guided by prevailing pedagogical principles.

Claims have been made that computer-assisted language learning (CALL) holds great promise for the future of research on practice, as it allows for massive and fine-grained data collection on L2 performance in longitudinal experimental designs, potentially in externally valid ways (DeKeyser, 2007a). In CALL, moreover, CF can be delivered much more consistently than in face-to-face settings, and *ad infinitum*. So, research on practice by means of CALL tools might offer an answer to the three methodological challenges outlined above: careful manipulation and control of the learning environment in ecologically valid settings, consistent CF, and—pending instructional designs that couple form focus with meaning focus—embedding controlled practice in meaningful L2 use.

The current experimental study investigated the effects of controlled practice supported by explicit CF on L2 grammar learning in a time frame of two months. Practice lasted one month, and took place in L2 classrooms and in learners' home settings. Practice was operationalized by means of mini-games (Cornillie & Desmet, n.d.), and was embedded within meaning-focused reading and discussion activities. The integration of mini-games and authentic L2 texts affords, first, controlled practice of specific linguistic constructions with a great deal of repetition, supported by consistent explicit CF. Moreover, practice supported by mini-games and meaning-oriented tasks holds the potential to involve learners in meaningful L2 processing, and to support learners' intrinsic motivation to practise (Ryan & Deci, 2000).

6.1 Theoretical issues and previous research

This section gives an overview of the theoretical issues concerning controlled practice. We first define the term 'controlled practice', and then present two competing views on automatization in a L2 through controlled practice. This is followed by an overview of empirical research on L2 automatization. We then argue for the added value of metalinguistic information provided in controlled practice, and note that the complexity of this information needs to be taken into account. We conclude with the potential of mini-games for the future of L2 practice, and note methodological limitations of previous research.

6.1.1 Defining controlled practice

The term 'controlled practice', often used interchangeably with the more controversial term 'drill (and practice)', refers to all activities in a L2 that focus on specific linguistic constructions and that involve a considerable amount of recycling of constructions, feedback, and often time pressure, with the goal of developing explicit knowledge about these constructions as well as skills in the L2 (DeKeyser, 2007b). Controlled practice is often associated exclusively with narrow forms of drilling, but in fact, it covers a wide range of activities in the L2 that likely involve quite different kinds of L2 processing. Activities for controlled practice vary, first, with respect to the concurrence (or dissociation) of form and meaning processing. In this respect, controlled practice may be mechanical, in which case L2 learners do not need to process the meaning of the utterance to complete the task; it may be meaningful, namely when the task requires the learner to comprehend the L2 on both a structural and semantic level; or it may be communicative, in which case learners need to convey personal meaning rather than reproduce prefabricated and highly predictable responses. The key difference between the latter two types of drills is that the teacher does not know in advance which utterances to expect. For more information on the distinction between meaningful, mechanical, and communicative drills, we refer to Paulston & Bruder (1976); for their respective utility for L2 development see DeKeyser (1998) and Wong & VanPatten (2003).

Further, while the term ‘practice’ is sometimes reserved to mean output practice only, controlled practice can involve receptive skills or productive skills. Next, practice can be oral or written, and focus on various formal aspects of the L2 (phonological, morphological, syntactic, or lexical form). Thus, controlled practice can take many forms, which are likely to induce different cognitive L2 processes. In what follows, we will use the term ‘controlled practice’ in its broadest possible sense.

Even though activities for controlled practice are quite varied, critiques typically focus on one specific type of controlled practice, namely productive oral pattern drills that are mechanical in nature. Lightbown (2008) aptly describes such activities as “the kind of mechanical drill in which students repeat sentences that are related only by the fact that they share some grammatical pattern” (pp. 28-29). As is well known, such drills were popularized by audiolingualism, and have been shown to be ineffective and sometimes even disadvantageous for the development of communicative L2 competence (for a review see Wong & VanPatten, 2003), i.e. the ability to express personal meaning fluently and accurately. Oral mechanical drills as championed by audiolingualist approaches to L2 teaching indeed fly in the face of current empirical findings on L2 learning and pedagogical practice. They can be disregarded simply on the grounds that they do not engage learners in processing and using the L2 to comprehend and convey meanings, and merely involve them in the learning of meaningless linguistic patterns. With the effects of mechanical drills being clear, little is known about the usefulness of meaningful and communicative drills for L2 learning. The dearth of empirical findings on the latter types of drills can be explained by SLA researchers’ scepticism towards controlled practice since the communicative turn in L2 pedagogy, in particular towards practice in production skills (Larsen-Freeman, 2003, pp. 102–106).

6.1.2 Controlled practice and the development of L2 knowledge: two competing theories of automatization in a L2

Research into automatization in a L2 through controlled practice has been approached either through the theoretical lens of skill acquisition theory (SAT) (DeKeyser, 2008)—known as Adaptive Control of Thought in the more general literature on human development (Anderson, 1992)—or instance theory (Logan, 1988). In this section, we present the main tenets of these theories, point at their differences, and motivate our choice for SAT as the theoretical backdrop for the current study.

A skill acquisition perspective on L2 learning posits that the development of a specific skill is the gradual process of moving through a series of stages that differ with respect to the effort used and the type of knowledge that is relied on to perform the skill. More specifically, the typical trajectory in L2 skill learning comprises three stages. Initially, explicit L2 knowledge, (typically rule-based knowledge about grammatical constructions) is developed through explicit learning. Usually, this is accomplished by providing the learner with information through some form of explicit teaching. Such knowledge is available to awareness and can be verbalized or *declared*, hence the term 'declarative knowledge'. This is followed by a first phase of practice, in which explicit knowledge is applied consciously and with great effort to concrete L2 items. This phase allegedly results in the development of *procedures*, i.e. condition-action pairs which encode the rules in behaviour and comprise knowledge on what needs to be done under specific circumstances. The main advantage of this proceduralization stage is efficiency of retrieval: knowledge about a particular L2 construction becomes available as a ready-made chunk in memory, ready to be called upon when the conditions for its use reoccur, and to be retrieved quickly. The final stage consists of automatizing procedural knowledge through continued practice, which requires increasingly fewer conscious cognitive resources, and sometimes even leads to loss of initial explicit knowledge. The benefit is that over time, controlled practice results in knowledge that is robust and "accessible in the same way as implicitly acquired knowledge" (i.e. knowledge developed through implicit learning) (DeKeyser, 2005, p. 328), as manifested in fluent and error-free performance.

The downside of the skill acquisition process is that automatization is 'trapped' in a specific skill: it does not transfer well to other skills (DeKeyser, 1997). In L2 development, the skill-specificity of automatization explains why a good L2 writer is not necessarily a fluent L2 speaker, and that learners trained according to audiolingual principles are not necessarily good at communicating in the L2—the reason being that the latter are primarily trained in parroting form-form mappings rather than in conveying personal meanings.

In a nutshell, skill acquisition in a L2 is intended to culminate in the development of knowledge that is virtually indistinguishable from implicit L2 knowledge—this knowledge is sometimes even labelled 'implicit' (Hulstijn, 2007)—as evidenced by ever more accurate and fluent performance as a function of the amount of practice, "while temporarily leaning on declarative crutches" (DeKeyser, 1998, p. 49), and for specific L2 skills. In SAT, thus, explicit knowledge is critical and interacts with implicit processes: "the controlled use of declarative knowledge guid[es] the proceduralization and eventual automatized implicit processing of language as it does in the acquisition of other cognitive skills" (N. C. Ellis & Larsen-Freeman, 2006, p. 569). Yet, the theory also allows for the incidental build-up of an implicit knowledge base without initial declarative knowledge, i.e. as the by-product of communicative L2 use.

Logan's (1988) instance theory of automaticity is different from SAT in its central claim that automatization is memory-based (sometimes called *item-based* or *instance-based*[4]) rather than rule-based. More concretely, the theory hypothesizes that initial performance may be guided by rules, but predicts that with increasing practice, not so much knowledge of the underlying rule (i.e.

[4] In what follows, we will use the term 'item-based'.

procedural knowledge) gets automatized, but rather the retrieval of concrete L2 instances from memory.

While less radical versions of rule-based and item-based theories of automaticity have been proposed (for a review, see DeKeyser, 2001), the two theories remain conceptually opposed. In SAT, declarative knowledge is crucial, whereas it is irrelevant in Logan's instance theory. For the current study on the role of CF in controlled practice, SAT has the most explanatory potential, for CF is inextricably bound to the development and fine-tuning of declarative (metalinguistic) knowledge (Carroll, 1995, 2001), especially if it is output-prompting (i.e. provides only negative evidence).

6.1.3 Empirical research on automatization in a L2

Empirical research on automatization in a L2 is rather scant. In this section, we summarize the results of two exemplary studies.

In an attempt to empirically test SAT in L2 learning, DeKeyser (1997) conducted a computerized and laboratory-based experiment on the effects of meaningful controlled practice on the automatization of grammar rules for comprehension and production skills in an artificial yet natural language-like L2. Over a period of 8 weeks, he observed gradual increases in accuracy rates and decreases in response times for both skills, which followed a power law, consistent with SAT. In line with SAT theory, the effects of practice were highly skill-specific, as comprehension practice did not transfer well to production skills, and vice-versa.

Robinson (1997) investigated the learning of the constraints on dative alternation in English as a L2 from an item-based perspective on automatization. No evidence was found of automatization in the form of increased response times as a function of increasing frequency of presentation, which may be due to the relatively short length of practice (30 minutes). However, the study showed that learners who were provided with declarative (rule-based) information prior to practice (see also section 6.1.4) were significantly more accurate than learners in three other conditions in judging the grammaticality of ungrammatical sentences that were new (i.e. not offered in the practice sessions). The other conditions comprised implicit learning—used in the very restricted sense of reading stimuli and responding to questions on their formal characteristics as in artificial grammar learning studies (e.g. Reber, 1989)—meaning-oriented learning, and meaning-oriented learning with focus on form by way of input enhancement. The instructed group was better than the other groups on new ungrammatical items, which suggests that the declarative knowledge taught initially had helped learners to develop generalized knowledge of the constructions, rather than knowledge of specific items. Moreover, the study documented significantly slower response times for new items than for items offered in practice, confirming Robinson and Ha's previous study (1993), and—perhaps more interestingly—significantly faster response times for the instructed learners on new items than for all other groups, whereas the response times for the practice items did not differ significantly between the instructed and meaning-oriented groups. Taken together, these results imply that the instructed learners may have relied on a different knowledge base for the new items (i.e. rule-based knowledge) than for the practice items (i.e. memory-based knowledge), and that this knowledge helped them in terms of accuracy and in terms of a measure of fluency (response times).

6.1.4 Transfer of controlled practice, and the role of declarative knowledge and CF

In section 6.1.2, it was noted that the effects of practice are skill-specific according to SAT. In order to promote transfer of practice from one skill to another, SAT emphasizes the importance of, first of all, similar task conditions (e.g. receptive vs. productive skill, oral vs. written mode). This relates to the issue of transfer-appropriate processing: learners are likely to perform better at tasks that involve cognitive L2 processes similar to those active in training tasks (Lightbown, 2008). Secondly, SAT posits a role for declarative knowledge for enabling transfer. As declarative knowledge is abstract, it may help to bridge the differences between tasks that are dissimilar in terms of cognitive processes: "knowledge that is overly contextualized can reduce transfer; abstract representations of knowledge can help promote transfer" (Bransford et al., 2000, p. 41). DeKeyser notes that L2 instruction needs to foster both declarative and procedural knowledge, since "solid abstract declarative knowledge [...] can be called upon to be integrated into much broader, more abstract procedural rules, which are indispensable when confronting new contexts of use" (1998, p. 100).

Declarative knowledge taught before practice may not be sufficient, however, and may need to be repeated during proceduralization, or even during automatization. As Nick Ellis (2005a) phrased it: "as with other implicit modules, when automatic capabilities fail, there follows a call recruiting additional collaborative conscious support" (p. 308). During performance, this need for conscious support may be satisfied by CF, which is considered an essential component of practice (Leeman, 2007). For some learners at some stages of development, limited CF in the form of 'knowledge of results' may suffice, such as signals of communication breakdown in face-to-face settings (i.e. requests for clarification; e.g. Lyster & Ranta, 1997; McDonough, 2007), or disconfirmations of grammatical accuracy in computerized controlled practice (Schulze, 2003). Such prompts may remind them of the declarative information taught before practice. For other learners, and perhaps for more complex grammar rules, declarative information may need to be repeated, rephrased, or elaborated upon, which calls for more elaborate CF that comprises metalinguistic clues or explanations.

As a number of meta-analyses on CF have shown in recent years (Li, 2010; Lyster & Saito, 2010; Russell & Spada, 2006), different types of CF have often proved to be differentially effective. More specifically, there is some support that CF which comprises metalinguistic explanation can be quite effective, although the findings are somewhat equivocal, potentially due to the many variables that interact with CF type.

Carroll & Swain (1993) compared, among other CF types, the relative effectiveness of knowledge of results (KR) CF and metalinguistic CF in controlled practice activities on English dative alternation, which required Spanish-speaking learners to decide whether a particular verb alternated and to orally produce grammatical sentences accordingly. They found that CF helped learners to outperform a control group on recall tasks for both practice items and new items, and that the metalinguistic CF group did better in terms of accuracy than the KR group, a group that was supported by recasts (a communicative type of CF which comprises knowledge of results feedback and model responses), and a group that received implicit prompts. Of further interest is the classroom-based study by R. Ellis, Loewen, & Erlam (2006), which found metalinguistic CF to be significantly more beneficial than recasts for realizing transfer from focused practice tasks on English past tense use to delayed tests that were intended to measure explicit and implicit L2 knowledge. This study also provided evidence of generalization of practice to new items on post-tests, which applied particularly to the metalinguistic CF group. Taken together, this research indicates that metalinguistic CF can be powerful for enabling learners to develop generalized, rule-based knowledge, as well as transfer their learning from practice tasks to follow-up tasks. Additionally, R. Ellis *et al.*'s (2006) study suggests that future research needs to consider measuring both explicit and implicit L2 knowledge on transfer tasks.

6.1.5 Structural complexity and complexity of declarative information

A variable that is likely to impact on the effectiveness of controlled practice with CF is the linguistic complexity of the structures that are being practised. The problem with linguistic complexity is that it is hard to define (DeKeyser, 1998). As a result, experimental research on the role of linguistic complexity in explicit instruction has produced mixed results so far (Graaff & Housen, 2009).

Closely related to linguistic complexity is the conceptual complexity of the declarative information available to learners, which may or may not mirror linguistic complexity, depending on whether the regularities are clear-cut or rather probabilistic. For obvious reasons, declarative information presented in pedagogical grammars often simplifies linguistic patterns. The predominant position is that simple rules make the best candidates for explicit instruction (DeKeyser, 1998). Hulstijn & de Graaff (1994) argue that the advantage of explicit instruction is greater in the case of complex rules than in the case of simple rules, because learners are likely to pick up simple formal phenomena spontaneously. However, this argument seems to pertain more to the efficiency of classroom time than of absolute effectiveness—each minute spent on a simple rule cannot be spent on a more complex rule. Further, Hulstijn (2007) proposes that in the foreign language classroom, grammar rules need to be as short and simple as possible, as human beings can only handle a limited amount of declarative knowledge at a time.

6.1.6 The potential of mini-games for the future of L2 practice

As we have seen, more practice means more recycling of constructions, and greater potential for automatization. Yet, given time constraints, controlled practice in classroom L2 learning is seen as problematic (DeKeyser, 2007a), so the question arises as to how learners may be engaged in practice activities outside of classrooms, without much external regulation by their teachers.

Dörnyei (2009) argues that "the key to the effectiveness [of controlled practice] is to design interesting drills that are not demotivating" (p. 289), and sums up a varied range of techniques such as using variation in the recycling of utterances, making drills personally relevant to learners, or using CALL or games. Clearly, controlled practice in a L2 that is driven by intrinsic motivation, with which digital game-based learning environments have long been associated (e.g. Malone, 1981), may be quite powerful for the development of automatized L2 knowledge. Good game-based instructional designs that satisfy

L2 learners' basic psychological needs (i.e. the need for competence, autonomy, and relatedness) and which induce a sense of immersion in the virtual world (Ryan et al., 2006) may lead to self-catalysed behaviour in practice.

Here, there might be clear value in mini-games for language learning (Cornillie & Desmet, n.d.). Such games can be played in brief sessions, and are constrained in scope. Hence, they seem particularly relevant for the focused practice of lexical, grammar, or pronunciation skills. Moreover, they give immediate feedback, which offers potential for fine-tuning declarative knowledge, and are fast-paced, which may urge learners to develop automaticity of skill. If such games can support learners' intrinsic motivation, and in turn their willingness to practice outside of the classroom, they may have an essential role to play in the future of language education.

To the best of our knowledge, no empirical studies have addressed automatization of grammar skills through mini-games. However, strong empirical support for the usefulness of mini-games to further lexical development in a L2 comes from Cobb & Horst (2011). Using the suite of mini-games shipped with the popular commercial *My Word Coach* series designed for explicit vocabulary learning, the authors carried out an ecologically valid experiment with 50 young English L2 learners in Canada. The various games focused on both form and on form-meaning connections. Growth was measured using a battery of pre-and-post-tests that targeted form recognition, meaning recognition, free production, and speed of lexical access. Two months of game use resulted in huge gains in vocabulary recognition in comparison with normal vocabulary growth, increased speed of lexical access, and more use of English words in a storytelling task. The study shows that intensive practice with mini-games helped to develop knowledge both at the declarative level (larger vocabulary recognition) and at the procedural level (faster access to already known words), and that the practice effect transferred to the more complex skill of storytelling. From the perspective of SAT, the latter result is particularly noteworthy, given the differences between the skill applied in practice (written comprehension) and the skill used in the follow-up storytelling task (spoken L2 production).

6.1.7 Methodological limitations of previous research

While discussing the findings of his study, Robinson (1997) notes a general limitation of laboratory-based studies on controlled practice:

> It may have been that the use of artificial verbs interfered with the process of interest in the incidental and enhanced conditions by making it difficult to process the sentences for meaning in the way natural language samples are processed. Similarly, restricting the stimuli to sentences, rather than extended text, may have reduced the challenge they posed to the learner, causing only shallow processing, whereas deeper processing may have produced different results (P. Robinson, 1997, p. 243).

DeKeyser has similarly insisted on several occasions that practice which is not genuinely meaningful does not qualify as productive for L2 development, and that the findings of lab-based studies should not be used as recommendations for L2 pedagogy (1997, 1998, 2007b). Further, he has made pleas for research that "combines the degree of control of a psycholinguistic experiment with the validity of research on real second language learning, and [...] that, on top of that, takes a process, that is, a developmental, longitudinal, perspective" (DeKeyser, 1998, p. 60).

6.2 The current study

The current study was an attempt to move beyond the state-of-the-art in research on controlled practice in two ways, namely on a methodological level, and in terms of theoretical issues. First, on a methodological plane, the current study tried to combine controlled experimental trialling with ecologically valid and meaning-oriented L2 practice. This was done by means of a text, read and discussed in authentic classes for its meaning, of which the content served as the basis for practice with mini-games in class and in learners' homes. Practice was logged by the system in order to control for the effect of time on task.

On a theoretical level, the objective of the study was to compare the relative effectiveness of 'knowledge of results' CF and metalinguistic CF provided in the mini-games on transfer in terms of task type (i.e. tasks that are considered to draw on different types of L2 knowledge) and on transfer in terms of content (i.e. generalization transfer; items offered during practice vs. new/non-practice items). Further, we wanted to compare the effectiveness of practice with CF for grammar problems of which the declarative rules varied in complexity. The study aimed to answer the following research questions:

1. How does meaningful receptive practice with mini-games and CF affect the development of L2 grammar knowledge?
2. How does the type of CF, namely metalinguistic (ML) CF or knowledge of results (KR) CF, influence transfer of practice on the development of L2 grammar knowledge?
3. How does practice result in generalized knowledge?
4. Are the effects of practice with CF differentially effective for grammar problems that vary in declarative rule complexity, and if so, how?

The following hypotheses were formulated:

1. Practice results in automatized L2 grammar knowledge as manifested in high accuracy rates and fast response times on transfer tasks.

2. The effects of practice supported by ML CF will be higher than the effects of practice with KR CF on transfer tasks that allow for the use of explicit knowledge and thus allow learners to monitor their performance. There will be no such difference on transfer tasks which are thought to preclude the use of explicit knowledge.
3. On tasks that are considered to prevent the use of explicit knowledge, the effects of practice will be higher for items offered in practice than for non-practice items. For the former items, namely, there will be a memory effect of practice.
4. The effects of practice will be stronger for grammar problems which comprise simple rule explanations.

6.3 Method

6.3.1 Research design and participants

The study was carried out over a period of two months, and adopted a between-subjects experimental design with repeated measures. The participants were Dutch-speaking learners of English in the fifth and sixth form of general secondary education in Flanders, Belgium. Typically, these learners are between 16 and 18 years old, and have received three or four years of formal instruction in English, respectively, which is meant to correspond roughly to the intermediate level (B1) of the Common European Framework of Reference for Languages (Council of Europe, 2011). The participants were drawn from 11 intact classes, eight of which were assigned to the treatment group (*N* = 125). The remaining three classes formed a control group (*N* = 61), which was included to account for the potential effect of the repeated tests, as well as that of regular classroom instruction between the tests. Participants in the treatment group were assigned at random to one of two treatment conditions, which differed with respect to the type of CF that was provided during practice (knowledge of results vs. metalinguistic CF). Assuming that classes may differ in terms of prior knowledge, language aptitude, or motivation, assignment to the treatment conditions was handled in this way in order to even out the potential confounding effect of the class on the treatment type.

6.3.2 Target problems and grammar instruction

The treatment consisted of learning and practising the constraints on quantifiers in English (henceforth QNT) and 'verbs with two objects' (V2O), known more commonly as dative alternation (Carroll & Swain, 1993) [5]. These particular linguistic problems were chosen for three reasons: they are frequent in English, the constraints on their use do not apply in Dutch, and ungrammatical realisations of these problems are frequent in the interlanguage of Dutch-speaking learners, particularly for QNT (Tops et al., 2001). Furthermore, the two problems differ in linguistic complexity, which could determine how learners may benefit from explicit instruction (see section 6.1.5). Additionally, acquisition of the principles underlying V2O is considered difficult for L1 and L2 learners alike (Gropen, Pinker, Hollander, Goldberg, & Wilson, 1989; Mazurkewich & White, 1984), and, given the high productivity of V2O in English use, it is a typical example of a learnability issue (i.e. the fact that learners come to know and produce more sentences than they are exposed to in the input). Hence, for L2 learners in particular, acquisition of V2O is considered to require negative evidence, for instance in the form of systematic practice with negative feedback (Carroll & Swain, 1993). V2O was also selected because L2 learners acquire it late, and because it is typically not instructed in L2 English curricula (R. Ellis, 2009a). The participants of this study had not received any overt instruction on V2O prior to the experiment. Thus, their knowledge of this problem was likely to be implicit only.

[5] In fact, this is an unfortunate label. What Carroll & Swain (1993) and others (Gropen et al., 1989; Mazurkewich & White, 1984) have labelled 'dative alternation' comprises both the dative alternation (with the preposition *to*) and the benefactive alternation (with the preposition *for*) (e.g. Levin, 1993).

For the instruction of QNT, we adopted a conservative approach for the constructions *less* + uncountable vs. *fewer* + countable, and *least* + uncountable vs. *fewest* + countable. Recognizing that communicative English L2 grammars often relax the rules for this distinction—with the quantifier *less* in particular becoming increasingly more frequent in combination with countable NPs in present-day English use, especially in informal registers (Leech & Svartvik, 1994, pp. 50, 273–274, 360)—we chose to instruct the rules as they were typically described in course books supported by the Flemish curriculum for English, as these focus on formal registers of the L2 with which learners are less familiar on the basis of out-of-class learning. More specifically, we told learners that *less* and *least* were becoming more frequent in English with countable nouns, and that, if they applied the rules strictly, they couldn't go wrong. Teaching exceptions to these rules might have confused the learners, and is likely to have compromised our interest in the relative effectiveness of controlled practice for grammar problems that vary in complexity.

The grammar explanations for DOC were based on Carroll & Swain (1993), which traces back to the formal analysis presented in Mazurkewich & White (1984). This analysis holds that the syntactical alternation between the double object construction (DOC) and the prepositional construction (PC) is governed by two constraints: a morphophonological constraint and a semantic one. The participants of the current study were instructed that the DOC was only possible with short (i.e. one-syllable) verbs or with longer verbs (two syllables and more) which had initial stress and if the sentence expressed transfer of possession, and that all other cases required use of the PC. The formulation of the morphophonological constraint is a pedagogical simplification, though. The more accurate explanation is that "the core class of alternating verbs have exactly one foot, e.g. (give), a(ssign), whereas the nonalternating verbs have two or more feet, e.g. (ex)(plain) and (do)(nate)" (Anttila, Adams, & Speriosu, 2010). According to the instructions of Carroll & Swain (1993), which is based on syllabic stress, verbs such as *assign, allow, award, extend, permit, reserve*, etc. would not alternate. For pedagogical reasons, though, we stuck to the simplified rule, and avoided verbs that formed exceptions to this rule. Moreover, we also told participants that the PC was always grammatically correct, whereas there are PCs that are certainly less frequent or even marked. More recent descriptive research of the distribution of the DOC and PC shows that actual usage is much more probabilistic (De Cuypere & Buysse, n.d.), with a host of factors determining speakers' preference for one construction over another in cases where they were both possible according to our instructions. Evidently, pedagogical grammars strive for different objectives than descriptive grammars, and as long as the general tendencies are reflected in the rule, simplifications of the probabilistic regularities are not too problematic.

Table 1 presents an overview of the constructions for each of the two linguistic problems, along with sample sentences used in the practice materials.

Table VI-1: overview of the constructions

linguistic problem	**construction**	**sample sentence from practice materials**
quantifiers	*many, few, fewer, fewest* + countable	*Charley has fewer shares in the company.*
	**much, little, less, least* + countable	**Copies of Coca-Cola use less ingredients.*
verbs with two objects	$V^{\text{monosyllabic}}$ + NP + NP	*John Pemberton taught me some tricks on how to make Coca-Cola.*
	$V^{\text{polysyllabic, initial stress}}$ + NP + NP	*Father promised me the rights to the name 'Coca-Cola'.*
	*$V^{\text{polysyllabic, final stress}}$ + NP + NP	**Pemberton revealed me the secret formula.*
	*$V^{\text{no transfer of possession}}$ + NP + NP	**Legend says that Charley stirred his father the first brew of Coca-Cola.*

6.3.3 Experimental procedure and practice materials

Data collection took place from January until March 2014. Figure VI-1 provides an overview of the experimental procedure. In phase 1, after being invited to participate in a study on learning English grammar, the participants filled out consent forms, and completed a computerized pre-test, comprising a timed grammaticality judgment test (TGJT) and a written discourse completion test (WDCT) (see section 6.3.4). Because this study concerned a natural language, in contrast with e.g. DeKeyser's (1997) study which used an invented miniature linguistic system, any effects of prior knowledge needed to be controlled for, especially since assignment to the control and treatment groups could not be done at random.

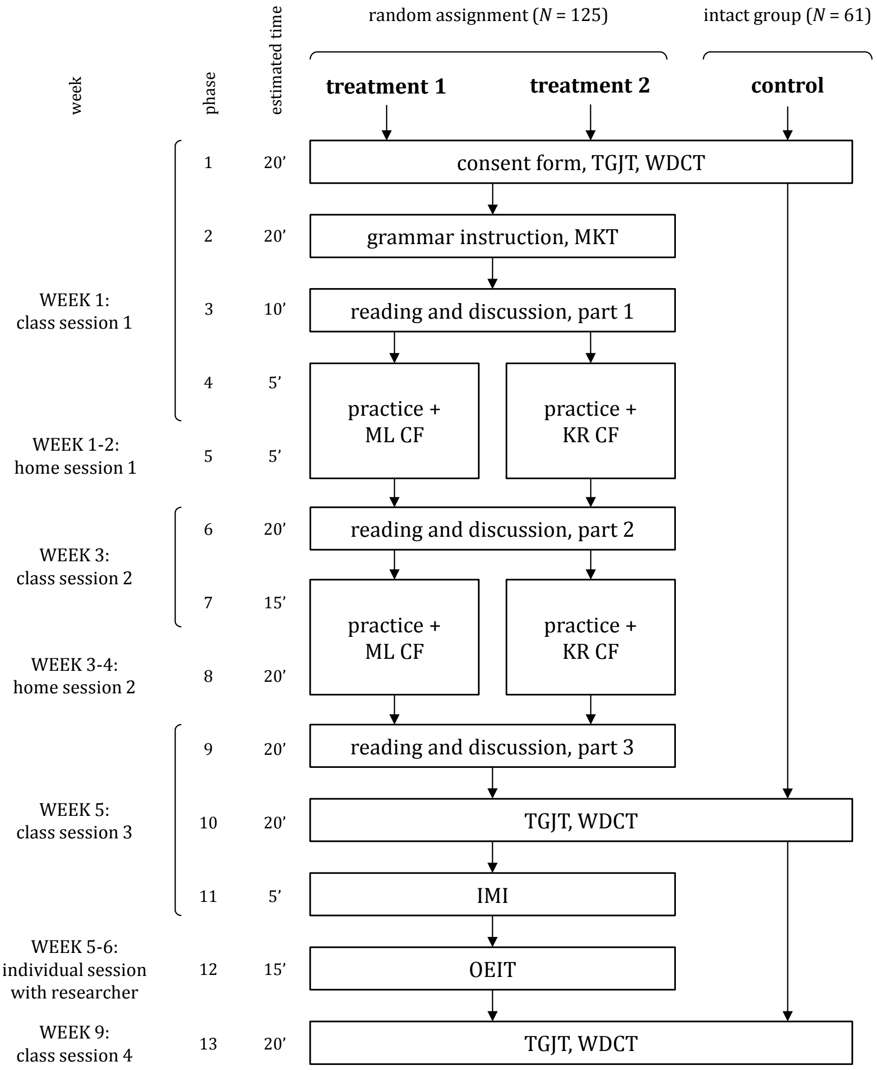
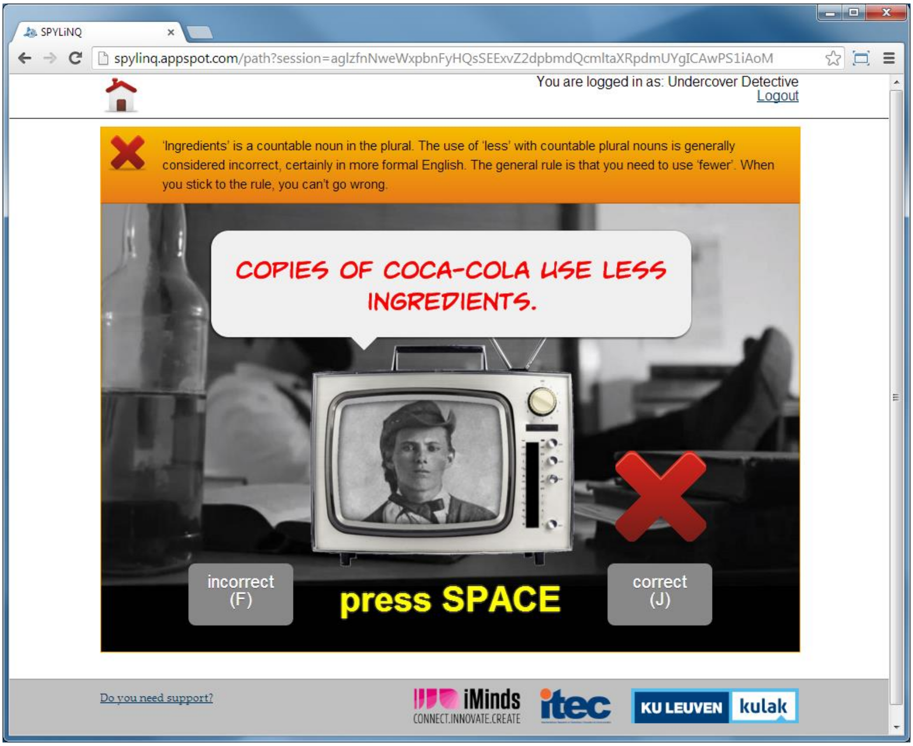


Figure VI-1: experimental procedure

In phase 2, the researcher provided explicit rule explanation on the two grammatical problems to the two treatment groups. In Skill Acquisition terms, this step comprises the declarative knowledge building phase. Rule instruction was done inductively in order to engage the learners more actively, in accordance with recommendations from Ranta & Lyster (2007, pp. 150–151), but relied on decontextualized exemplars, which were distributed to learners on paper (see Appendix 3). The metalinguistic terminology used in this phase was consistent with the learners' course books, and parts of the rule instruction were repeated or elaborated in the learners' mother tongue on an as-needed basis. After the inductive and collaborative identification of the rules, the researcher provided explicit instruction supported by schemata projected for the entire class (see Appendix 4). Subsequent to the rule instruction, learners took a metalinguistic knowledge test (MKT) (see section 6.3.4.3). This was considered necessary, as it might explain how learners proceduralized this knowledge during the practice phase, and how they might benefit from CF. Learners were allowed to take notes during the entire rule instruction phase, but were asked to return their sheets with exemplars and any notes before taking the MKT.

Following this phase, learners in the treatment groups read the first section of a mystery text in the L2 written by the researcher and based on the early history of Coca-Cola (Pendergrast, 1997) (see Appendix 5), answered comprehension questions, and discussed the text in class. The function of the reading comprehension and discussion tasks was to introduce the background context of the linguistic constructions that were to be practised, so that during the practice tasks, learners were more likely to also process the constructions for their meaning. Hence, this reading task immediately preceded the first practice tasks.

In phases 4 and 5, learners in the treatment groups practised the grammar problems by means of a mini-game presented on computer. The learners were told that their task was to solve the mystery introduced in the text, and that the mini-game was intended to involve them as the detective in the story, interrogating witnesses and potential suspects by means of a special device called a 'tele-interrogator'. The utterances of the interviewees were presented in written form only, and were drawn from the next chapter of the mystery text, thus offering learners a preview of what was going to happen in the story. Although the format of the mini-game was interrogation, hence involving a focus on meaning, the learners were required to judge whether the sentences were well-formed according to the explicit grammar instruction given earlier. Learners used the J and F keys on the computer keyboard to indicate whether they considered the sentences grammatical or ungrammatical, respectively. This format was designed for two reasons. First, save from its potentially stronger meaning-focus, the format closely mirrors the grammaticality judgment tasks used in pre- and post-testing, allowing to measure near transfer in terms of task type. Secondly, the format was very limited with respect to interactivity: in order to perform the task, learners did not need to take into account any elements on the screen other than the sentences and the CF—such as obstacles to be evaded or targets to be hit. While this might have been a more interesting task for the young learners, more advanced element interactivity might have constituted a dual task condition, detracting from the learning content and hindering transfer. This decision was taken in line with recent research which showed that game interactivity may induce extraneous cognitive load and hamper L2 learning (deHaan et al., 2010).



spylinq.appspot.com/path?session=aglzfnNweWxpbnFyHQsSEExvZ2dpbmdQcmltaXRpdmUYgICAwPS1iAoM

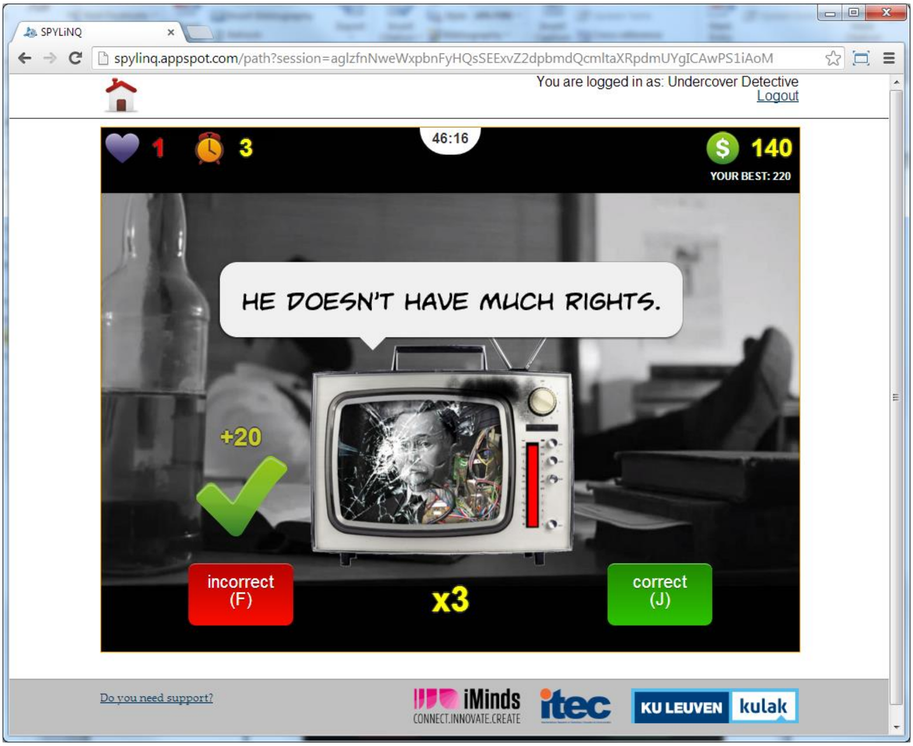
You are logged in as: Undercover Detective

‘Ingredients’ is a countable noun in the plural. The use of ‘less’ with countable plural nouns is generally considered incorrect, certainly in more formal English. The general rule is that you need to use ‘fewer’. When you stick to the rule, you can’t go wrong.

COPIES OF COCA-COLA USE LESS INGREDIENTS.

Figure VI-2: tutorial version of the mini-game, with metalinguistic CF

Learners got two versions of the mini-games to practise with. They first practised with a ‘tutorial’ version of the mini-games, which lacked time pressure and positive feedback (i.e. no points awarded for correct responses), and in which immediate CF was given on their grammaticality judgments. For the learners in one treatment group, this immediate corrective feedback consisted of knowledge of results CF only (KR CF), visualized as a green checkmark or a red cross, with audio support; learners in the other group got metalinguistic explanations in addition (ML CF) (see Figure VI-2). In addition, ungrammatical sentences were highlighted in red as CF for both groups. Practice with the tutorial version of the mini-games was intended to give learners the opportunity to apply their declarative knowledge about the grammatical structures, fine-tune it through interaction with CF, and proceduralize their declarative knowledge. In accordance with SAT, learners need to get ample time in the proceduralization phase; they should not be rushed (DeKeyser, 1998, p. 55). Therefore, the mini-games were not fast-paced. According to DeKeyser, however, proceduralization does not take long: “being required to use a rule a limited number of times to process a set of sentences is all it takes” (2007a, p. 290). Thus, learners practised each grammatical problem once, and separately, by means of 12 sentences (4 grammatical ones and 8 ungrammatical ones). The items for the constructions were offered in a random order. After the completion of these 12 items, learners were shown an overview of their responses with KR CF or ML CF, depending on the treatment condition.



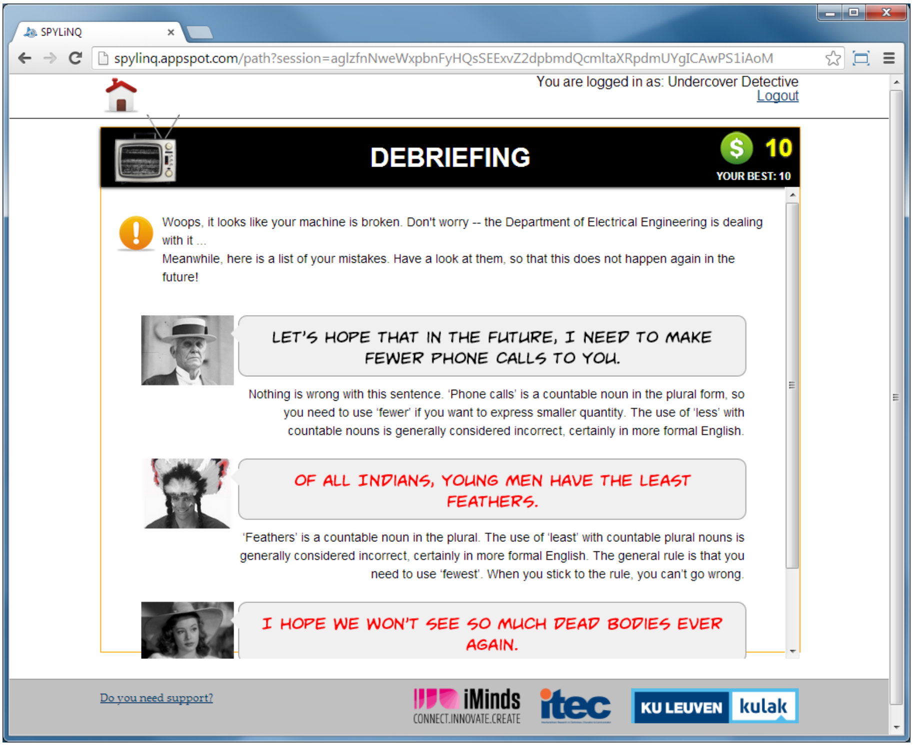
spylinq.appspot.com/path?session=aglzfnNweWxpbnFyHQsSEExvZ2dpbmdQcmltaXRpdmUYgICAwPS1iAoM

You are logged in as: Undercover Detective

HE DOESN'T HAVE MUCH RIGHTS.

Figure VI-3: full version of the mini-game

At the end of the first classroom session, learners in the treatment groups practised with the full version of the mini-games (see Figure VI-3), namely with time pressure, immediate KR CF, positive feedback in the form of points awarded for correct responses, and ‘positive failure feedback’ that was intended to support learners’ motivation when they failed (McGonigal, 2011). This meant that their interrogation device was damaged with each incorrect response, supported by animations and sound effects, and broke down if they eventually made more than five mistakes. If learners made fewer than five mistakes in one and a half minute, the task was stopped. Subsequently, an overview was displayed of the items that were answered incorrectly, allowing participants to further develop and fine-tune declarative knowledge. For the first treatment group, this screen comprised metalinguistic explanations (see Figure VI-4); the other treatment group only saw a list of incorrect responses, with an indication of whether the sentences were grammatical or not. So, both treatment groups saw KR CF in the speeded mini-game, but once the task was stopped, learners in the first treatment group got delayed ML CF in addition.


SPYLiNQ
spylinq.appspot.com/path?session=aglzfnNweWxpbnFyHQsSEExvZ2dpbmdQcmltaXRpdmUYgICAwPS1iAoM
Logout
DEBRIEFING
$ 10
YOUR BEST: 10
Do you need support?
iMinds CONNECT.INNOVATE.CREATE
itec
KU LEUVEN kulak

You are logged in as: Undercover Detective

Woops, it looks like your machine is broken. Don't worry -- the Department of Electrical Engineering is dealing with it ...
Meanwhile, here is a list of your mistakes. Have a look at them, so that this does not happen again in the future!

LET'S HOPE THAT IN THE FUTURE, I NEED TO MAKE FEWER PHONE CALLS TO YOU.

Nothing is wrong with this sentence. 'Phone calls' is a countable noun in the plural form, so you need to use 'fewer' if you want to express smaller quantity. The use of 'less' with countable nouns is generally considered incorrect, certainly in more formal English.

OF ALL INDIANS, YOUNG MEN HAVE THE LEAST FEATHERS.

'Feathers' is a countable noun in the plural. The use of 'least' with countable plural nouns is generally considered incorrect, certainly in more formal English. The general rule is that you need to use 'fewest'. When you stick to the rule, you can't go wrong.

I HOPE WE WON'T SEE SO MUCH DEAD BODIES EVER AGAIN.

Figure VI-4: full version of the mini-game, overview of incorrect responses with ML CF

The time pressure in the mini-games was intended to stimulate learners to automatize learners' receptive knowledge of the constructions. The time allowed to judge each sentence was based on the same approach as for the TGJT (see section 6.3.4.1), with the baseline ranging between 3 and 6.48

seconds depending on sentence length, plus 10 seconds and divided by the square root of a number between 1 and 5. For each series of four consecutive correct responses, this number was raised to a maximum of 5, which increased the time pressure and multiplied their score for subsequent correct responses. Response times and accuracy rates were logged for individual responses, in view of constructing a measure of the degree to which knowledge of the individual linguistic problems was automatized.

After the first class session, learners in the treatment groups got opportunities for additional, voluntary practice with the mini-games at home or at school. The practice materials for QNT and V2O were interwoven, in order to provide equal opportunities for practice for both grammatical problems. The learners used anonymous codes to sign in to the website, and the system continued to log their behaviour. Personal accounts (such as social authentication through Facebook or Google) might have provided more control over whether learners used their own login codes, but this was not done out of a concern with learners' privacy. Learners were asked to practise at least 20 minutes at home before the next class session. In order to engage them, a 'leaderboard' was shown, comparing the learner's personal highest score for each grammatical problem with the five highest scores in an anonymous way, and the participating teachers regularly reminded them of their assignment.

In the first part of the second classroom session, two weeks later, learners were engaged in reading comprehension and discussion of the next episode of the mystery text (phase 6). In this phase, the text served a double purpose: first, as in the first session, it was intended to make learners focus on meaning during the subsequent practice phase. Secondly, the text now included one grammatical sentence for each verb used in the practice material for V2O. In line with recommendations from DeKeyser (1998, p. 59) and Lantolf & Thorne (2006, p. 302), this was meant to help learners further automatize their knowledge of V2O while reading and discussing the text. Attempts were made to also include items for the QNT constructions, but these failed, as the text started to feel too artificial. The researcher did try to elicit learners' use of the constructions for QNT (as for DOC) during the meaning-oriented discussions in class. In order to keep learners' attention focused on meaning as much as possible during the reading and discussion phases, no input enhancements were made on the text (such as highlighting of the linguistic phenomena, glosses with additional explanation, etc.), and the researcher tried to avoid pronouncing the constructions with particular stress during the discussion. Following the discussion, the learners again practised the constructions for QNT and V2O, but now with new practice items, which were based on the final part of the story. At the end of this class session, learners were again asked to practice 20 minutes at home before the next session.

In the third classroom session, 2 weeks later, learners in the treatment groups read and discussed the conclusion of the mystery text, which again included examples for V2O, and took the first post-test, which was identical to the pre-test except for the order of the items (see section 6.3.4). They also completed a questionnaire which contained a scale on intrinsic motivation based on the Intrinsic Motivation Inventory (IMI) (Plant & Ryan, 1985) and the Player Experience of Needs Satisfaction model (Ryan et al., 2006). The control group only took the language post-test.

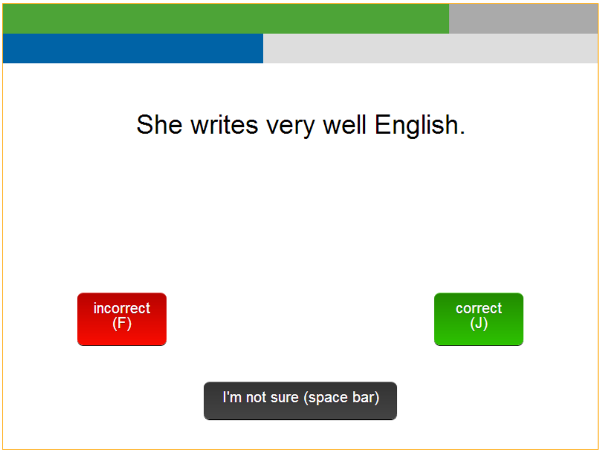
Between the first and delayed post-tests, which concluded the experiment one month later, 69 learners were selected from the treatment groups to participate in an individual session with the researcher to complete a spoken language test known in the literature as the oral elicited imitation test (OEIT) (Erlam, 2009). Care was taken to ensure that there was a balanced number of participants from both treatment groups, as well as a spread in the amount of time that the participants had spent on practice, so that the effects of the different types of CF as well as of time on task could be investigated. The next section provides more detail on the content and procedure of this test.

6.3.4 Language tests and questionnaire data: format, scoring, and reliability

Four tests were used to measure development in terms of different aspects of L2 knowledge, and a questionnaire was used to gauge the intrinsic motivation of participants in the treatment groups. Two tests constituted the repeated measures of the study and were offered three times (see Figure VI-1), namely a timed grammaticality judgment test (TGJT) and a written discourse completion test (WDCT). These tests were completed on the computer. The metalinguistic knowledge test (MKT) was paper-based. The oral elicited imitation test (OEIT) was audio-recorded.

6.3.4.1 Timed grammaticality judgment test (TGJT)

Because of the time limit that TGJTs impose on learners' responses, they are considered to predispose learners towards using implicit L2 knowledge (Gutiérrez, 2013; Loewen, 2009). In this study, they may offer a measure of how well the participants of this study had automatized L2 knowledge through practice. As in the practice tasks, the stimuli were written, and the TGJT used a layout very similar to the practice tasks described in section 0, but without CF and without the graphics related to the mystery context (see Figure VI-5). Another difference was that participants could skip items by pressing the space bar, in case they were unable to make a judgment; this option was provided in order to control for guessing. Participants used the J key on the computer keyboard to indicate grammatical correctness; if they thought the sentence was ungrammatical, they had to use the F key. This is in contrast to Loewen's (2009) approach, but in line with psycholinguistic studies that are based on the assumption that people are likely to confirm well-formedness of stimuli with their dominant hand (e.g. Keuleers, Brysbaert, & New, 2010). Participants indicated whether they were dominantly right- or left-handed on the consent form, allowing us to control for this potential effect. A blue progress bar visualized the time remaining to judge the current item; the system automatically advanced to the next item if the participant was unable to respond within time. A green progress bar indicated progress through the test.



She writes very well English.

Figure VI-5: layout for the timed grammaticality judgment test

The TGJT was piloted in December 2013 in order to select the best items for the learning study. The participants of this pilot study were 96 students in the first year of the applied economic sciences programme at a Flemish university, who are typically between 1 and 2 years older than the participants in the actual learning study. The pilot test consisted of 112 items, covering QNT (48 items), V2O (48 items), and 16 distractor items drawn from the Marsden project (R. Ellis et al., 2009). For QNT and V2O, half of the items were drawn from the practice materials (see section 0); the other half comprised novel items, aimed at measuring generalization transfer in the learning study. The items were offered in 8 sets of 14 items (six items for QNT, six for V2O, and two distractor items). The sets were counterbalanced for the participants, which allowed to control for any effects of learning or fatigue. The order of items within the sets was randomized, but equal for all participants. The time limit for each item was between 3 and 6.98 seconds, based on sentence length (number of characters) and earlier research from Loewen (2009)—whose time limit was based on response times from native speakers of English—and Gutiérrez (2013). Participants were allowed a 15-second break between each item set. The responses from six participants were missing, due to data collection problems. The data from the remaining 90 participants were suitable for analysis.

Items that received a response within the time limit were matched with a key, and received a score of 0 or 1, depending on the outcome of the match. Items that did not receive a response within the time limit, as well as skipped items, were scored as 0. The assumption was that if participants did not respond, or did not do so within the time limit, they did not have sufficiently well-developed grammar knowledge pertaining to this item. This is in line with Gutiérrez (2013)—it is not very clear what Loewen (2009) did with missing responses on the TGJT—but may be somewhat problematic for two reasons. First, the test was fast-paced, and secondly, it had proven impossible, particularly for V2O, to control the lexical complexity of the items by using only relatively frequently used verbs (e.g. only verbs from the 2000 most frequently used words in English; see e.g. Waring & Nation, 1997). Further, responses that had a response time shorter than 1 second were recoded as missing values. Observations during the test learnt that participants were sometimes a fraction of a second too late with their responses, in which case the system registered a response for a previous item. So, by deleting all responses with a response time under 1 second, these responses were deleted. Table VI-2 shows the distribution and scoring approach of the response types for the pilot TGJT.

Table VI-2: distribution of response types and scoring for pilot TGJT

Response type	Correct (J)	Incorrect (F)	Skipped	Out of time limit	Removed
Proportion (in percentages)	55.74	32.92	6.77	2.98	1.60
Scoring	matched with key (0 or 1)	matched with key (0 or 1)	0	0	missing value

Visual inspection of the scores of the item sets over time revealed no clear effect of learning or fatigue, so we retained the entire data set for the selection of items. For the construction of the final test, items were retained on the basis of the following criteria: items had to have an item-total correlation as high as possible, there had to be a spread of difficulty, and there had be an equal distribution of items for QNT and V2O on the one hand, as well as an equal distribution of practice items and non-practice items (see Table VI-3). A reliability analysis showed that the final selection of the items for QNT and V2O had an internal reliability coefficient of .73 (Chronbach's α), which is acceptable in research on learning.

Table VI-3: number of items in TGJT according to type, linguistic problem, and difficulty (stars indicate frequency bands of difficulty)

	QNT	DOC
Practice items	* 2 ** 8 *** 2	* 3 ** 6 *** 3
Non-practice items	* 1 ** 8 *** 3	* 3 ** 5 *** 4
All items	24 (17 ungrammatical, 7 grammatical)	24 (20 ungrammatical, 4 grammatical)

The TGJT used in the learning study consisted of 6 sets of 9 items, each comprising 4 items for QNT, 4 for V2O, and 1 distractor item (see Appendix 6). Administration of the TGJT in the learning study was done in the same way as in the pilot study.

Reliability analyses pointed out that the internal consistency of the TGJT (without the distractor items) was .44 in the pre-test (Cronbach's α), .89 in the first post-test, and .88 in the delayed post-test. Reliability of the pre-test is much lower than the .70 acceptability threshold, and markedly lower than the reliability of the post-tests. This can be attributed to four possible causes. First, after the pre-test, a number of learners said that they experienced difficulties with the time pressure, and with operating the keys, resulting in key-presses that were erroneous, too late, or perhaps more skips. This may have introduced irrelevant variability in the test. At the time of the post-tests, participants in the treatment groups were likely to be more familiar with the task type, hence reducing the number of such responses. The smaller number of such responses (skipped items, responses that were out of time, and removed responses) on both post-tests lends support to this interpretation (see Table VI-4). Secondly, the participants in the learning study were one to two years younger than the participants in the pilot study, and were by and large less academically oriented. Thus, they may have possessed less relevant linguistic knowledge, or perhaps knowledge of more informal registers of the L2. Third, since the TGJT was aimed at measuring implicit L2 knowledge, the learners were not told what the tests were about. After the treatment, however, participants in the treatment groups obviously knew that the tests targeted QNT and V2O, and may have drawn on explicit knowledge rather than implicit knowledge, even though the test items were offered in a mixed order and included distractors. Thus, it is conceivable that on the post-tests, learners focused only on the grammaticality of the items in terms of the two linguistic problems that were practised, whereas in the pre-test, they might have paid attention to various other aspects, introducing more variability, and decreasing the reliability of the test. A fourth possible cause for the lower reliability of the pre-test is that the participants in the learning study took the test under slightly different conditions than those in the pilot study, namely in the atmosphere that characterizes secondary education classrooms (arguably more rowdy than a university context), and with less advanced computer hardware. However, since the reliability of the pilot TGJT was .73, and reliability of the post-TGJTs was more than acceptable, we did not consider the low reliability of the first TGJT highly problematic.

Table VI-4: distribution of response types on the TGJT used in the learning study, per test time (expressed in percentages)

	Correct (J)	**Incorrect (F)**	**Skipped**	**Out of time limit**	**Removed**
Pre-test	54.32	34.52	7.55	2.81	0.79
Post-test 1	48.27	44.54	4.72	1.83	0.64
Post-test 2	49.58	42.97	5.29	1.75	0.42

6.3.4.2 Written discourse completion test (WDCT)

The purpose of the WDCT was to elicit productive use of knowledge of the target constructions, moving beyond decontextualized fill-in-the-blanks tasks and translation tasks. Still, the task was written, and not time-pressured, so learners were able to draw on their explicit knowledge. Hence, the WDCT could constitute a measure of explicit knowledge, implicit knowledge (particularly in the event that learners were not aware of the target structures being tested), or both.

Learners got a description of a situation, and were required to complete a sentence, taking into account a number of constraints, such as the requirement to use a particular noun phrase. For QNT, only situations were given that ask for formal English use, since the distinctions *less* vs. *least* and *least* vs. *fewest* are made most strictly in formal English, as the learners were instructed. A sample item for QNT is:

> *You work as a pharmacist for Johnson & Johnson, and have just improved the recipe of a painkiller. You are in your boss's office, and want to convince him to produce your improved version of the painkiller. You think that you can convince him by saying that the quantity of <u>ingredients</u> needed to make the painkiller is much smaller now. You say:*
>
> *"My new recipe is better than the old one, because it uses __________".*
>
> ! You must use the underlined words.

Due to time constraints, piloting the test on a large number of learners was impossible. Yet, 16 items were given to the first class that participated in the experiment, of which eight were kept for the remaining classes on the basis of inspections of the item-total correlations. Four items included verbs (for V2O) or nouns (for QNT) used in the practice materials, the other four items used other vocabulary items in order to measure generalized knowledge (see Appendix 7). As with the TGJT, the items were clustered in sets, which were offered in a counterbalanced order.

The researcher coded the learners' written responses on the WDCT manually. Items without responses were scored as missing values. Since the purpose of the test was to elicit functional use of grammar, and the items described communicative situations, spelling mistakes were ignored. A fully grammatical response was awarded 3 points. Grammatical, but marked constructions were given 2 points. Avoidance (i.e. non-use) of the target constructions received no points. For example, for V2O, acceptable constructions with a marked word order (e.g. *Have you reported to them the car accident?*) or a non-target preposition (e.g. *Have you reported the car accident at them?*) in the PC were coded as grammatical, but marked (2 out of 3 points). Similarly, for QNT, use of a grammatically correct quantifier in a sentence that was not entirely grammatical was given 2 points (e.g. *Team B wins, because they made the fewer mistakes than team A*). Use of a non-target but grammatically correct quantifier was considered avoidance (e.g. *My new recipe is better than the old one, because it uses many ingredients* in a context that required the use of *fewer*). The assumption was that in such cases, learners were avoiding the choice between *less* or *fewer* because they were not sure about the distinction. Full avoidance of the target construction received no points (e.g. *Have you reported the car accident yet?*). This is somewhat problematic, certainly for the pre-tests (see Table VI-5) and for the control group (see Table IV-6), since non-use of the target construction does not necessarily imply that the learner is unaware that the construction **Have you reported them the car accident yet* is ungrammatical. Moreover, the tests were not piloted on native speakers, which compromises their validity. However, the learners were instructed to produce full constructions—in this case, they were told to use the pronoun *them*—and it was assumed that by coding such instances as non-use, the effects of practice would be evident.

Table VI-5: avoidance on WDCT over time (expressed in percentages)

	pre-test	post-test 1	post-test 2
V2O	9.89	5.31	4.68
QNT	11.95	4.40	3.93

Table VI-6: avoidance on WDCT per experimental condition (expressed in percentages)

	control	practice with ML CF	practice with KR CF
V2O	9.25	5.34	5.56
QNT	7.23	7.60	5.65

Following reliability analysis, one of the final eight items was removed because it had a low item-total correlation. Closer inspection showed that this item on V2O allowed both the DOC and the PC, even though the instructions told learners to only use a preposition when they really thought it was necessary. Reliability of the test containing the 7 remaining items was .52 at the pre-test (Cronbach's α), .69 at the first post-test, and .71 at the final post-test. The low reliability on the pre-test can be explained in similar ways as for the TGJT. Reliability of the post-tests is lower than the post-TGJTs, and may be due to the fact that the test contained fewer items, and was not fully piloted in advance.

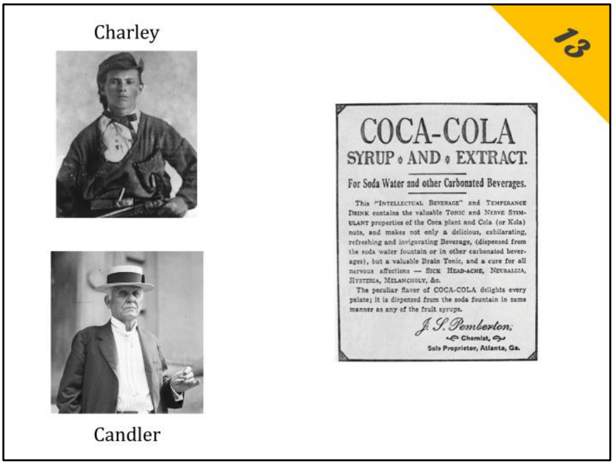
6.3.4.3 Metalinguistic knowledge test (MKT)

The MKT relied to a large extent on metalinguistic terminology and stimulated conscious reflection on grammar rules. Hence, it was most likely to be a pure measure of explicit knowledge. It comprised four items, of which two concerned QNT, and the other two pertained to V2O. Each item consisted of an ungrammatical sentence, and four possible explanations for the ungrammaticality of the sentence. Learners were instructed to choose the best possible explanation. They also had the possibility to skip the question, which was scored as 0. Reliability of the test was .22 (Cronbach's α), which is likely due to the fact that it was not piloted in advance, and to the small number of items.

6.3.4.4 Oral elicited imitation test (OEIT)

The OEIT was intended to measure far transfer to productive skills, and—although still a very constrained task—comes closest to measuring "unplanned language use" (R. Ellis et al., 2006, p. 351). It largely followed the format used in the Marsden project (Erlam, 2009), and involves a dual focus on meaning and on form. Learners listen to a set of statements, some of which are grammatically incorrect. Their task is to repeat the stimuli in a grammatically correct way, after judging the truth value of the statements. This design feature aims for the task to be reconstructive, rather than purely imitative, and hence to draw on learners' implicit knowledge. In line with the format developed in the Marsden project (Erlam, 2009), test takers were not told explicitly that some of the statements uttered by the researcher were grammatically incorrect, but only that they had to repeat the statements in correct English. The participants took the OEIT in individual sessions with the researcher or one of the participating teachers.

In comparison with the original format, the OEIT used in the current study might have had a stronger focus on meaning, as it was conducted in the form of a role-play. The test-taker took the role of the detective, formulating his ideas about the mystery case out loud in his office, and the participant's task was to play the detective's parrot, repeating its owner's speech. The operationalization of the OEIT as a role-play against the backdrop of the detective story was meant to focus learners' attention on meaning during the language test. Also, the OEIT described in Erlam (2009) was introduced to participants as a 'beliefs questionnaire' on a wide range of topics. In the current study, the statements were all related to the mystery text in class, which made them thematically more coherent. The participants thus had to indicate (on paper) whether they thought the statements were true according to the story. Another difference with the latter format was that participants were supported visually by means of slides presented on a computer. These slides comprised pictures for the vocabulary used in the stimuli. For the items on V2O, the pictures were not ordered horizontally so as not to prime the participants into using either the DOC or PC (see Figure VI-6).



COCA-COLA
SYRUP AND EXTRACT.

For Soda Water and other Carbonated Beverages.

This "INTELLECTUAL BEVERAGE" and TEMPERANCE DRINK contains the valuable TONIC and NERVE STIMULANT properties of the Coca plant and Cola (or Kola) nuts, and makes not only a delicious, exhilarating, refreshing and invigorating Beverage, (dispensed from the soda water fountain or in other carbonated beverages), but a valuable Brain Tonic, and a cure for all nervous affections — SICK HEAD-ACHE, NEURALGIA, HYSTERIA, MELANCHOLY, &c.

The peculiar flavor of COCA-COLA delights every palate; it is dispensed from the soda fountain in same manner as any of the fruit syrups.

J. S. Pemberton;
Chemist,
Sole Proprietor, Atlanta, Ga.

Figure VI-6: visual support for the oral stimulus

**Charley revealed Candler the secret recipe of Coca-Cola.*

Unfortunately, using the OEIT as a measure of implicit L2 knowledge proved untenable. As a possible effect of the treatment, some participants quickly realized that the task was intended as a language test, once the researcher had said that they were required to repeat the statements in correct English. If not at that point in the test, many learners found out later, as is evident in the following transcript:

> Researcher: *Charley has less shares than Frank Robinson.*
> Learner: *Charley has <err> do I have to repeat or correct the sentence?*
> Researcher: *You have to say it in correct English.*
> Learner: *Ah OK! Charley has fewer shares than Frank Robinson.*

Hence, whenever there was time, the researcher held a short debriefing session with the learner, in which he gauged the learner's perceptions of the task, and whether the learner had become aware that they had been tested on grammar.

Cases in which it had become evident, either during the test or in the debriefing sessions, that the participant had attended to grammatical form were marked as 'form awareness'. Cases in which the participants said they had not realized the test was about grammar were coded as 'no form awareness'. Cases in which there was no clear evidence of form awareness, or no data at all, were coded as missing data. Admittedly, this is a relatively crude measure. Participants may have become aware of the purpose of the test in the instruction phase, or later, after completing a number of test items. Moreover, determining whether participants had realized what the test was intended to measure relied to some extent on interpretative observation and self-report. Table VI-7 shows the distribution of form awareness with respect to the experimental condition.

Table VI-7: distribution of form awareness in OEIT
with regard to experimental condition

	practice with ML CF	practice with KR CF
Form awareness	21	20
No form awareness	2	6
Unclear / no data	11	9
TOTAL	34	35

The learners' spoken responses were audio-recorded, transcribed, and coded. In cases where learners had self-corrected, the final attempt was retained, because it was evident that the majority of learners had monitored their speech—this was confirmed in the debriefing sessions—which implies that they had probably been drawing mainly on their explicit knowledge. Retaining the first attempt for analysis, as was done for instance in R. Ellis *et al.*'s (2006) study, may have resulted in an even more varied mix of implicit and explicit L2 knowledge, compromising the construct validity of the test. The coding approach of the constructions was similar to that of the WDCT. 3 points could be gathered for each response. A fully grammatical construction received 3 points initially, or 2 points if it was marked. On top of this, we subtracted 1 point for responses that showed signs of avoidance. For QNT, this meant that learners lost 1 point if they did not use a countable noun in the plural. For instance, if the target construction was *too many coca leaves*, and the learner uttered *too much coca cola,* he missed a point. Similarly, for V2O, learners lost 1 point if they did not use the same verb class. For instance, if they changed the verb *promise* to *give*, they missed a point because this was seen as a form of avoidance, even if changing the verb class may have been the result of their explicit knowledge and practice. Full avoidance of the target construction (e.g. use of *more* rather than *much* or *many*) received 0 points. The following instances were scored as missing values: skipped items, items where the teacher intervened too explicitly (e.g. by repeating the statement after the learner had judged the meaning of the sentence), instances where the learners had obvious lexical difficulties or were unable to repeat the sentence, or sentences that sounded meaningless. The proportion of missing values with respect to the total number of responses was 8.05 % ; the proportion of avoidance was 3.22 %.

Construction of the items for the OEIT was partly inspired by the reliability analysis of the items in the pilot TGJT. The test consisted of four practice items, and six items each for QNT and V2O (see Appendix 8). The items for QNT and V2O were mixed and clustered in two sets, which were offered in a counterbalanced order. As part of the reliability analysis, two items were removed for V2O, which allowed both the PC and the DOC and had very low item-total correlations (between -.088 and .076). Another item on QNT had a very low item-total correlation (.088), which can be explained by the fact that its head noun *headache* can be both countable and uncountable. This item was also removed from the test. Following removal of these items, internal consistency of the 9 remaining items for QNT and V2O was .81 (Cronbach's α).

6.3.4.5 Summary of instruments

Table 7 presents a summary of the instruments, with for each instrument, the construct it is intended to measure, based on assumptions formulated in the Marsden project (R. Ellis, 2009b), and the variables. For the language tests, in addition, the type of transfer from practice is listed, and the types of items. Non-practice items were included to measure generalization (i.e. system learning). Analysis of the questionnaire results (IMI) is outside the scope of this chapter.

Table VI-8: summary of instruments

instrument	targeted construct(s)	variables	transfer of practice	types of items
TGJT	implicit L2 knowledge	- accuracy rates - response times	near transfer	- practice items - non-practice items
WDCT	explicit L2 knowledge	- accuracy rates - response times	far transfer to written production	- practice items - non-practice items
MKT	explicit L2 knowledge	accuracy rates	—	—
OEIT	implicit L2 knowledge	accuracy rates	far transfer to spoken production	practice items
IMI	instrinsic motivation	7-point Likert-scale responses	—	—

6.4 Results

6.4.1 Data preparation, and correlations between the language tests

Following the reliability analyses (see section 6.3.4), all accuracy scores were recomputed on a scale from 0 to 1, and test averages were computed for each participant for both accuracy rates and response time data. As for the data set comprising the TGJT and the WDCT, participants that did not have results for the three test times were discarded, in order to obtain a fully balanced data set. This resulted in 52 participants for the control group, 56 participants for the practice group that received ML CF, and 59 participants for the practice group supported by KR CF.

Table VI-9 shows the correlation matrix for the accuracy rates on the language tests for the treatment groups only. The matrix shows stronger significant correlations between the TGJT and the WDCT on the post-tests (large-sized) than on the pre-tests (medium-sized). Moreover, there are significant correlations between the pre-tests and both post-tests of the WDCT, but not so for the TGJT. This trend is not evident in the control group (see Table

VI-10), where the TGJT and WDCT do not correlate significantly on the three test times, and where the repeated measures of each test type correlate, except for the delayed post-test of the TGJT, which does not correlate significantly with its pre-test. Furthermore, for the treatment groups only, there were significant correlations between the accuracy rates and response times on both post-tests (Pearson's *r* -.49 and -.47, respectively), whereas no correlation was found on the pre-test. This all suggests that in the treatment groups, the TGJT may have measured a different construct on the pre-test than on both post-tests. Another noteworthy finding is that the OEIT correlates significantly with the post-tests of the TGJT and WDCT—in particular with the first post-test of the WDCT—but not with the pre-tests. This implies that the OEIT may measure a construct similar to the construct measured by the post-tests of TGJT and WDCT.

Table VI-9: Pearson's correlation coefficients for the language tests (treatment groups), adjusted for multiple comparisons using Holm's method (** *p* < .01, * *p* < .05)

	1.	**2.**	**3.**	**4.**	**5.**	**6.**	**7.**	**8.**
1. TGJT pre	—	.36*	.07	.22	.18	.29	.24	.11
2. WDCT pre		—	.25	.25	.41**	.26	.25	.44**
3. MKT			—	.21	.30*	.26	.27	.37**
4. TGJT post 1				—	.66**	.39*	.81**	.57**
5. WDCT post 1					—	.51**	.60**	.73**
6. OEIT						—	.40*	.39*
7. TJGT post 2							—	.68**
8. WDCT post 2								—

Table VI-10: Pearson's correlation coefficients for the language tests (control group), adjusted for multiple comparisons using Holm's method (** *p* < .01, * *p* < .05)

	1.	**2.**	**3.**	**4.**	**5.**	**6.**
1. TGJT pre	—	.05	.49**	.12	.27	.18
2. WDCT pre		—	.12	.55**	.08	.53**
3. TGJT post 1			—	.21	.69**	.27
4. WDCT post 1				—	.32	.62**
5. TJGT post 2					—	.31
6. WDCT post 2						—

6.4.2 Results for TGJT and WDCT

6.4.2.1 Effects of practice on mean accuracy rates and mean response times

Table VI-11 contains the summary statistics for the TGJT and WDCT. Four one-way ANOVAs showed that there were no significant differences between the groups on the pre-tests in terms of mean accuracy rates and response times.

Table VI-11: summary statistics for TGJT and WDCT

	group	**pre-test**	**post-test 1 mean accuracy**	**post-test 2 mean accuracy**
TGJT, mean accuracy	control (*N* = 52)	*M* = .41 *SD* = .08	*M* = .39 *SD* = .09	*M* = .39 *SD* = .10
	practice ML CF (*N* = 56)	*M* = .39 *SD* = .09	*M* = .68 *SD* = .17	*M* = .65 *SD* = .19
	practice KR CF (*N* = 59)	*M* = .39 *SD* = .08	*M* = .60 *SD* = .15	*M* = .59 *SD* = .16
TGJT, mean response time	control (*N* = 52)	*M* = 3.48 *SD* = .28	*M* = 3.35 *SD* = .33	*M* = 3.23 *SD* = .35
	practice ML CF (*N* = 56)	*M* = 3.47 *SD* = .31	*M* = 2.93 *SD* = .49	*M* = 3.02 *SD* = .46
	practice KR CF (*N* = 59)	*M* = 3.36 *SD* = .30	*M* = 2.91 *SD* = .42	*M* = 2.93 *SD* = .37
WDCT, mean accuracy	control (*N* = 52)	*M* = .30 *SD* = .20	*M* = .31 *SD* = .18	*M* = .31 *SD* = .21
	practice ML CF (*N* = 56)	*M* = .35 *SD* = .23	*M* = .59 *SD* = .31	*M* = .58 *SD* = .32
	practice KR CF (*N* = 59)	*M* = .29 *SD* = .23	*M* = .46 *SD* = .28	*M* = .44 *SD* = .28
WDCT, mean response time	control (*N* = 52)	*M* = 53.42 *SD* = 13.58	*M* = 32.26 *SD* = 8.14	*M* = 25.37 *SD* = 6.29
	practice ML CF (*N* = 56)	*M* = 55.21 *SD* = 14.97	*M* = 37.01 *SD* = 13.03	*M* = 30.08 *SD* = 8.87
	practice KR CF (*N* = 59)	*M* = 50.7 *SD* = 11.26	*M* = 35.85 *SD* = 12.66	*M* = 28.27 *SD* = 8.09

In order to test hypotheses 1 and 2, repeated measures ANOVAs were run on both the TGJT and the WDCT data (one for mean accuracy, one for mean response time on each test). In the analysis of the WDCT, which was not time-pressured, mean response times were normalized using logarithmic transformation in order to eliminate the effect of extreme high values and, as such, to create a more normal distribution required by the statistical procedures.

The repeated measures ANOVAs for the TGJT revealed main effects of time and condition that were significant at the .0001 level, and interaction effects between test time and condition for mean accuracy (F(4, 328) = 44.573, $p <$ .0001) and mean response time (F(4, 328) = 11.217, $p <$ .0001) (see Figure VI-7). For the WDCT, there were main effects of time and condition significant at the .0001 level, and there was a significant interaction effect between test time and condition on mean accuracy (F(4, 328) = 7.6876, $p <$ .0001). For the response times on the WDCT, there was only a main effect of test time (F(2, 328) = 421.77, $p <$ .0001). The main effect of condition was not significant (F(2, 164) = 2.37, p = .097), nor was the interaction between test time and condition (F(4, 328) = 2.35, p = .0544) (see Figure VI-8).

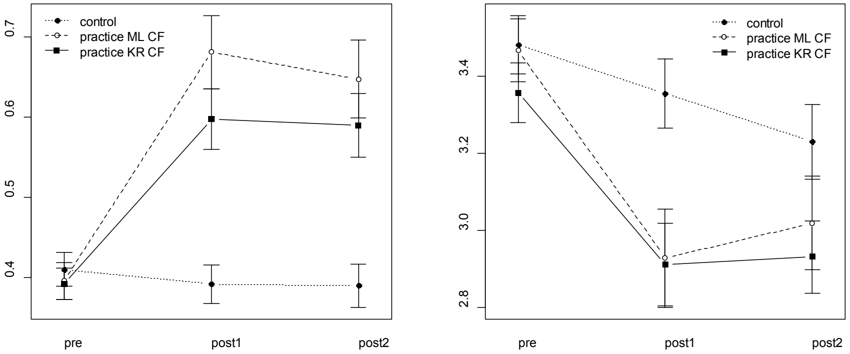


Figure VI-7: plots of mean accuracy (left) and mean response time (right) on the TGJT, by test time and experimental condition

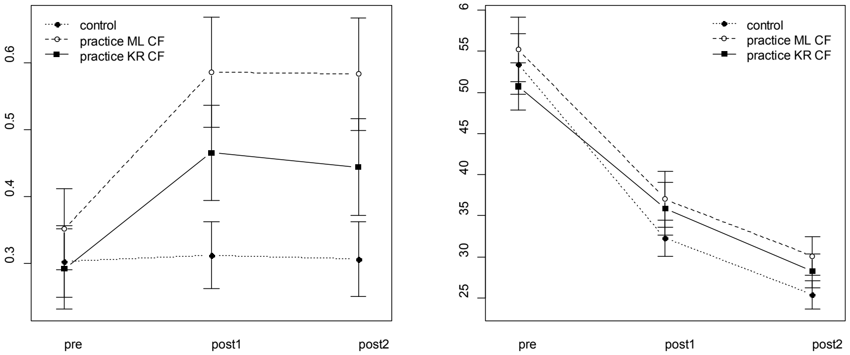


Figure VI-8: plots of mean accuracy (left) and mean response time (right) on the WDCT, by test time and experimental condition

Visual inspection of the residuals showed no abnormalities, so post-hoc analyses were subsequently performed to find out which conditions differed from each other at the different test times. The post-hoc contrasts for the accuracy rates on the TGJT showed that both treatment groups significantly outperformed the control group on both post-tests (*p* < .001), and that the ML CF group did significantly better than the KR CF group on the immediate post-test (*p* < .01), but not on the delayed post-test (*p* = .069). As for the response times, the treatment groups were significantly faster than the control group on the immediate post-test (*p* < .001) and on the delayed post-test (*p* < .001 for the KR CF group; *p* < .05 for the ML CF group), but there was no significant difference between the treatment groups on the immediate (*p* = 1) or delayed post-test (*p* = .636).

Post-hoc contrasts for the WDCT showed that the control group was significantly less accurate on both post-tests than the ML CF group (significant at *p* < .001) and the KR CF group (significant at *p* < .05), and that the ML CF group outperformed the KR CF group on both post-tests (significant at *p* < .05). No significant differences were found between the groups in terms of response times.

Table VI-12 presents a summary of the statistically significant group differences on the post-tests for the TGJT and the WDCT.

Table VI-12: summary of significant effects

	mean accuracy rates	**mean response times**
TGJT post 1	practice ML CF > control (***) practice KR CF > control (***) practice ML CF > practice KR CF (**)	practice ML CF > control (***) practice KR CF > control (***)
TGJT post 2	practice ML CF > control (***) practice KR CF > control (***)	practice ML CF > control (*) practice KR CF > control (***)
WDCT post 1	practice ML CF > control (***) practice KR CF > control (*) practice ML CF > practice KR CF (*)	
WDCT post 2	practice ML CF > control (***) practice KR CF > control (*) practice ML CF > practice KR CF (*)	

6.4.2.2 Effects of practice with respect to item type

Following the repeated measures ANOVAs on the mean accuracy rates and response times for the three experimental conditions, four multi-level analyses were performed on the singular responses of the participants in the two treatment groups. The first of these analyses were carried out in order to account for the differences in item type (practice items vs. non-practice items). Three independent factors were included, namely test time, treatment type, and item type, and participant was included as a random factor.

As for the effect of item type on the accuracy scores of the TGJT, a significant main effect was found for item type ($F(1, 16338) = 89.7214$, $p < .0001$), but not for the interaction between item type and test time ($F(2, 16338) = 1.1264$, $p = .3242$). For the response times on the TGJT, there was a significant main effect of item type ($F(1, 16164) = 513.636$, $p < .0001$), and a significant interaction effect between item type and test type ($F(1, 16164) = 48.172$, $p < .0001$) (see Figure VI-9). Post-hoc comparisons for both accuracy rates and response times on the TGJT showed that the effect of item type was significant at the three test times ($p < .001$).

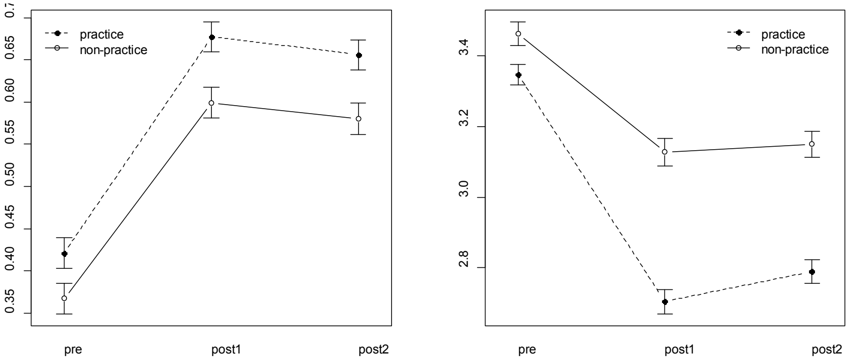


Figure VI-9: plots of mean accuracy (left) and mean response time (right) on the TGJT, by test time and item type (only for the treatment groups)

On the WDCT, item type did not have a main effect on the accuracy rates ($F(1, 2258) = 1.5078$, $p = .2196$), nor was there a significant interaction effect ($F(2, 2258) = 2.1165$, $p = .1207$). For the (log-transformed) response times, item type had only a main effect ($F(1, 2290) = 32.106$, $p < .0001$) (see Figure VI-10). The post-hoc comparisons revealed significant differences between the response times of the item types on the three test times ($p < .01$).

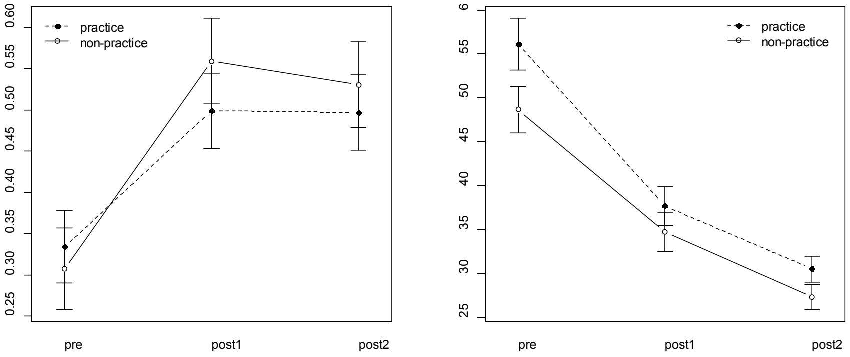


Figure VI-10: plots of mean accuracy (left) and mean response time (right) on the WDCT, by test time and item type (only for the treatment groups)

There were no interaction effects between item type and treatment type on either of the language tests.

6.4.2.3 *Effects of practice with respect to linguistic problem*

Similar multi-level analyses were run in order to investigate whether the effects of practice depended on the complexity of the linguistic problem. On the accuracy scores of the TGJT, there was a significant main effect of linguistic problem ($F(1, 16338) = 67.0431$, $p < .0001$), and an interaction effect with test time ($F(2, 16338) = 62.2514$, $p < .0001$) (see Figure VI-11). These differences were significant at the three test times, as shown by post-hoc comparisons ($p < .001$).

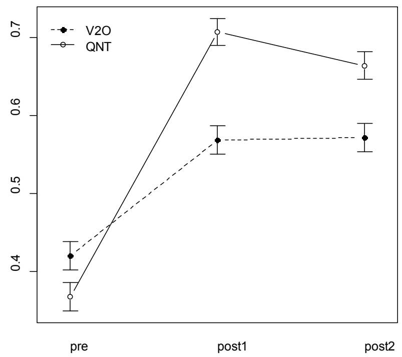


Figure VI-11: plot of mean accuracy on the TGJT,
by test time and problem (only for the treatment groups)

On the response time data, a significant three-way interaction effect was observed between linguistic problem, treatment type, and test time (*F*(2, 16164) = 4.464, *p* < .05) (see Figure VI-12). The plots show that evolution of the response times for V2O was slightly different between the groups.

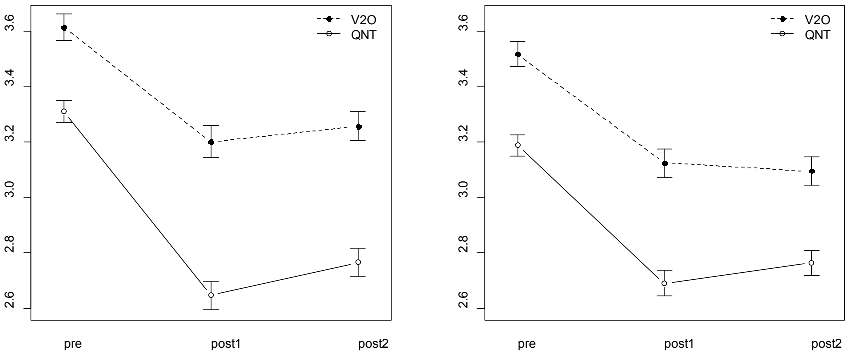


Figure VI-12: plots of response times on the TGJT for ML CF group (left) and KR CF group (right), by test time and problem

A similar trend can be observed for the WDCT data. As for the for accuracy rates, there was no main effect for linguistic problem (*F*(2, 2258) = 0.4419, *p* = .5063), but a significant interaction effect between test time and problem (*F*(2, 2258) = 16.4846, *p* < .0001) (see Figure VI-13). Post-hoc comparisons showed that the differences between the accuracy rates for the linguistic problems were only evident on the pre-test (*p* < .001) and the first post-test (*p* < .05). For the response times, a significant main effect of problem was present (*F*(1,

2290) = 6.736, p < .01), as well as a significant three-way interaction between problem, test time, and treatment type (F(2, 2290) = 4.804, p < .01) (see Figure VI-14).

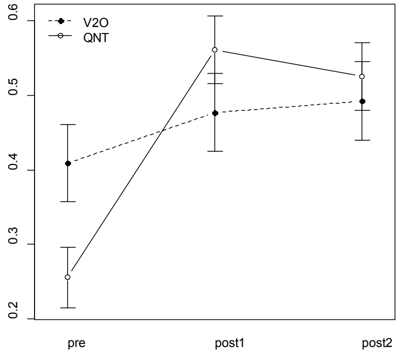


Figure VI-13: plot of mean accuracy on the WDCT,
by test time and problem (only for the treatment groups)

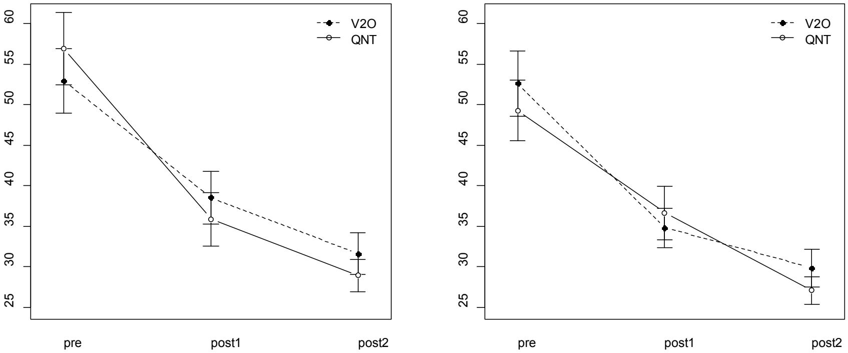


Figure VI-14: plots of mean response times on the WDCT for ML CF group (left) and KR CF group (right), by test time and problem

6.4.3 Results for OEIT

To investigate whether the participants had attended to meaning on the OEIT, we examined the mean scores of their responses on the truth value of the statements, which they formulated on the basis of their reading of the mystery story. This was done separately for the participants that had been aware that the task was a grammar test and for the unaware participants (see section

6.3.4.4), because the former participants may have focused more on form, perhaps resulting in lower scores on the meaning-oriented part of the test. The summary statistics show a slightly higher mean for the aware group and, within the aware group, a median of .86, which is higher than the mean of .74. This suggests that the participants had largely attended to meaning, certainly in the in the aware group. An independent-samples t-test of the means of the different groups showed that form awareness did not affect the participants' scores on the meaning judgments ($t(10.002) = .9214$, $p = .38$).

Table VI-13: mean accuracy for meaning judgments on the OEIT, by form awareness

	M	***Mdn***	***SD***
no form awareness	.66	.57	.23
form awareness	.74	.86	.23

Table VI-14 shows the mean accuracy rates on the OEIT by form awareness and treatment type (practice with ML CF vs. practice with KR CF). Separate regression analyses were run for the aware and unaware participants, since the former participants are likely to have drawn mostly on explicit knowledge, while the latter may either have used implicit knowledge, or simply have repeated the stimuli sentences verbatim.

Table VI-14: mean accuracy on the OEIT, by form awareness and treatment type

	M	***SD***
no form awareness, ML CF (*N* = 2)	.11	0
no form awareness, KR CF (*N* = 6)	.19	.12
form awareness, ML CF (*N* = 21)	.58	.22
form awareness, KR CF (*N* = 20)	.62	.20

The mean accuracy rates of the OEIT were regressed, for each group separately, onto two main predictors and three control variables. The main predictors were treatment type, and the participants' time on task, expressed as the number of minutes spent judging the sentences during practice (range between 2.9 and 85.7 minutes; $M = 26.31$; $Mdn = 22.00$). The latter variable was log-transformed in the regression analyses. The control variables included measures of participants' prior implicit L2 knowledge, as measured by the mean accuracy rates on the first TGJT, and of their explicit knowledge, i.e. their results on the first WDCT and MKT. The continuous variables of these independent predictors had no strong inter-correlations (see Table VI-15), so they could be jointly used in the regression analyses.

Table VI-15: Pearson's correlation coefficients for the predictors used in the regression analyses for OEIT (comprising the dataset of all participants that participated in the OEIT; *N* = 69), adjusted for multiple comparisons using Holm's method (** *p* < .01, * *p* < .05)

	1.	**2.**	**3.**	**4.**
1. minutes of practice	—	.00	.11	.07
2. TGJT (pre-test) mean accuracy		—	.34*	.09
3. WDCT (pre-test) mean accuracy			—	.28
4. MKT mean accuracy				—

Following a first regression analysis for the aware participants *(N* = 41), two outliers were removed from the dataset. In the debriefing sessions, these two participants had reported that they had noticed grammatical mistakes in the stimuli sentences, but had not corrected them because they thought the task was to simply repeat the sentences. As a result, we considered these two data points irrelevant for the analysis. After removal of these outliers, the distribution of the residuals was more normal. The 5 predictors jointly explained 32 percent of the variance in the mean accuracy rates (adjusted R^2 = .32, *F*(5, 32) = 4.523, *p* < .01). The mean accuracy rates were positively affected by the time spent on practice (β = .084, *p* < .05) (see Figure VI-15) and by performance on the pre-test of the WDCT (β = .322, *p* < .01). The type of CF given in practice did not have a significant effect (β = .034, *p* = .531), nor did performance on the first TGJT (β = .083, *p* = .763) or MKT (β = .166, *p* = .097).

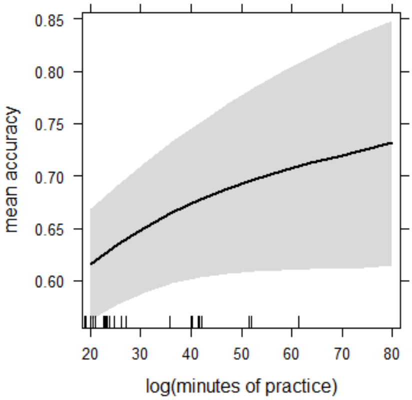


Figure VI-15: plot for relation between time spent on practice and mean accuracy on the OEIT

The same regression model was applied to data set of the participants who were not aware that the OEIT was a grammar test (N = 8). This model was not significant (adjusted R^2 = .76, $F(5, 1)$ = 4.752, p = .33).

6.5 Discussion

6.5.1 Hypothesis 1

Practice results in automatized L2 grammar knowledge as manifested in high accuracy rates and fast response times on transfer tasks.

The first hypothesis was largely confirmed. Controlled practice resulted in higher accuracy rates on tests of near transfer (measured by the TGJT), far transfer to written production (WDCT), and far transfer to spoken production (OEIT). The latter finding only applied to the participants that had been aware they were being tested on grammar. There was also evidence of automatization as manifested by faster response times on the post-tests of the TGJT, as well as by the significant correlations between the mean accuracy rates and response times on that test.

There is strong evidence to believe, however, that none of the post-tests measured implicit knowledge of the participants that practised the grammar problems. Five findings underpin the interpretation that all post-tests measured explicit knowledge instead (whether automatized or not): first, the higher standard deviations of the accuracy rates on the post-tests of the TGJT (almost double the size of the pre-tests) as well as the WDCT for the treatment groups, which suggests that learners responded with less certainty and systematicity on the post-tests, a presumed characteristic of explicit knowledge (Zobl, 1995); secondly, the lack of correlations between the pre-test of the TGJT and its two post-tests; third, the large-sized correlations between the TGJT and WDCT on the post-tests in comparison with a smaller correlation on the pre-test; fourth, the medium- to large-sized correlations between the OEIT on the one hand, and the post-tests of the TGJT and WDCT on the other hand; and fifth, the observation that most participants of the OEIT were aware they were being tested on grammar, even if the test involves the strongest focus on meaning of all tests. Therefore, the finding that the treatment groups did better on the post-tests of the TGJT than on the pre-test, and that they outperformed the control group should probably be seen as evidence of automatization of explicit knowledge, rather than evidence of automatization impacting on implicit knowledge. In all, it seems reasonable to believe that there was no transfer of practice to implicit knowledge.

The following transcript from the OEIT supports this idea (<*err*> indications an audible hesitation ; <*int*> indicates an interjection in Dutch):

> Researcher: *Pemberton Junior has less screws loose than me.*
> (Learner writes down that the statement is a lie according to the story.)
> Learner: *<err> John Pemberton Junior has fewer <err> fewer <err> <err> <laughs> I don't know the name <err> has fewer <err>*
> Researcher (in Dutch): *What is the Dutch word?*
> Learner: *Vijzen.*
> Researcher (in Dutch): *Can you give another word in Dutch?*

(Researcher helps with the Dutch vocabulary.)
Learner: *Schroeven.*
Researcher (in Dutch): *Can you say it in English now?*
Learner: *<err> less <err> sc- scarfs <laughs> <err> than <err> me <int>*
Researcher: *Can you repeat it once again?*
Learner: *<err> John Pemberton Junior has less <err> scarfs <err> <int> than me.*

This excerpt comes from a session with a learner who afterwards admitted having been told by another learner that the role-play was a grammar test. It shows that the learner initially produces the quantifier *fewer*, after hearing the stimulus statement containing the construction *less screws*, which was ungrammatical according to the rule instruction. However, upon having trouble producing the head noun *screws*, he again uses the quantifier *less*: once while producing the construction, and once again when repeating the full sentence. This suggests that the learner was closely monitoring the researcher's statements in the role-play, likely focusing on the form-form pairings, drawing on declarative knowledge when correcting this particular construction, and reverting to implicit knowledge when he was producing form-meaning connections. In other words, in this particular excerpt, and potentially likewise for the other participants that were aware the OEIT was a grammar test, the task did not engage learners in joint form-meaning processing, but in alternations between form- and meaning-focus.

6.5.2 Hypothesis 2

The effects of practice supported by ML CF will be higher than the effects of practice with KR CF on transfer tasks that allow for the use of explicit knowledge and thus allow learners to monitor their performance. There will be no such difference on transfer tasks which are thought to preclude the use of explicit knowledge.

The second hypothesis was also largely confirmed, even if this was due mainly to the fact that the TGJT did not function as intended (see results for the first hypothesis). On the WDCT, i.e. the test that created maximal conditions for using declarative knowledge, the group that was reminded of the declarative information during practice by means of ML CF outperformed the group that was withheld this information in terms of mean accuracy rates. This effect was evident on both post-tests.

For the TGJT, which was assumed to preclude the use of declarative knowledge because it was time-pressured, this effect was also present on the first post-test, but disappeared on the delayed post-test. It seems likely that, despite the time pressure, the learners were also capable of using declarative knowledge on the TGJT to some extent. This lends support to the idea that TGJTs are not pure measures of implicit knowledge (Loewen, 2009). Further, this shows that contextual factors, such as learners' participation in explicit instruction with practice, may make TGJTs even less suitable for gauging implicit knowledge. Evidently, the similarity of the practice tasks with the TGJT and their metalinguistic nature are very likely to have contributed to this. What is more, the finding that the added value of metalinguistic CF disappeared on the delayed post-test may suggest that this automatized explicit knowledge was not robust.

The presence of ML CF in the practice tasks did not affect transfer to the OEIT. Potentially, learners had more difficulty retrieving declarative knowledge on the OEIT. This may be due to the fact the OEIT was a complex task, requiring them to focus both on meaning and on form, and with some amount of time pressure.

6.5.3 Hypothesis 3

> *On tasks that are considered to prevent the use of explicit knowledge, the effects of practice will be higher for items offered in practice than for*

non-practice items. For the former items, namely, there will be a memory effect of practice.

The effects of practice on the accuracy rates of the TGJT were as strong for non-practice items as for practice items. This suggests that learners had developed generalized knowledge during practice, which seems to run counter to the hypothesis. Learners did, however, respond faster to the practice items on both post-tests than on the pre-tests, which confirms the hypothesis. Since there were no interaction effects with the treatment type, metalinguistic CF may not have aided generalization on this test. Perhaps, rather than having relied on declarative knowledge while judging new items, the learners may have judged the new items on the basis of their similarity with the items offered in practice.

The results for the accuracy rates on the WDCT cannot confirm or disconfirm the hypothesis, since no significant effects were found involving the item type. The results for the response times are difficult to explain. Given that there is only a main effect of item type, the results suggest that the learners did not become faster at responding to these items than on the pre-test. Perhaps this is due to the fact that the WDCT was a far transfer task, requiring learners to apply their knowledge acquired in practice in a more complex and productive task. Memory effects from practice may be less evident on such tasks.

6.5.4 Hypothesis 4

The effects of practice will be stronger for grammar problems which comprise simple rule explanations.

The fourth hypothesis was univocally confirmed. The effects were clearly stronger for the simple rule (QNT), which is hardly surprising.

However, it is noteworthy that the accuracy rates for QNT decline somewhat on the post-tests, and that those for V2O remain more or less stable, and even seem to rise somewhat on the WDCT. Moreover, on the WDCT, the differences between QNT and V2O disappeared on the delayed post-test. Given that QNT is the more simple rule, this is remarkable. This trend can be explained by the often heard comment from teachers that the rules for QNT are known to be resistant to instruction—Dutch-speaking learners of English tend to forget them time and again. A second—and likely related—explanation is that conservative rules for QNT, as often included in pedagogical grammars, no longer reflect actual English language use. As noted in section 6.3.2, the quantifiers *less* and *least* are being used increasingly more frequently with countable nouns in less formal registers of English. Flemish youth are being exposed at a relatively young age to less formal English through television series, movies, and *off-the-shelf* (i.e. non-educational) games, and empirical research suggests that their long-term exposure to these media is associated with their development in a L2 as measured by oral translation tasks (Kuppens, 2010). Given the learners' exposure to less formal registers of English, 'backsliding' for constructions involving the quantifiers *less* and *least*, subsequent to explicit practice according to more conservative rules, may be expected. This would mean that on the delayed post-test, learners may have been relying more on implicit knowledge again when judging these constructions. The accuracy plots for the different constructions on the TGJT (see Figure VI-16) seem to back this argument.

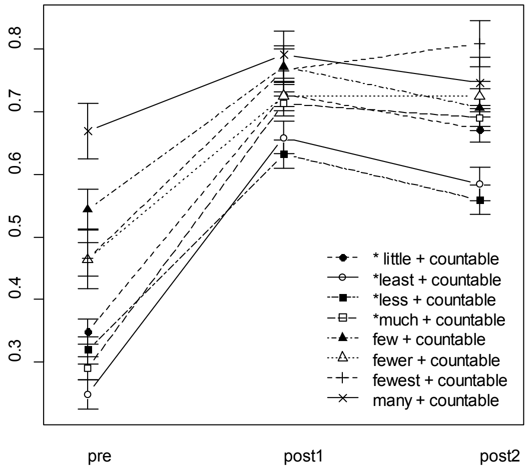


Figure VI-16: plots of mean accuracy for the items on QNT on the TGJT, by test time and construction

The mean accuracy on the TGJT of the items related to V2O did not decline on the delayed post-test, despite the rules for V2O being far more complex than those of QNT. A first possible explanation for this is that judging the grammaticality of V2O on the basis of the provided rule instruction involved some focus on the meaning of the sentences, as one step in the reasoning required learners to decide whether the sentence expressed transfer of possession. Making this decision is not possible without attending to meaning. Hence, learners may have been more engaged in form-meaning processing for V2O than for QNT, of which the grammaticality can be judged purely on the basis of form-form pairs. A second—and highly speculative explanation—is that the more stable scores for V2O on the delayed post-test are an effect of implicit learning, as constructions of V20 appeared with a relatively high frequency in the reading text (24 items to be exact; see section 6.3.3, phases 6 and 9, and Appendix 5), and were repeated by the researcher in the discussion activities in class. This may have resulted in more robust implicit knowledge.

6.6 Conclusion

This study intended to investigate to what extent controlled practice with CF can benefit L2 learning. Results show that controlled practice with CF can result in automatized and accurate knowledge in making grammaticality judgments and accurate knowledge in written transfer tasks, especially if metalinguistic CF is included in practice. This implies that learners who practise their knowledge of grammar with similar tasks as the ones used in this study are likely to become more grammatically accurate on writing tasks, as well as to notice mistakes quickly while revising texts. There also was a small transfer effect from the practice tasks to a more complex productive speaking task. Given the great differences in skills applied in practice and on the transfer task, this small effect is in line with Skill Acquisition Theory and with the Transfer-Appropriate Processing hypothesis (Lightbown, 2008).

There was no evidence of transfer to implicit knowledge, given the evidence that the tests intended to measure such knowledge failed to capture it. At best, the tests used in this study may have measured the degree to which the participants' grammar knowledge (not skill) was automatized. DeKeyser (2007b) notes that "automatized knowledge is not exactly the same as implicit knowledge", and further explains that lack of awareness is not a prerequisite for automaticity (p. 4). Perhaps by means of focused practice on more meaning-oriented tasks, implicit knowledge could eventually be developed, side by side with automatized procedural knowledge, as Hulstijn's version of the non-interface position has it (2002). But before this can be investigated in research on real L2 learning, more valid and fine-grained measures of explicit and especially implicit L2 knowledge will need to be developed. Zoltan Dörnyei (2009) seems hopeful on this point:

> Let me point out that there are recent arguments that claim that with our increasing understanding of the cognitive operations in the brain, many existing distinctions previously described in purely functional, binary terms, such as the explicit-implicit or the declarative-procedural

> distinctions, can now be characterized in a more graded manner [...]. Thus it is likely that the explicit-implicit duality will be replaced by a more refined framework before long. (p. 135)

Naturally, this study on controlled practice was limited in a number of respects. First, while the instructional design was intended to involve learners in meaningful controlled practice by embedding the content and format of the practice materials in an engaging mystery story, it is more likely that there was an alternation between meaning focus in the reading and discussion activities and strong form focus in the practice tasks. The practice tasks were essentially metalinguistic in nature, involved comprehension skills rather than production skills, were fast-paced, and did not require the learners to attend to meaning. This will probably have resulted in learners tuning out of meaning once they had the chance to do so, even if the researcher emphasized that learners had to pay attention to the content of the practice items in order to resolve the mystery. Assuming that this really was so, it is likely that the learners were engaged in mechanical practice—a "very peculiar [and] 'language-like behavior'", to use DeKeyser's (1998, p. 53) words, which consists of scanning L2 input for formal analogies and pairing forms with forms, without taking meaning into account. This would have precluded any potential for developing, proceduralizing, and automatizing L2 skills.

Thus, future iterations on this instructional design need to make sure that learners attend to meaning in the practice tasks. The format of the OEIT used in this study seems like a good candidate for this, because it also requires learners to produce a response in terms of meaning. Repetitive role-plays like this format play a crucial role in Gatbonton & Segalowitz' (2005a) instructional design model known as ACCESS, which is intended to bridge communicative meaning focus with controlled practice of forms. Moreover, the format of the OEIT seems sufficiently limited in terms of the range of expected responses in learners' spoken production, so perhaps automatic speech recognition may have something to offer here in the future.

Another issue is one of ecological validity. Ungrammatical constructions of V2O and QNT are not likely to raise many problems in communication. Consequently, a focus on accuracy for these constructions may not be entirely justified, at least not if the objective of instruction is to support the development of implicit knowledge for speaking purposes. Obviously, some sacrifices need to be made in terms of ecological validity in the favour of research methodology.

A methodological limitation was that the TGJT was written, which not only caused a considerable amount of stress for all learners, as well significant problems for learners with dyslexia, but this was also problematic because the rules for V2O required making a judgment on an auditory aspect (verb stress). Oral TGJTs may help to get rid of this problem; see also comments in the discussion section of Loewen's (2009) book section.

A final limitation is that we do not have fluency measures for the OEIT data. This is in part due to our data collection method: learners' speech was recorded on tape, and we did not have separate timestamps for learners' responses to the meaning of the stimuli. This entails that considerable additional effort is required in order to process these data, which is outside the scope of our project. In the future, a computerized version could record the time it takes to judge the truth value of the sentence, and perhaps detect pauses and hesitations in the learners' spoken production by means of automatic speech recognition. In this way, speech rates can be computed, both pruned and unpruned, resulting in a measure of fluency.

Acknowledgements

The conceptual design of the technology-based practice environment used in this study, but not its development, was partly realized through interaction with the Games Online for Basic Language learning (GOBL) project (519136-LLP-2011-NL-KA2-KA2MP), funded with support from the European

Commission. This publication reflects the views only of the author, and the Commission cannot be held responsible for any use which may be made of the information contained therein.

The PhD candidate developed and deployed the technology through the web application framework Google App Engine. The client-side application was built with HTML5, and relied on techniques and some source code from the *HTML5 Game Development* course from Udacity (CS255). Icons are from the Coquette icon set. The practice activity contains a still from the film noir project *The Big Smoke* (copyright Kenneth Gawne). The omission of any other due credit is entirely unintentional.

We wish to thank dr. Wilfried Cools and Prof. dr. Wim Van Den Noortgate for their kind methodological advice.

Chapter VII

Discussion and conclusion

7.1 Introduction

In this day and age, digital games are all around us. We may be concerned about their potential negative effects on human behaviour and development, and we may hence try to resist them, but they are here to stay. Taking this into account, recent research on human engagement in digital games has marked a shift from trying to understand the relationship between gaming and various forms of negative behaviour such as aggression, social isolation, and overuse, towards a more positive research agenda that emphasizes the potential affordances of gaming and game mechanisms for supporting physical, psychological, as well as social change and well-being (e.g. Bogost, 2011; McGonigal, 2011; Walz & Deterding, 2014). One such recent strand of research is based on Self-Determination Theory (see Przybylski, Rigby, & Ryan, 2010; Rigby & Ryan, 2011), and is on the quest of explaining the appeal of digital games in terms of their ability to satisfy basic psychological needs that are assumed to apply universally to human beings.

When digital games satisfy our needs, a lot of this has to do with feedback. This research project dealt with feedback which signals to an individual that a particular action did not achieve a predetermined goal, known as *negative feedback* in the psychological field of concept learning (Schachter, 1991). Negative feedback represents only a portion of the feedback that we get in gameful experiences. For instance, games also give us positive feedback in the form of verbal praise and extrinsic rewarding; there is outcome feedback, which shows us what goals we have accomplished in a (typically fictional) virtual world; there are *leaderboards*, comparing our scores with those of our competitors. Moreover, in online games, there is feedback generated by algorithms that were designed by game developers, and feedback given by peers. These different types of feedback may all mediate human engagement and development in different ways.

However, when it comes to digital games specifically engineered for learning, negative feedback that is provided consistently is likely to play a key role. Learning from failures is widely acknowledged a good learning principle (Cannon & Edmondson, 2012), and serious games designer Marc Prensky writes that “doing and failing—or trial and error—is a primary way to learn” in games, and that feedback plays an essential role in helping learners to learn from their mistakes (2001, pp. 158–159).

The objective of this research project was to investigate the effectiveness of negative feedback in the area of digital game-based language learning. As was explained at large in the second chapter, negative feedback in the domain of language learning is typically known as ‘corrective feedback’ (CF), and is geared specifically towards helping learners achieve formal accuracy in linguistic performance. This research project started from the hypothesis that there is no simple cause-and-effect relation between CF and second language (L2) development in games, but that a number of factors are likely to determine whether CF will help learners to become more proficient language users. More specifically, the effectiveness of CF for L2 learning was claimed to depend on the type of CF given and on which type of L2 knowledge it results in, while learners’ perceptions of CF as an element of the learning environment, and their perceptions of themselves as receivers of CF, were expected to mediate the instructional effectiveness of CF.

This chapter starts by summarizing the main findings as they relate to the central research questions articulated in the second chapter. We then discuss the limitations of this research project. Finally, we propose directions for future research on the development of automaticity in a L2 supported by technology-enhanced learning environments that rely on games and gameful instructional designs.

7.2 Summary of results

This section summarizes the main results of this research project in light of the central research questions formulated in the second chapter. We also discuss two notable methodological results.

7.2.1 Research question 1

How useful do learners find CF in digital game-based language learning?

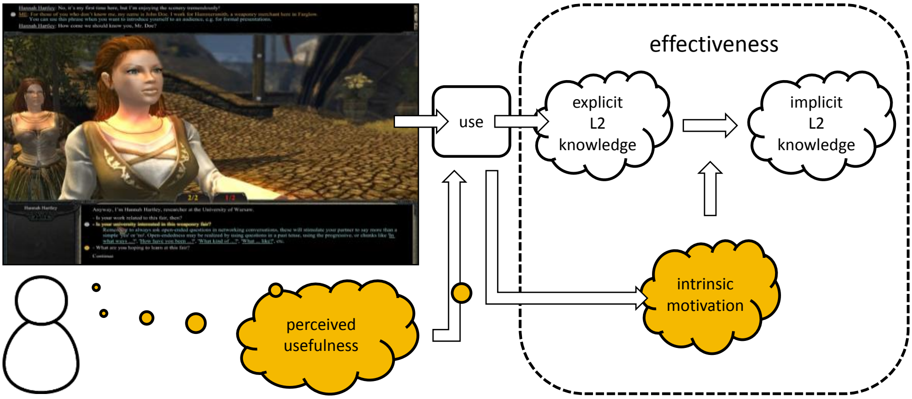


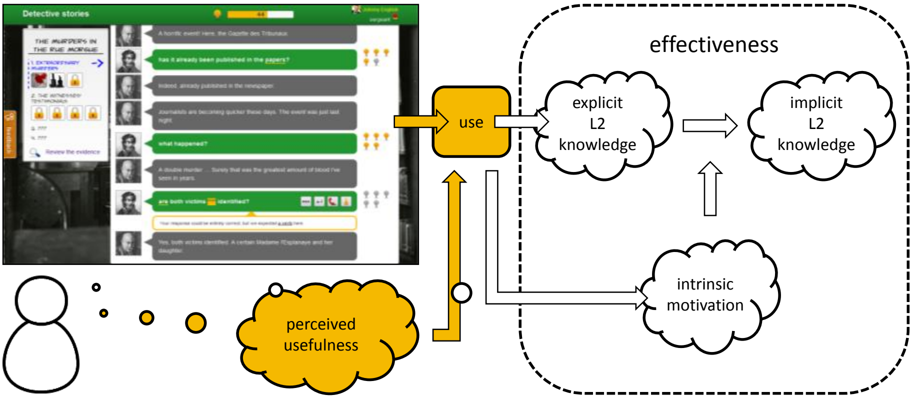
Figure VII-1: learning environment and scope of the first study

The first central research question was answered positively in the first study. Results from interviews held with learners after their experience with an immersive 3D game designed for the instruction of English pragmatics, in combination with data from questionnaires, showed that learners considered the CF embedded in the experience useful for learning, as well as for realising transfer to contexts outside the game. Moreover, explicit, rule-based metapragmatic explanations given immediately in the game dialogues were generally found more useful for learning than and preferred to more implicit simulations of communicative CF provided in the game dialogues. Lastly, three parameters related to learners' self-perceptions were found to correlate with learners' perceptions of explicit CF as measured by the questionnaires. This was, as anticipated, in the positive direction for prior intrinsic goal orientation (i.e. learners' intrinsic interest for learning English) and perceived competence as the result of playing, but also, contrary to expectation, in the positive direction for learners' game experience, defined as the degree to which learners were immersed in the experience of playing the game, felt captivated by its vividness, and felt generally good as the result of the experience. While no such relation was found for implicit CF, learners commented in the interviews that the latter CF type absorbed them in the virtual world represented in the game, showing the impact of their actions, and that a combination of explicit and implicit CF seemed best to them. The study also allowed us to fine-tune our conceptualisation of the construct of 'intrinsic motivation', with a view to investigating its potentially mediating role in the effectiveness of CF.

The finding that learners found CF helpful for their learning, in particular explicit CF, is consistent with results from previous research in a wide range of instructional L2 contexts (e.g. Chenoweth, Day, Chun, & Luppescu, 1983; Nagata, 1993; Radecki & Swales, 1988; Schulz, 2001). However, the finding that this applied to an environment in which the spotlight was on situated, agentive, and meaningful interaction in the L2 and on playful immersion in experience is new, and seems to bode well for the design of educational games for L2 learning. It also suggests—pending empirical validation in game-based language learning settings, but in accordance with a convincing body of educational research (Kirschner, Sweller, & Clark, 2006)—that completely unguided discovery L2 learning is likely to be less effective than more structured and explicit L2 instruction.

7.2.2 Research question 2

How does the perceived usefulness of metalinguistic CF in digital game-based language learning explain the actual use of such CF?


Detective stories

THE MURDERS IN THE RUE MORGUE

1. EXTRAORDINARY MURDERS

2. THE WITNESSES' TESTIMONIALS

Review the evidence

A horrific event! Here, the Gazette des Tribunaux

has it already been published in the papers?

Indeed, already published in the newspaper.

Journalists are becoming quicker these days. The event was just last night.

what happened?

A double murder ... Surely that was the greatest amount of blood I've seen in years.

are both victims identified?

Your response could be entirely correct, but we expected a verb here.

Yes, both victims identified. A certain Madame l'Espanaye and her daughter.

Figure VII-2: learning environment and scope of the second study

This research question was addressed in the second study, and was informed by the view adopted by many current theories on human learning that learners do not necessarily process instructional cues such as CF simply because they are included in instructional designs (e.g. Butler & Winne, 1995). We chose to investigate this question in an interactive murder mystery with semi-open written activities, which allowed learners plenty of possibilities to revise their L2 production through interaction with immediate 'knowledge of results' feedback and optional metalinguistic information.

However, the question could not be answered on the basis of the data. This may be due to methodological issues. In the learning environment used for this study, CF was given on learners' written responses in the form of highlighting, provided immediately when learners submitted a response, combined with optional metalinguistic prompts that were available for each highlighted word. Use of optional metalinguistic CF was measured by dividing the number of clicks on words by the total number of words highlighted in the sentences. Perceived usefulness of this optional metalinguistic CF was measured by means of questionnaires after learners had worked in this environment for 25 to 40 minutes. A correlation analysis revealed that there was no association between perceived usefulness of optional metalinguistic CF and learners' actual use of this CF.

Two possible methodological explanations were given for this lack of association. A first possible explanation is that learners' perceptions were inaccurate with respect to their actual learning processes, e.g. they may have found the CF more useful than their actual use indicated, or perhaps their reports on the questionnaire were biased. A second explanation is that learners restricted their judgments on the questionnaire to the highlighting CF that was shown immediately, even though the questionnaire specifically addressed the optional megalinguistic CF. We will revisit this issue in section 7.3.4.1, and will propose ways of circumventing it in future research.

Moving beyond the general conceptual framework of this PhD project (see chapter 2), the study did find a positive association between the use of optional metalinguistic CF and prior declarative knowledge (i.e. knowledge about metalinguistic terminology and grammar rules). In other words, learners who had a greater command of metalinguistic terminology and grammar rules used the optional CF more. This finding is somewhat in line with results from Brandl's (1995) and Heift's (2002) studies in CALL, although in these studies, it was learners' performance in the L2 that explained use of optional CF in grammar practice tasks, which may or may not correlate with declarative knowledge. In any event, the current finding has clear implications for the pedagogical implementation of similar technology-mediated tools for language practice in classroom-based L2 learning, namely that learners need to be equipped with the necessary declarative knowledge to decode and use the information provided in metalinguistic CF.

7.2.3 Research question 3

How does vivid CF affect learners' intrinsic motivation and their willingness to practise in digital game-based language learning?

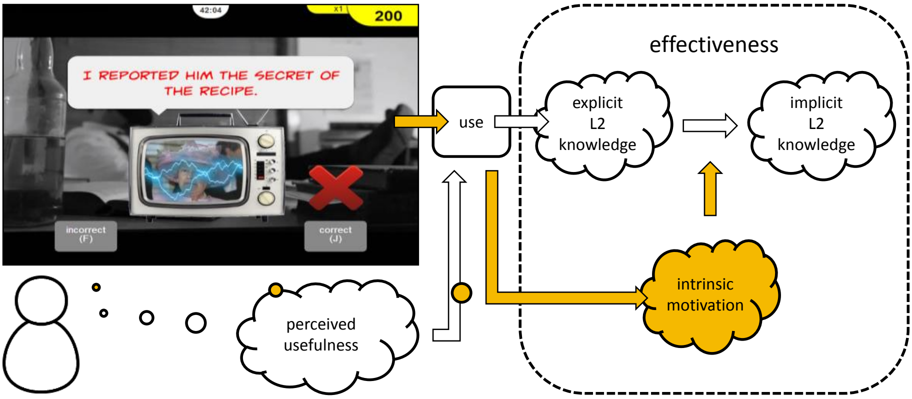


Figure VII-3: learning environment and scope of the third study

This research question was investigated in the third study. To address this question, we used speeded mini-games with immediate 'knowledge of results' CF, followed by delayed metalinguistic information; three different versions of this mini-game were developed (one without fantasy and with plain CF, one with fantasy and with plain CF, and one with fantasy and vivid CF) in order to assess the effects of CF embedded within a fantasy environment as well as of vividness of CF on learner motivation.

On the basis of the results of this study, there is tentative support for including fantasy and vivid CF in mini-games for language learning in order to heighten learners' intrinsic motivation as well as their willingness to practise. The data showed that when learners received CF that was embedded within a fantasy context, this elevated their level of perceived competence, and also found that vivid CF in a fantasy context, operationalized by means of animations and sound effects, engendered the highest degree of perceived immersion. Further, perceived competence and immersion, which are considered antecedents of intrinsic motivation in digital game-based experiences (Ryan et al., 2006), were significantly related to participants' intrinsic motivation (i.e. interest and enjoyment), which in its turn was strongly related to their willingness to practise language with such mini-games in the future.

However, there was no conclusive evidence that vivid CF by itself impacted on perceived competence or immersion, only that vivid CF in combination with fantasy increased immersion, and that fantasy without vivid CF strengthened competence. Moreover, post-experimental interviews with volunteers indicated that the animations and sound effects distracted from the learning content, and even frustrated learners to a certain extent.

7.2.4 Research question 4

How does continued practice with CF in digital game-based language learning assist learners in developing L2 grammar knowledge?

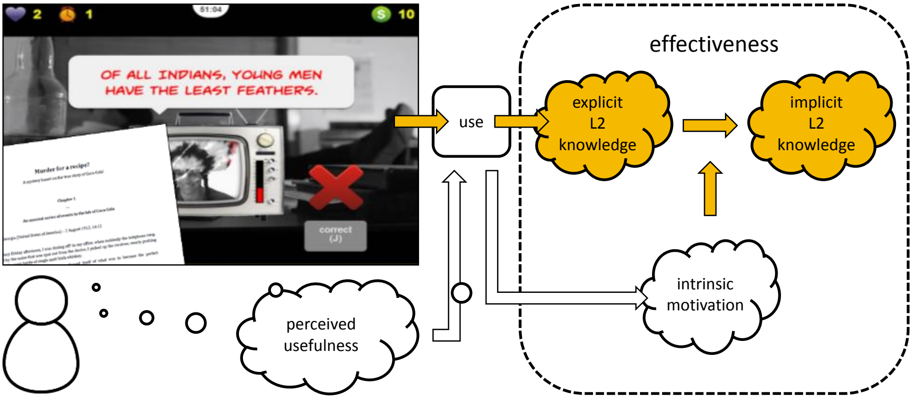


Figure VII-4: learning environment and scope of the fourth study

L2 practice in the fourth and final study was equally supported by mini-games, but was now embedded in meaningful reading and discussion tasks. This was done with a view to engaging learners in meaningful L2 processing during controlled practice. Vivid CF was also provided, but was reduced in terms of perceptual salience, so as not to interfere with learners' cognitive processing during practice.

The results showed that intensive practice with CF supported by mini-games and a mystery story helped learners to develop L2 grammar knowledge that was useful for their performance on various transfer tasks. There was evidence of transfer and generalization of learning to a follow-up task that was highly similar to the fairly simple and mechanical practice tasks (i.e. near transfer), as well as evidence of transfer to more complex written and spoken follow-up tasks (i.e. far transfer). Moreover, on the near transfer task, the knowledge developed during practice was quickly available, especially for items that were offered in practice, but also for novel items.

Moreover, the effects of metalinguistic CF (i.e. CF that reminded learners of the grammar rules explained prior to practice) were stronger than the effects of CF which did not include any metalinguistic explanation (i.e. 'knowledge of results' CF). The added value of metalinguistic CF was especially evident on the written test of far transfer, which maximized the potential for use of explicit knowledge because learners had ample time to think and apply their knowledge of the grammar rules. A more unexpected finding was that metalinguistic CF also had a positive effect on learners' accuracy scores on the immediate test of near transfer, in which the potential for using explicit knowledge was limited—but not ruled out—because the task was time-pressured; this effect disappeared on the post-test. Finally, metalinguistic CF did not benefit learners' accuracy scores on the spoken test of far transfer, which also involved a certain amount of time pressure as well as a stronger meaning focus than the other two tests.

To summarize, these findings suggest that practice with CF helped learners to develop knowledge that was accurate and quickly retrievable—an open question remains to what extent this knowledge can be considered 'automatic' (see section 7.3.4.4)—and that metalinguistic CF aided learners in realising transfer to follow-up tasks, excepting a highly meaning-focused and complex spoken grammar task. This is consistent with the hypothesis raised in the conceptual framework of this PhD project (see the second chapter), namely that practice with output-prompting CF caters particularly to the development of explicit knowledge, and that metalinguistic CF plays a key role in supporting learners to transfer their learning to other tasks by way of monitoring their performance on the basis of explicit knowledge.

This would seem to corroborate Krashen's (1981) Monitor Theory—namely that explicit and implicit knowledge in L2 learning are completely dissociated, and that explicit learning and practice cannot cater for the development of implicit L2 knowledge, or 'acquisition' in Krashen's terms—were it not for the fact that the practice tasks used in this study were essentially metalinguistic and probably mechanical in nature. Future research is needed with meaningful technology-mediated practice tasks, potentially followed by communicative practice tasks in class, in order to examine whether practice that involves joint form-meaning processing eventually leads to automatized procedural knowledge that is nearly indistinguishable from implicit knowledge. Another argument against the claim that this study showed that explicit learning and practice cannot lead to acquisition, is that the tests used in this study are not likely to have measured implicit knowledge at all. The next section deals with this issue.

7.2.5 Methodological results

In addition to the results that were found with a view to answering the main research questions, two methodological findings emerge from this project. The first relates to the tests used to measure L2 knowledge in the fourth study, the second concerns the measurement of learners' behaviour in practice.

7.2.5.1 The measurement of 'implicit L2 knowledge' (study 4)

A substantial challenge for the fourth study was to develop measures of implicit L2 knowledge of the linguistic items that formed the object of practice, in view of the skill acquisition perspective on L2 development, which states that continued practice results over time in knowledge that is virtually indistinguishable from implicit knowledge in terms of performance (i.e. high accuracy rates and fast response times). To this end, two tests were developed based on formats designated by R. Ellis (2005, 2009b) as indices of implicit knowledge, namely a timed grammaticality judgment test (TGJT) and an oral elicited imitation test (OEIT). The TGJT is assumed to measure implicit L2 knowledge, as it largely limits learners' possibilities to retrieve explicit knowledge by means of strict time constraints for each test item. The OEIT as described by Erlam (2009) is considered a measure of implicit L2 knowledge because it primarily draws a learner's focus on meaning, because there is some delay between the stimuli and their reproduction, and because it involves some amount of time pressure. Recall that the TGJT was used both as a pre-test and post-test in the fourth study, the OEIT was only used as a post-test.

According to R. Ellis (2005, 2009b), lack of conscious awareness of grammar rules is a critical condition for a task to be a valid measure of implicit L2 knowledge. In his conceptualization of implicit knowledge, intuitive awareness (or 'feel') is involved, and conscious awareness of rules is not at play. However, as was shown in our fourth study, the fact that the learners had participated in explicit instruction and practice of the grammar rules which formed the content of the language tests seems to have compromised their usability as measures of implicit knowledge. In contrast with the pre-tests, after the treatment, learners were obviously aware that they were being tested on two particular target problems, so it is highly probable that they relied more on explicit knowledge while completing the post-tests. Analyses of the standard deviations of the accuracy rates, the intercorrelations between the different test types before and after the treatment, and observations of and debriefing with the participants, provided additional evidence that the post-tests of the TGJT were less pure measures of implicit knowledge than its pre-test, and that the OEIT equally was not a good index of implicit knowledge. Thus, our results show that the TGJT and OEIT formats designed by R. Ellis (2005, 2009b) did not function as suggested in the literature, i.e. as measures of implicit knowledge.

An open question is whether any test which learners complete subsequent to an explicit training phase can at all measure implicit knowledge, in particular if instruction is focused on a small number of grammatical structures, as the learners are highly likely to have awareness of which structures are being tested. This may even be the case if the structures are scrambled in the test and if distractor items are included. The TGJT has been recently used in another experiment in SLA to measure the effects of instruction and practice (de Vries, Cucchiarini, Bodnar, Strik, & van Hout, 2014), but in this study it is not clear to what extent this test was intended as a (relatively pure) measure of implicit L2 knowledge. Further, the OEIT was used in an intervention study by R. Ellis, Loewen & Erlam (2006), who treated this task as "a measure of unplanned language use" (p. 351). However, these researchers do not seem to question the construct validity of the test subsequent to their treatment phase, pointing to the small number of self-corrections for the treatment groups on the post-test, which according to them indicates that learners were not monitoring their speech. This does not seem to be conclusive evidence, since highly automatic knowledge may help learners to plan their speech quickly, resulting in few hesitations in actual production (response times on the OEIT, or other measures of fluency, were not analysed in this study). Admittedly though, the treatment in R. Ellis, Loewen & Erlam (2006) was less explicit than in our study, which may have 'primed' their participants to a smaller extent than in our study.

To some extent, however, the finding that the TGJT and OEIT probably did not measure implicit L2 knowledge is a false problem. DeKeyser (2005) does not consider lack of awareness a key condition for ascertaining the effect of practice on automatization:

> The point is whether the declarative knowledge that results from explicit learning processes can be turned into a form of procedural knowledge that is accessible in the same way as implicitly acquired knowledge [… i.e.] that it be available with the same degree of automaticity as implicitly acquired knowledge. […] Moreover, it is quite possible that, after large amounts of communicative use and complete automatization of the rules, learners eventually lose their awareness of the rules. At that point they not only have procedural knowledge that is

> functionally equivalent to implicitly acquired knowledge, but even implicit knowledge in the narrow sense of knowledge without awareness. (p. 328-329)

Hence, according to DeKeyser, practice results first in automatized procedural knowledge, and perhaps later also in implicit knowledge—the difference between both types of knowledge, however, is deemed irrelevant in a skill acquisition perspective.

To summarize, assuming that the learners in our fourth study were conscious of the grammar rules while completing the post-tests—which the evidence seems to suggest—it is conceivable that we not measuring implicit knowledge, but automatized procedural knowledge. This methodological result has important implications for future intervention-based research in SLA: researchers need to be mindful that the TGJT and OEIT may not always function as measures of implicit knowledge.

Still, we can question the validity of the construct that the post-tests of the TGJT and OEIT may have measured instead of 'implicit knowledge', namely the construct of 'automatized procedural knowledge'. This issue will be further taken up as a limitation in section 7.3.4.4.

7.2.5.2 The measurement of learners' behaviour in practice

A second methodological result concerns the measurement of learners' behaviour in practice. All three technological environments that were developed in order to collect data on learners' perceptions and their L2 development recorded learners' behaviour in practice, such as the amount of time spent on the practice tasks, the responses which learners selected or produced, the optional CF which learners requested, up until learners' response times for individual items. Thus, learner behaviour was measured in fine-grained ways and unobtrusively, creating no interference with the learners' interaction in the L2. Moreover, these data were collected efficiently and cheaply, and were formatted in structured ways, virtually ready for subsequent analysis.

This creates opportunities for learner modelling, a technique used in intelligent tutoring systems to infer information on learners' cognition, individual differences, and perhaps learning styles on the basis of their behaviour, and to subsequently personalize their learning experience (for a review, see Vandewaetere, Desmet, & Clarebout, 2011). The application of learner modelling and intelligent adaptive tutoring in the field of game-based learning is a logical next step, and has attracted some attention in recent years (e.g. Kickmeier-Rust & Albert, 2010; Shute, Masduki, & Donmez, 2010; Vandewaetere, Cornillie, Clarebout, & Desmet, 2013).

7.3 Limitations of the research project

As is the case with any research project, this PhD project was limited in a number of respects. In this section, we first discuss its limitations in terms of ecological validity, followed by constraints in relation to generalizability and research design. Further, we give an overview of the data sets that remain to be analysed. Finally, we note limitations with respect to the measuring instruments, and L2 processes.

7.3.1 Ecological validity

In this subsection, we discuss the ecological validity of the research project in function of the experimental nature of the learning environments, and the characteristics of grammar instruction and practice.

7.3.1.1 The experimental nature of the game-based learning environments

As noted in the first chapter, the prototypes of digital game-based learning environments used in this research project were designed and developed in interaction with R&D projects that were primarily oriented towards economic and/or social valorisation. Further, studies 2, 3, and 4 were carried out in the real contexts of L2 classrooms. Both these contextual factors contribute to the ecological validity of this project.

Even so, all four studies used experimental games and proofs of concept rather than finished and implemented products that are designed primarily to satisfy the needs of real end-users (i.e. learners, teachers, and other stakeholders such as parents). The learning environment used in the first study was primarily intended as a proof-of-concept of an instructional design model (van Merriënboer & Kirschner, 2007) applied to the learning of complex language learning tasks, rather than as a viable product based on thorough human-centred design aiming to address genuine learner needs. The design of the environment used in the second study was quite technology-driven, as the objective was to investigate the affordances of natural language processing and crowdsourcing techniques for generating CF on semi-open written exercises (Desmet, 2007). The design and development of the mini-games that formed the centre point of the third and fourth studies perhaps come closest to user-centred design, as they were developed in interaction with two projects in which the needs of language learners and teachers were thoroughly surveyed (Strik, Drozdova, & Cucchiarini, 2013; Zaman et al., 2012). Still, the design of the learning environments used in studies three and four could have benefited from more iterations during the design process—for instance, more iterations may have enabled us to intercept design issues earlier, such as learners' rejection of the vivid CF, and their mechanical processing during controlled practice.

To the benefit of the ecological validity of research on game-based learning, the ideal approach is to use games that are based both on theory/pedagogy and on thorough empirical user research, which have proven to be actually used in real contexts and have been adopted by its target audiences, and have potentially passed the test of marketability. Examples of such products in the area of digital game-based language learning are few, namely the commercially available *My Word Coach* series, which is marketed by Ubisoft and comprises a suite of mini-games, the more recent—yet pedagogically questionable—mini-game-based apps *DuoLingo* and *Mindsnacks*, and the 3D immersive experiences developed by Alelo Inc., of which *Tactical Language and Culture Training System* is perhaps best known (Johnson, 2007). Of these products, only *My Word Coach* has been the subject of a sound empirical study published in a respected international journal on CALL (Cobb & Horst, 2011). The evaluation of products based both on sound pedagogical design and on thorough user-centred research in authentic L2 classrooms and more informal settings such as learners' homes provides the best guarantees for the generalizability of the findings of research.

Clearly, the iterative (instructional and user-centred) design, development, implementation, and evaluation of a digital game-based learning environment that is eventually adopted by its target audience goes far beyond the scope of a PhD project, but the issue of learner-centred design is an important one nonetheless. Already in 1991, Phil Hubbard, a pioneering CALL researcher with plenty of experience in designing and evaluating CALL materials, including games, argued that a particular activity in a L2 can only be considered a game when learners actually perceive of it as a such and play it for its own sake rather than for some external reason, and that the good intentions of the teacher or instructional designer are irrelevant in this respect (Hubbard, 1991). What is more, learners' perceptions, attitudes and beliefs towards digital game-based language learning were mainly measured by way of interviews and questionnaires in this project. In the field of human-centred design, reliance on such instruments in order to discover user needs is not considered a hallmark of good user-centred design, because users are typically unaware of their needs and/or are unable to verbalize them (Gould & Lewis, 1985).

Hence, for future studies with game-based language learning environments that aim to be ecologically valid, it is crucial that in the early phases of research—i.e. before theory-driven hypothesis testing that is often dominated by quantitative measurement, including questionnaire-based research—the intended target audiences are thoroughly charted, ideally by means of qualitative methods such as those used in the discipline of human-centred design (Gould & Lewis, 1985). A good outlook for research in CALL is that, over the past decades, thorough and qualitative attention for the learner, its environment, and the design process has moved from the fringes of the field to become a mainstay (Colpaert, 2010; Hémard, 2003).

A related limitation of this research project is the lack of involvement of a skilled game designer or interaction designer. Our project adopted an interdisciplinary approach, building bridges mainly between the discipline of second language acquisition research and educational technology with a focus on learning psychology. We also made use of gaming technology. However, given the fact that the project relied on gaming designs, elements and mechanics with a view to engender meaningful, intrinsically motivated, and effective language practice, a more diversified design team comprising professional game designers or students in game design would have been a real asset.

The latter point applies particularly to the issue of CF design. Marc Prensky wrote that “designing feedback to be less learninglike and more gamelike is often a big paradigm shift and challenge for Digital Game-Based Learning designers” (2001, p. 159). With the results of this project in mind, perhaps the challenge is not so much to balance cognitive and playful aspects of CF, as Prensky seems to suggest—certainly not if this implies reducing the level of instructional support provided in CF—but to design CF in such a way that it enhances both learning and motivation. In this project, we have explored ways in which CF can be made somewhat ‘gamelike’. Undoubtedly, however, people with plenty of experience playing and designing games can come up with a myriad more ways in which CF can support the motivation of language learners. Collaboration between game designers and instructional designers on this front is key.

7.3.1.2 The pedagogical focus on grammar and formal accuracy

In this project on the effectiveness of CF, by and large, a pedagogical focus on accuracy and grammar as a formal system predominated. CF can also be given on more function-oriented aspects of the L2, such as the use of pragmatic linguistic devices (see e.g. Fukuya & Zhang, 2002; Koike & Pearson, 2005; Sykes, 2009; Takimoto, 2006; as well as the first study in the current PhD project), which is perhaps more relevant given the current-day prevalence of communicative approaches in language teaching and learning.

The focus on formal grammar was mainly motivated by practical considerations. First, there continues to be scope for language as a formal system as well as for grammatical accuracy in current-day language teaching and learning, and the use of technology has a crucial role to play in this respect. Secondly, tutorial systems such as the ones used in this project lend themselves much more easily to teaching 'the parts' of language and to providing consistent CF on formal aspects of L2 performance, than to teaching more meaning-focused aspects of the L2. Further, the effects of instruction on the development of L2 grammar are arguably easier to quantify than the effects of instruction on, say, L2 pragmatic development.

Another limitation with respect to syllabus design is that the grammar instruction in this project may have been somewhat conservative, certainly in the fourth study. For instance, more progressive pedagogical grammars (e.g. Leech & Svartvik, 1994) reflect current tendencies in English language use better, and are hence increasingly less strict with respect to formal distinctions such as those between *less* and *fewer*, and *less* and *least*. As we argued in chapter 6, this sacrifice was made for the sake of experimental rigour, as our objective was to evaluate the usefulness of explicit grammar practice for simple versus complex grammar rules. This methodological choice also yielded interesting results on the delayed post-test, where learners' performance on the constructions *less/least* + countable NP, ungrammatical according to our rather traditional rule instruction, seemed to decline to some degree. This confirms the idea that grammar practice is especially relevant for linguistic phenomena that can be accurately and consistently captured by means of grammatical explanations.

7.3.2 Generalizability and research design

Further, a couple of limitations can be noted with respect to the overall research design of this project and the generalizability of findings. First, this PhD project comprises four empirical studies in technology-supported learning environments with three markedly different designs, namely an immersive 3D avatar-based game, text-based gamified dialogue tasks, and mini-games. The findings and conclusions of one environment may not apply to other environments used in this project, let alone to L2 learning environments beyond this project. In other words, the generalizability of findings of the current project to other types and designs of game-based language learning environments cannot be taken for granted.

Secondly, as argued in the second chapter, this research project adopted a conceptual framework of learning inspired by the Cognitive Mediational Paradigm (Winne, 1987). We put forward that the effectiveness of CF was likely to be mediated by learners' perceptions, namely by their perceptions about the instructional value of CF (i.e. perceived usefulness), as well as by their perceptions about themselves as individuals interacting with CF. The mediating role of perceptions, however, was not investigated directly. The interrelations between perceptions about CF and self-perceptions on the one hand, and instructional processes and development on the other hand, were investigated separately. The second study explored the relationship between perceived usefulness of CF and its actual use (but found no association due to methodological issues). The third study provided some evidence that features associated with vivid CF in mini-games may impact on learners' intrinsic motivation and on their willingness to practise, but did not investigate the (long-term) use of those mini-games in self-directed contexts. Hence, there is only indirect evidence that the effectiveness of CF may be regulated by learners' perceptions.

Further, the results of the current research project may not apply to other target audiences. Because the project relied on questionnaires and tests of metalanguage, and required a fair degree of familiarity with computer technology, the language learners that participated in this project were mainly intermediate-level and more academically-oriented language learners. Learners with less advanced digital literacy expertise, beginning language learners, learners who take language courses with a stronger emphasis on communicative effectiveness, or learners with a less analytic command of language may all benefit less from explicit instruction with CF.

Finally, save for the fourth study, all studies involved only relatively short periods of practice, as well as a focus on groups of learners rather than on individuals. A limitation of short-term studies is that they neglect the dynamic nature of learner characteristics. Individual differences are rarely stable, but emerge from and change on the basis of the dynamic interaction between the learners and their environment over time (Dörnyei, 2009). A case in point is that of motivation, "less a trait than a fluid play, an ever-changing one that emerges from the processes of interaction of many agents, internal and external, in the ever-changing complex world of the learner" (N. C. Ellis & Larsen-Freeman, 2006, p. 563). This applies almost certainly to learners' interactions with new technologies: in early stages, they may be highly intrinsically motivated as a result of the novelty of the technology. By way of example, the participants in the third study seem to have been more intrinsically motivated by the CF which comprised gaming elements, but this 'dancing bear' effect may quickly fade away with continued practice, or vivid CF may even become boring or downright frustrating if there is insufficient variation with increasing exposure. Therefore, the finding that learners seemed more motivated by the vivid CF implemented in the third study may not be taken as representative of more continued practice.

As for the number of participants, it was largely a pragmatic choice to work with intact classes (except for the first study), and this decision may benefit the generalizability of findings to game-based practice in authentic classrooms. However, it also implies that contextual information was not captured. This was especially the case in the second, third and fourth study. In these set-ups, it would have been highly interesting to observe individual learners while they worked with the technologies, and to afterwards confront them with their behaviour in practice, so as to collect more rich information on *why* particular features of the technologies were used, used in different ways than intended, not used, or perhaps abused.

7.3.3 Unanalysed data

A great deal of the 'big data' collected within this Ph.D. project was not analysed, namely the behavioural data gathered during practice by means of tracking and logging technologies. Potential data sets for future analysis include, from the first study, the time spent in each dialogue turn, as well as learners' choices for particular pragmalinguistic devices in the three similar dialogues, with a view to examining whether attention/exposure to CF resulted in improved performance in subsequent dialogues. Similarly, the data collected for the second study demand exploration of the relation between use of optional CF and uptake—admittedly, though, the practice period was rather short to investigate this. In the third study, the effects of fantasy and vivid CF on performance in practice were not analysed. It could be expected that learners performed less well when more gaming elements were present (fantasy, animations, and sound effects). And last but not least, as for the data from the fourth study, the accuracy rates and response times for the individual responses in practice remain to be analysed, with a view to investigating automatization over time as well as learning difficulty for particular linguistic constructions.

Further, the questionnaire data of the fourth study still need to be analysed. The questionnaires targeted learners' intrinsic motivation as the result of practice. It was hypothesized that learners who received less linguistic support during practice, i.e. the group who received only 'knowledge of results' CF but not metalinguistic information, would feel more helpless and less competent as the result of practice, as they were deprived of information that could help them to improve. Observations and reactions of learners in this group who had noted that some of their peers received grammar explanations showed that the former learners indeed felt discouraged to a certain degree. In the future, mediation analyses could be carried out in order to investigate whether these self-perceptions mediated L2 development.

7.3.4 Measuring instruments

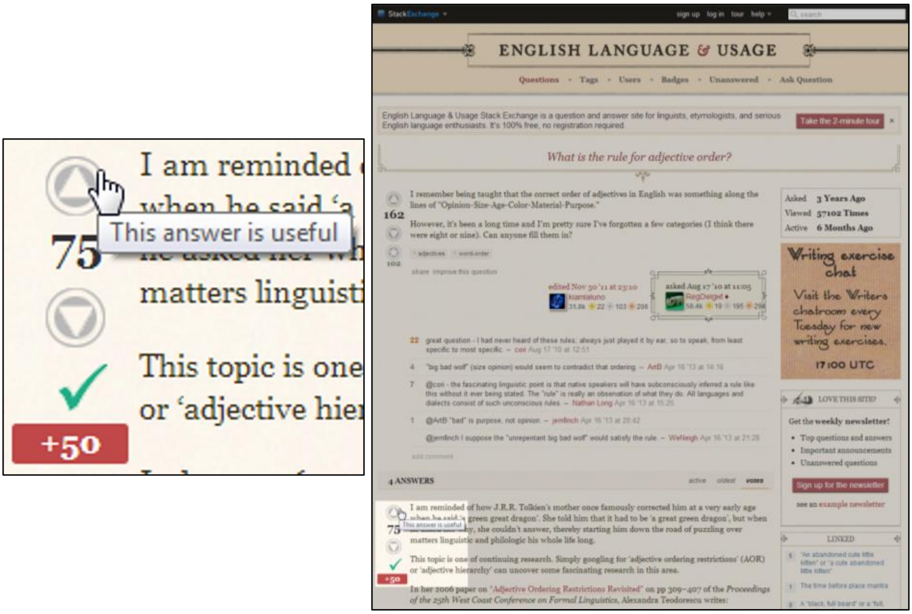
In addition to methodological issues in terms of ecological validity and generalizability, we also note limitations of the measuring instruments. In line with the order of presentation of the studies, we subsequently discuss the instruments used to measure learners' perceptions (studies 1, 2, and 3), use of CF (study 2), and L2 knowledge (studies 2 and 4), including automatized grammar knowledge (study 4).

7.3.4.1 Measurement of learners' perceptions

In this research project, learners' perceptions were measured by means of questionnaires and interviews. These measures have allowed us to draft a comprehensive picture of the effectiveness of CF in digital game-based language learning, but are subject to the limitation that they rely on self-report, which may be prone to memory bias. Moreover, such measures assume that all learners are equally capable of introspection. Observations of learners' behaviour with the technologies, thick descriptions, and other more ethnographic measures may yield different results.

A case in point is how perceived usefulness was measured in the second study. Recall that learners practised for 25 to 40 minutes, and were meanwhile supported by highlighting CF, which was given immediately, and metalinguistic hints, which became available after learners clicked on highlighted words. Learners were afterwards asked, by means of a questionnaire, to rank how useful they found the metalinguistic hints. As noted, the lack of association between perceived usefulness and actual use of CF may have been due to the fact that learners restricted their judgments to the highlighted words. In other words, the participants may have interpreted the questionnaire in different ways than the researcher had intended. The questionnaire should perhaps have been piloted in order to find out how participants interpreted the instructions, and to minimize the possibility that their interpretations deviated from the researcher's objectives. This is of course a general problem with questionnaire design, but may also be due to the fact that the questionnaire was given on paper, i.e. after participants had practised in the online environment.

A possibility for future research is to consider measuring learners' perceptions in the online environment itself, rather than afterwards on paper. Nowadays, this technique is adopted in support pages for (commercial) products as well as in online communities of practice, where users rate the usefulness of support provided by experts and peers (see Figure VII-5). In recent years, this approach has been extended to include reward mechanisms. A typical case is that users receive experience points and badges on the basis of peer rankings on the perceived usefulness of their contributions. Further, contributions that receive much support from peers are then presented on top of the page, and in this way are made to stand out from contributions which are considered less useful on average. This design feature seems to create interesting affordances for feedback in online language learning environments, and rhymes particularly well with a more dynamic view on grammar—coined 'grammaring' by Larsen-Freeman (2003)—as language learners are challenged to critically reflect on the feedback provided by both peers and experts, rather than to accept it at face value.



English Language & Usage Stack Exchange is a question and answer site for linguists, etymologists, and serious English language enthusiasts. It's 100% free, no registration required.

What is the rule for adjective order?

I remember being taught that the correct order of adjectives in English was something along the lines of "Opinion-Size-Age-Color-Material-Purpose."

However, it's been a long time and I'm pretty sure I've forgotten a few categories (I think there were eight or nine). Can anyone fill them in?

edited Nov 30 '11 at 23:10 kiamlaluno 31.8k 22 103 206

asked Aug 17 '10 at 11:05 RegDwight 58.4k 19 195 294

22 great question - I had never heard of these rules; always just played it by ear, so to speak, from least specific to most specific. – con Aug 17 '10 at 12:51

4 "big bad wolf" (size opinion) would seem to contradict that ordering. – ArtB Apr 16 '13 at 14:16

7 @con - the fascinating linguistic point is that native speakers will have subconsciously inferred a rule like this without it ever being stated. The "rule" is really an observation of what they do. All languages and dialects consist of such unconscious rules. – Nathan Long Apr 16 '13 at 15:25

1 @ArtB "bad" is purpose, not opinion. – jemfinch Apr 16 '13 at 20:42

@jemfinch I suppose the "unrepentant big bad wolf" would satisfy the rule. – WeNeigh Apr 16 '13 at 21:28

add comment

I am reminded of how J.R.R. Tolkien's mother once famously corrected him at a very early age when he said 'a green great dragon'. She told him that it had to be 'a great green dragon', but when he asked her why, she couldn't answer, thereby starting him down the road of puzzling over matters linguistic and philologic his whole life long.

This topic is one of continuing research. Simply googling for 'adjective ordering restrictions' (AOR) or 'adjective hierarchy' can uncover some fascinating research in this area.

In her 2006 paper on "Adjective Ordering Restrictions Revisited" on pp 309–407 of the *Proceedings of the 25th West Coast Conference on Formal Linguistics*, Alexandra Teodorescu writes:

Get the **weekly newsletter!**

- Top questions and answers
- Important announcements
- Unanswered questions

5 "An abandoned cute little kitten" or "a cute abandoned little kitten"

1 The time before place mantra

8 A "black, full beard" or a "full

Figure VII-5: measurement of perceived usefulness in the community of practice *English Language & Usage Stack Exchange*

Further, as concerns the use of Likert scales on the questionnaires and their subsequent analysis, psychometricians often disagree on whether Likert-type responses should be treated as interval data (i.e. ordered data with a fixed distance between each response type), arguing that they should be analysed as ordinal data instead (i.e. ordered data lacking a measure of distance between the response types). The distance between, say, a 7 and 6 on a 7-point Likert scale item may be different from the distance between a 4 and 5. In this project, 7-point Likert scales were used to measure learners' perceptions, and their responses were treated as interval data, which allowed us to investigate correlations between different perceptions, and between perceptions and other variables. This is in line with the research traditions of educational psychology that rely to a great extent on 7-point Likert scales, and in which so-called violations of statistical principles do not seem to increase the chances of drawing wrong conclusions (Norman, 2010).

7.3.4.2 Measurement of CF use

The results of this project do not reveal a clear picture of how use of CF was related to learning. This may have to do with how use of CF was measured in the second study, and with the lack of a measure of CF use in the fourth study.

In the second study, use of CF was measured by counting the number of times a learner clicked on a word to see a metalinguistic prompt, divided by the number of opportunities for such clicks. Arguably, this is a highly crude measure of CF use. Evidence of clicking does not necessarily imply that the information subsequently shown was actually processed. Moreover, as the study was rather short, it is possible that learners clicked on the highlighted words simply out of curiosity rather than for a real need for support. Other behavioural measures such as eye-tracking (Örnberg Berglund, 2012), and more qualitative measures such as think-aloud protocols and stimulated recall measures can yield a more complete understanding of how and especially *why* learners use CF.

What is more, clicking on and actually processing a metalinguistic prompt is likely to emerge from different linguistic-pedagogical needs compared to calling up a full metalinguistic rule explanation. In the former case, the learner may possess partial, or even fairly extensive, knowledge of a grammatical principle; in the latter case, the learner is likely to have a less well developed command of the grammar rule at stake. This idea is central in the concept of Dynamic Assessment in Sociocultural Theory of L2 learning, which will be dealt with in detail in section 7.4.2.

In this project, we did not apply the concept of Dynamic Assessment because it originates in a framework that focuses on regulation between humans. Dialogic regulation between human beings is characterized by rather different dynamics compared to scripted and more rigid interaction between computers and humans, which is much less sensitive to learner needs (for comments see Lantolf & Poehner, 2014, pp. 186–188), as was the case in our second study. Moreover, the materialization of Dynamic Assessment in human-computer interaction requires that the large majority of possible responsive moves of the learners are foreseen in the system's design. This requires considerably more effort in terms of data collection (i.e. actual productive language use of learners) and technology development (human language technology) than was possible in this research project.

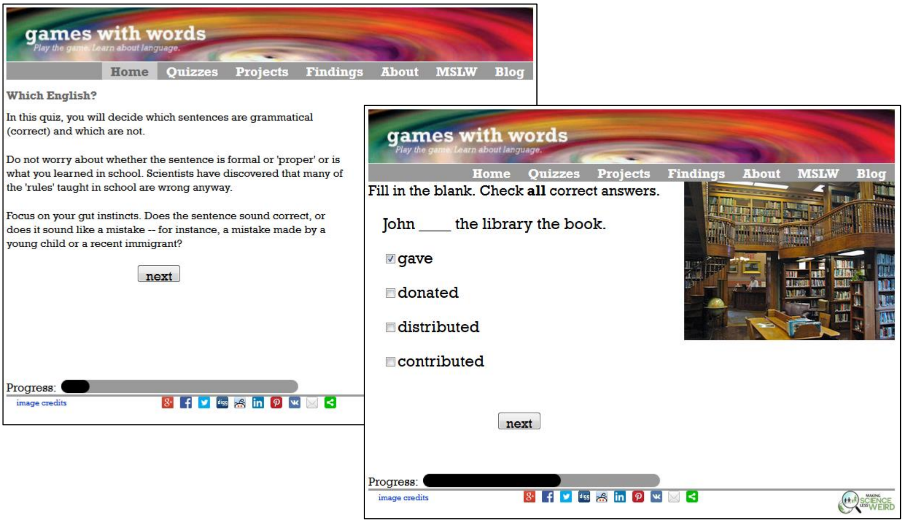
In the fourth study, it was found that metalinguistic CF given in explicit practice tasks aided performance on follow-up tasks more than 'knowledge of results' CF, even if all learners participated in a rule instruction phase prior to practice. On the basis of this finding, we may conclude that learners in the metalinguistic CF group attended to this type of CF, but we did not measure *how* this CF was actually used, for instance at which stages of practice, for which linguistic problems, and by which participants. Future studies with the practice tasks used in the fourth study could also measure use of CF, in order to allow a more fine-grained investigation of the effectiveness of CF over relatively extended periods of controlled practice, as a function of the complexity of the linguistic problems that are being practised, as well as in relation with different learner profiles.

7.3.4.3 Measurement of L2 knowledge: general limitations

In the remainder of this section, we first discuss two general limitations of the tasks used in this research project to measure L2 knowledge (this subsection), and then deal in more detail with the tasks that were used to measure automatized knowledge in the fourth study (section 7.3.4.4).

The selection and development of the tasks used for measuring L2 knowledge and development was largely based on R. Ellis' (2005) psychometric study. This was informed by practical considerations, and was in line with our objective to test the effectiveness of instruction and practice on the acquisition of particular L2 features. Task types were chosen that are fairly easy to operationalize and score. According to Norris and Ortega's (2000) classification, we used constrained constructed-response measures, such as the oral elicited imitation test (OEIT) and the written discourse completion test (WDCT) in study 4; selected-response measures, such as the metalinguistic knowledge test in studies 2 and 4; and metalinguistic judgments, such as the grammaticality judgment tests in studies 2 (untimed) and 4 (timed). We did not use free constructed-response measures, as our main interest was in investigating the effect of CF on the development of particular target language features, and elicitation of particular linguistic features in the latter type of outcome measures is known to be challenging. In other words, the task types were quite closed-ended. This may be seen as a limitation, since the corollary is that this project has little to say about the effects of practice on transfer to more communicative and open-ended tasks.

A second limitation is that more effort could have been spent on piloting the language tests for the fourth study, with a view to improving their construct validity. The main objective of this study was to investigate whether explicit instruction and controlled practice could help learners to develop knowledge of two particular aspects of English grammar (quantifiers and the double object construction). The linguistic content of the tests used to measure L2 development was based on descriptive and pedagogical grammars (we refer the reader to chapter 6 for more details). However, as was noted in section 7.3.1.2, pedagogical grammar rules and even linguists' descriptions of grammatical phenomena are not always entirely in line with actual (and ever dynamic) target language use. Combined with the fact that the tests were not piloted on L1 speakers, this may limit the construct validity of the language tests. Therefore, in future studies, the tests could be piloted on L1 users, in order to validate the linguistic content of the items. Perhaps a useful approach would be to distribute the test via social media and use crowdsourcing to attract large numbers of participants (like the *Games with Words* research initiative by MIT; see Figure VII-6), and to rely on IP address information and self-report measures in order to collect demographic data about the participants. In recent years, examples of such research initiatives have arisen which rely on gamification techniques in order to incentivize participants and collect large amounts of data, such as *The Great Brain Experiment* (Matterson, 2013).



Which English?

In this quiz, you will decide which sentences are grammatical (correct) and which are not.

Do not worry about whether the sentence is formal or 'proper' or is what you learned in school. Scientists have discovered that many of the 'rules' taught in school are wrong anyway.

Focus on your gut instincts. Does the sentence sound correct, or does it sound like a mistake -- for instance, a mistake made by a young child or a recent immigrant?

Fill in the blank. Check **all** correct answers.

John ____ the library the book.

☑ gave

□ donated

□ distributed

□ contributed

Figure VII-6: instruction screen and sample item of the online language test *Games with words*

7.3.4.4 Measurement of automatized L2 knowledge

In the fourth study, we measured participants' knowledge of the L2 after they had practised grammar rules, and found that this knowledge was most probably explicit—even on tasks that involved some amount of time pressure and are often considered measures of implicit L2 knowledge (R. Ellis, 2005, 2009b)—since learners were largely aware of the fact that they were being tested on grammar. Consequently, we argued in section 7.2.5.1 that we may have measured 'automatized procedural knowledge' on the TGJT and OEIT, instead of implicit knowledge.

However, this argumentation may be questioned. First, we can only make inferences on the development of automaticity on the basis of the TGJT data (and the practice data; see further). Consistent with predictions made by cognitive psychology, in particular Skill Acquisition Theory (DeKeyser, 2008), measures of automaticity always involve response times, since automatic processes are considered to be rapid. Response times were only investigated in the analysis of the TGJT data, and not for the OEIT data, given the considerable effort it takes to measure response times in speech, let alone in complex spoken tasks. In this project, there was no scope for extracting response times from the speech data or other measures of fluency, such as filled pauses, repeated words, or sentence restarts. Automatic speech recognition technology may be of use here (Stouten, Duchateau, Martens, & Wambacq, 2006). So, pending the investigation of fluency rates in the OEIT data, we can only say that practice helped learners somewhat on the OEIT in terms of accuracy.

Secondly, it is difficult to say whether the fast response times on the post-tests of the TGJT provide support for the argumentation that learners had developed automatized procedural knowledge. First, the response times from the practice sessions still need to be analysed. If these drop steeply in the initial phases of practice, followed by a more gradual decrease over time, then there might be evidence for the so-called ‘power law of practice’ (DeKeyser, 2008), which testifies to the effect of practice on automatization.

But in addition to the fact that further data analysis is needed, there are more fundamental issues at stake. The first fundamental issue concerns the nature of automatization. The second concerns the nature of the practice tasks in this study.

As for the nature of automatization, a key debate in the literature concerns the question to what extent skill acquisition involves a mere quantitative change (i.e. speeding up) or also a qualitative change, namely restructuring of the underlying cognitive mechanism in carrying out a particular task (Segalowitz, 2005). To distinguish these two types of change, Segalowitz & Segalowitz (1993) have proposed the ‘coefficient of variation’, which comprises the standard deviation divided by the mean response time. In the case of qualitative change, this coefficient is supposed to decrease, while it should remain stable in the case of a mere speeding up. This merits further investigation.

The second fundamental issue is the nature of the practice tasks in the fourth study. As noted, the content of the practice items was drawn from a mystery story in order to focus the learners' attention on meaning during practice, and the practice task was a gamified version of a grammaticality judgment task, comprising CF and some amount of time pressure in order to stimulate automatization. However, this time pressure and the fact that no meaningful response was required during or after practice may have led learners to tune out of meaning. Performing grammaticality judgment tasks is considered to involve three subsequent processes, namely semantic processing, noticing, and reflecting (R. Ellis, 2004), but in this case, task design was such that semantic processing was not a necessary condition to complete the task. If semantic processing did not occur, or was severely under pressure, then learners were probably primarily engaged in behaviour that involved strengthening associations of form with form. This is highly likely for the practice items on quantifiers, which can be judged on the basis of surface grammatical features. If no focus on meaning was involved, then this would have led to the speeded-up processing of declarative knowledge, rather than automatization of procedural knowledge (i.e. knowledge involved in meaning-oriented language use). In other words, something may have been automatized, but not likely the kind of behaviour which is considered useful for the development of communicative ability.

Ultimately, although performance measures such as accuracy rates and response times can nowadays be measured relatively easily and in ecologically valid settings, exclusive reliance on such measures may not provide sufficient information on the types of knowledge that are assumed to be involved in task processes (e.g. automatized declarative knowledge vs. automatized procedural knowledge). If SLA researchers are willing to sacrifice ecological validity for the benefit of construct validity, then methods of brain research seem promising in this respect. Neuroimaging techniques can assist researchers in identifying the neurological systems on which learners rely while performing certain tasks, and in subsequently inferring the types of knowledge that are involved in L2 processing. Worthy of mention here are Ullman's declarative-procedural model of L2 learning, which adopts a non-interface position, as well as his empirical research using brain research techniques (see comments in Dörnyei 2009, pp. 161–162 and Lantolf & Poehner 2014, pp. 74–78).

7.3.5 L2 processes: intended meaning focus does not mean actual meaningful use

The issue of meaningful L2 use highlighted in the previous section brings us to another limitation of this research project. One of the main reasons for carrying out this project in language learning environments that are game-based was that such environments may catalyse learners' involvement in highly meaningful and situated L2 use. In the real world, when language users make poor linguistic choices, they often face bad consequences. This idea formed the basis of early designs in digital game-based language learning, such as *London Adventure* (Phillips, 1986) and a series of language learning games in the genre of 'interactive participatory drama' (Hubbard, 2002). In contrast with other contemporary popular media formats such as film, games typically involve individuals as characters in a story. In games that draw heavily on narrative, which may or may not be enacted through language, the choices that players make critically matter, as they shape the development of the characters which they identify with. If, then, these choices are made on the basis of information given in the form of aural or written language, then interaction with a language-intensive and story-driven game is likely to engage players in highly meaningful L2 processing. Not attending to linguistic input, such as 'click-through' behaviour in dialogues involving the player's character, will result in less favourable consequences for the player's character. And even if players seek out such dire consequences in games—for instance out of sheer curiosity what harm might befall their avatar—then they might want to know *why* this or that happened, which again is likely to prompt meaning-oriented L2 use.

Adopting this rationale, considerable effort in this research project was put into designing the L2 practice activities such that they would stimulate meaning-focused L2 use. The first study used an immersive 3D environment that revolved around making appropriate pragmatic choices in simulated dialogue tasks, which influenced how the non-player characters in the game reacted to the learner's choices. In the second study, learners interacted through the written word with an authentic murder mystery, namely E.A. Poe's *Murders in the Rue Morgue*, which was unveiled gradually as the learners were using language. The third study revolved around mini-games, and marked a shift towards language practice based on a more discrete-item approach. Since the focus of this study was not on L2 development, no serious attempt was made to make the language drills meaningful, other than the fact that the practice content was loosely related to a yet to be written story about the theft of the highly secret recipe of Coca-Cola. However, debriefing interviews held after practice revealed that learners were intrigued by the idea of such a background story. This idea was elaborated in the fourth study, which focused on L2 development supported by continued practice with mini-games. The researcher wrote a mystery text, which was rooted in the early history of the Coca-Cola Company, and in which the disappearance of its notorious recipe was intended as an 'inciting incident' to involve learners in the story. The episodes of this text were linked to the practice content, and the idea was that learners advanced in the story by practising their grammar via the mini-games. So, the latter design also involved an intended focus on meaning.

On the basis of our observations of learners and comments made by learners and teachers, it is fair to say that learners attended to meaning at least to some degree while they were practising their grammar skills. Yet, there is considerable room for improvement. The designs developed in the first and second study were perhaps most meaning-oriented of all four studies. However, if these designs are implemented in more longitudinal research designs, it is conceivable that learners will quickly notice that their choices do not meaningfully affect the development of their virtual character, which may lead to their abandonment of meaning focus and to more mechanical practice behaviour. From a conceptual point of view, providing meaningful feedback on learners' choices that are (pragma-)linguistically inappropriate makes great sense. However, the conceptualisation of such feedback and its implementation are extremely complex, time-intensive, and rather impracticable for educational technology projects, which typically have to cope with much smaller budgets than those of commercially focused game development projects. In the fourth study, embedding the drills in otherwise meaningful-oriented L2 use was a first step towards meaningful practice, but likely there was an alternation between meaningful reading and discussion on the one hand, and mechanical controlled practice on the other hand.

The bottom line for the engineering of digital game-based language learning spaces is that a designed meaning focus, intended by the instructional designer(s), does not necessarily imply that learners will actually process the language in meaningful ways. Task design needs to be such that learners cannot tune out of meaning, and that meaning and form focus go hand in hand. In the next section, possibilities will be suggested to accomplish this.

7.4 Directions for future research

In the previous sections, we summarized the results of the research project, which suggest that game-based practice with explicit CF can be quite powerful for the development of grammatical accuracy in a L2, and that this need not necessarily get in the way of playful immersion in experience. Further, we noted limitations of the project; here, we briefly reprise three key constraints that may inspire future research.

First, the studies largely failed to capture contextual information on individual L2 practice, since they were relatively short and focused on the group level rather than on individual learners. Secondly, no clear picture emerged of why CF was used, not used, or used in other ways than intended, and how the use of CF impacted on L2 development. And third, the effects of receptive practice in simple, strongly form-focused tasks did not seem to transfer well to spoken skills in more meaningful and complex tasks. It is conceivable that teachers will not spend much effort on game-based practice if it does not help their learners to develop knowledge that is useful for performance in more complex and productive, ideally spoken tasks in the L2. In other words, game-based practice is only likely to be adopted if it catalyses the development of communicative automaticity (accuracy *and* fluency) in a L2.

In this section, we provide directions for future research that take these findings and limitations into account. The central notion that will guide this discussion is that of *(learner) agency*—which is similar to a construct from Self-Determination Theory that was not addressed in this project, namely *autonomy* (Deci & Ryan, 2000). Not surprisingly, it seems to be that in this project—regardless of the fact that the designs of the learning and testing activities were intended to elicit particular forms of behaviour deemed favourable for learning—learners ultimately made their own choices, and did not always behave as intended. This applies both to how learners approached the tasks, and to how they interacted with the CF.

Cases in point include learners' behaviour in the practice and testing tasks in the fourth study, their use of CF in the second and fourth studies, and their rejection of vivid CF in the third study. In the fourth study, the learners approached the practice tasks in rather mechanical ways, despite the fact that the practice items were thematically related to a mystery story read for its meaning in class and hence created opportunities for meaningful L2 processing. Learners dealt in very similar ways with the oral production and role-playing task that was used as a disguised post-test of grammar subsequent to practice. Although this task did not only create opportunities for learners to attend to meaning, but also involved them in meaningful role-play and actually required them to process the stimuli for meaning, learners seemed to switch to and fro between meaning-focus and form-focus. In TBLT-speak, learners were “displaying rather than using language [in meaningful ways]” (R. Ellis, 2003, p. 8), or, as Jager (2009) phrased it so aptly, they were “‘regurgitating’ pre-selected expressions and grammatical structures” (p. 200) but not conveying *personal* meaning. Likely, they opted to do so because they unmasked the task as a grammar test. Further, learners also made their own choices when interacting with CF.

The second study presented some evidence that not all learners chose to work through the CF, but displayed ‘gaming the system’ (Baker et al., 2008) behaviour, i.e. they peeked at correct responses or repeatedly requested hints instead. Moreover, because we did not co-regulate individual learners when they interacted either with the optional CF or with correct responses and hints, we could not determine why learners engaged in such behaviour. The same applies to the fourth study, in which no qualitative data was collected on how or why learners attended (or did not attend) to the CF.

Finally, in the third study, the learners who were interviewed after practising with three versions of a gamified practice task, one of which included vivid CF, indicated that they would rather practise without vivid CF because it frustrated them. This was in contrast with the intention to design CF that was humorous and playful.

We could argue that these findings all testify to the central tenet of the Cognitive Mediational Paradigm (e.g. Butler & Winne, 1995), namely that learners make their own choices in learning and while interacting with externally provided feedback, regardless of whether the instructional design is intended to engage them in meaningful task processing. Large effect sizes for particular instructional cues (such as feedback), as reported in the literature, justify their inclusion in the instructional design, but do not guarantee that learners will always use such cues, let alone that they will invariably benefit from them.

Therefore, it is critical that future studies on the effectiveness of CF in digital game-based language learning explicitly allow room for individual learners to display agency in the engineered (i.e. determined by the designer) learning spaces. This may entail that researchers who operate within a mindset of experimental, interventionist research must be willing to loosen the reins somewhat, and to first and foremost focus on scaffolding and observing the individual learner rather than controlling, measuring, and comparing on the level of groups of individuals. We propose, then, that future designs must strive to enable agentive participation of learners on three levels: negotiation of meaning that is genuinely communicative, careful negotiation of linguistic form, and negotiation of gameful designs.

In the remainder of this section, we give a brief outline of two conceptual frameworks that may contribute to enabling these three types of participation. The first is an instructional design model known under the acronym ACCESS (Gatbonton & Segalowitz, 1988, 2005b), which targets the development of spoken automaticity in a L2 and forges controlled practice activities with principles of communicative language teaching (more particularly TBLT). The second is a conceptual framework on dynamically provided feedback derived from Sociocultural Theory of L2 learning (Lantolf & Thorne, 2006), which affords a more fine-grained investigation of how negotiated use of CF in controlled practice tasks may related to L2 development.

7.4.1 From alternation between meaning- and form-focus to negotiation of meaning, and form-focus skilfully embedded in genuinely communicative spoken L2 practice: the ACCESS instructional design model

In this project (especially in the fourth study), we observed that learners by and large displayed L2 processing behaviour that is deemed less favourable for L2 learning: they chose to tune out of meaning once they engaged with strictly form-focused practice activities, and even in complex tasks, learners managed to switch swiftly between exclusive meaning-focus and exclusive form-focus.

This reflects the most central design challenge of current-day communicative language teaching, i.e. the incorporation of explicit focus-on-form and controlled practice activities within an otherwise meaning-oriented approach to L2 instruction. In Diane Larsen-Freeman's (2003) experience, many teachers are pragmatic, and integrate both approaches into their teaching, as in our project. Still, she writes that a dichotomous thinking between teaching form (i.e. the parts of language, linguistic units, constructions) and function (i.e. the achievement of a non-linguistic purpose while using the parts of language) continues to predominate the field of language teaching at large, and that this dichotomy can be observed at the local level of the classroom:

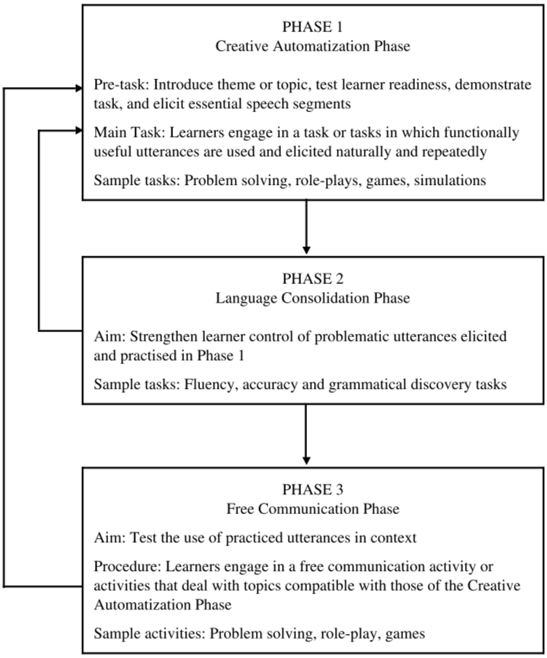
> We may include both foci—function and form—but we do not routinely integrate them. Typically, a teacher or a textbook will use both activities that are primarily communicatively focused and activities that primarily deal with the parts of language—yet these will occur in different lessons, or different parts of lessons, or in different parts of a textbook unit. In other words, even at the microlevel of a lesson, the two approaches remain segregated. (Larsen-Freeman, 2003, p. 7)

This segregation applies particularly to controlled practice. An essential feature of controlled practice—the epitome of "activities that primarily deal with the parts of language"—is the frequent repetition of particular linguistic units and constructions. Such repetition is key for L2 learning, given the evidence that L2 development is highly sensitive to frequency (N. C. Ellis, 2002). However, the problem on a practical level is that the openness which characterizes communicative language teaching resists repetition. Or, the other way round, the challenge for instructional design is how to make repetition truly communicative.

Gatbonton & Segalowitz (1988, 2005b) propose an instructional design model that is intended to bridge these two seemingly juxtaposed teaching approaches. Their model was developed against the backdrop of automaticity accounts of SLA, and is known under the acronym ACCESS, which stands for Automatization in Communicative Contexts of Essential Speech Segments. The model is communicative and exemplar-based: automatization concerns the retrieval and production of *essential speech segments* in genuinely communicative contexts, i.e. the automatization of formulaic language and *constructions* (conventionalized form-meaning mappings), rather than structures, patterns, or rules. The objective of practice is that learners will eventually produce such speech segments with greater accuracy and fluency.

A typical ACCESS lesson comprises three phases (see Figure VII-7), which differ in their degree of form focus, but which invariably involve the learner in expressing personal meanings, rather than in repeating prefabricated expressions. Moreover, there is scope for highly agentive participation of learners at all three stages of the lesson.

In the first phase, called *Creative Automatization*, learners are first introduced to the topic ('pre-task'), and then engage in a task that elicits the natural repetition of functionally useful utterances ('main task'), typically by way of problem solving tasks, role-plays, or games. R. Ellis (2003) would probably classify such activities as 'focused language tasks'. An example of such a task in ACCESS is the activity of taking a class photograph, in which students tell each other which pose to assume for the photo (Gatbonton & Segalowitz, 1988). This activity clearly comprises both elements that are needed for communicative automatization: the non-linguistic purpose of getting the class photo taken, and the inherent repetition involved in getting every student into position (triggering the use of location words, verbs of command, and modals, for instance). The phase is called 'Creative Automatization' because "students themselves generate (create) communicative intentions and produce the correspondingly appropriate utterances based on their understanding of the communicative situation" (Gatbonton & Segalowitz, 1988, p. 476). The teacher's role in this phase is twofold: in the pre-task, s/he diagnoses learners' initial knowledge and provides them with model speech segments on an as-needed basis; in the main-task, s/he elicits productive and repetitive use of these segments in genuinely communicative interaction.



PHASE 1
Creative Automatization Phase

Pre-task: Introduce theme or topic, test learner readiness, demonstrate task, and elicit essential speech segments

Main Task: Learners engage in a task or tasks in which functionally useful utterances are used and elicited naturally and repeatedly

Sample tasks: Problem solving, role-plays, games, simulations

PHASE 2
Language Consolidation Phase

Aim: Strengthen learner control of problematic utterances elicited and practised in Phase 1

Sample tasks: Fluency, accuracy and grammatical discovery tasks

PHASE 3
Free Communication Phase

Aim: Test the use of practiced utterances in context

Procedure: Learners engage in a free communication activity or activities that deal with topics compatible with those of the Creative Automatization Phase

Sample activities: Problem solving, role-play, games

Figure VII-7: The three phases of the ACCESS instructional design model (from Gatbonton & Segalowitz, 2005)

In ACCESS, the Creative Automatization phase is optionally followed by a phase called *Language Consolidation*, in which the teacher can give more formal focus to specific utterances that may have caused problems in the earlier phase, including the provision of declarative information. This phase thus aims to strengthen learners' control of the previously used utterances in terms of fluency, accuracy, and grammatical knowledge. Moreover, it creates scope for the use of more constrained and explicit language drills. However, also here, meaning focus is a key requirement, and the activity must connect content-wise to the previous, more communicative phase.

An ACCESS lesson concludes with a phase of *Free Communication*, in which learners are encouraged to express their personal meanings on the topic of the lesson more freely. In this phase, it is not critical that learners use the target utterances, although such use may occur naturally because there is a link with the previous phases in terms of the topic.

In light of the findings and limitations of this research project highlighted earlier in this section, the key affordance of ACCESS is that meaning and form focus are tightly integrated, and that it allows for involving the learner in highly agentive participation throughout the three stages of ACCESS lessons. The fact that in ACCESS, meaning- and form-focus go hand in hand is consistent with theoretical views on the integration of synthesis and analysis formulated by e.g. Skehan, DeKeyser, and Ortega (for discussion see Van den Branden, 2007), and with the view that focus on form works best when the distance with authentic L2 use is as short as possible: "optimal practice in the foreign language classroom should be interactive, truly meaningful, and with a built-in focus on selective aspects of the language code that are integral to the very nature of that practice" (Ortega, 2007, p. 184).

Finally, learners' agentive participation in ACCESS lessons can be heightened through the integration of games in the three phases of L2 practice, from strongly immersive and meaning-focused games to more focused mini-games—provided that the latter involve the learner in meaningful L2 processing. Moreover, the instructional design model is cyclical, allowing teachers and learners to collaboratively evaluate how practice went in class, and to inform future iterations through short loops of design, implementation and evaluation. Thus, learners, teachers, and designers can negotiate the gameful designs of the instructional environment. Key in this respect is the free communication phase, which may enable the engagement of learners in meaningful, negotiated, and collaborative remixing of the texts and designs offered previously in game-based practice, consistent with current societal trends in the adoption of games—a particularly notable example of which is *fan fiction* (R. W. Black, 2009; Thorne, Black, & Sykes, 2009)—which are "evoking a shift away from models of learning based on information delivery toward theories of human development rooted in experiential problem solving and complex and spatially distributed forms of collaboration" (Cornillie, Thorne, et al., 2012, p. 245).

7.4.2 From CF exposure to use of CF as negotiation of form: Sociocultural Theory and Dynamic Assessment

A second notable limitation of this project, as noted above, was that no clear picture was established of learners' interaction with CF, particularly with respect to the different ways of dealing with (different types of) CF, and the reasons for doing so. We also lacked qualitative measures of CF use. In other words, learners were *exposed* to various CF types, but we did not measure well whether they truly interacted with the CF, when they paid attention to it, and, if they engaged with CF, *why* they did so. As a result, we were unable to determine how use of CF related to L2 development. This, again, relates to learner agency: learners do not blindly accept externally provided feedback, but seek help at specific stages of development, for specific reasons, and for specific linguistic phenomena. A framework that may help to comprehend learners' agentive use of CF better is Sociocultural Theory (SCT) (Lantolf & Thorne, 2008). SCT is a theory of L2 development that originates in the work of the psychologist Vygotsky, and which pays close attention to the interaction of learning individuals with their socio-cultural surroundings, of which externally provided CF is one particular aspect.

As a starting point, SCT sees human beings as inextricably embedded in their environment. Mental functioning and development (including learning a L2) are considered processes that capitalize on the human's *mediation* with his or her environment, in which "language use, organization, and structure are the primary means of mediation" (Lantolf & Thorne, 2008, p. 201). One such form of mediation is *regulation*, which refers to how the individual's behaviour is defined with respect to his/her environment, and how it changes over time. L2 development, then, is seen as the gradual (but unstable) process of a learner moving from more passive regulation by environmental artefacts (both physical and symbolic) and other humans (peers, parents, teachers, etc.) to complete self-regulation, defined as "the ability to accomplish activities with minimal or no external support" (Lantolf & Thorne, 2008, p. 204).

Since SCT attributes significant importance to the individual's mediation with his or her environment, the theory does not exclusively see L2 development as a 'result' that can be separated from the learner's interaction with a specific instructional environment. Rather, SCT focuses on observing development *in* the learner's interaction with the L2 environment itself. A central notion here is the *zone of proximal development* (ZPD), which is defined most accurately as "the distance between the actual developmental level as determined by independent problem solving and the level of potential development as determined through problem solving under guidance or in collaboration with more capable peers" (Vygotsky, 1978, as cited in Lantolf & Thorne, 2006, p. 266). As may be gathered, SCT sees CF as an instance of such external support or guidance.

From the notion of the ZPD, then, it follows that the amount and type of CF on the basis of which a learner can independently solve a linguistic problem are indicative of the learner's L2 development. A learner who is able to solve a problem with only minimal/implicit CF is considered more advanced than a learner who needs more detailed and explicit CF. At the end of the scale is the learner who needs to be given the correct form. Hence, the learner's interaction with and use of (output-prompting) CF become a primary lens through which L2 development can be observed. It stands to reason that any investigation of L2 development supported by CF has to take into account whether and how L2 learners use CF—indeed, "for assessment to be formative the feedback information has to be used" (P. Black & Wiliam, 1998, p. 16)—but the rationale proposed by SCT is particular: it argues that use of CF is not an antecedent, but also an actual measure of development. This idea has been meticulously operationalized in the instructional procedure known as *Dynamic Assessment*

(DA), and regulatory scales have been developed with which L2 development can be measured on the basis of the learners' use of different types of CF, ranging from implicit to explicit (e.g. Aljaafreh & Lantolf, 1994).

The implementation of such a graduated CF scale in technology-mediated L2 learning, particularly in instructed (tutorial) CALL environments, seems highly promising for research into the effectiveness of CF (Lantolf & Thorne, 2006, pp. 331–335). In this project, we chose not to apply this framework to tutorial CALL tools because its implementation requires serious technological efforts, because DA requires foreseeing (virtually) every possible response from the learner, and because we focused on practice at the classroom level—our resources were limited so that we could not co-regulate individual learners as they practised in class. Yet, both study 2 and study 4 showed that a closer look of learners' use of CF is required in order to come to a better understanding of which particular types of CF aid learning and how, at which particular stages of development, and for which particular linguistic phenomena.

Future studies into learners' use of graduated prompts, then, may yield a closer and more dynamic look into the effectiveness of CF than interventionist studies in CALL that adopt an experimental (between- or within-subjects) design, in which different CF types are included in different experimental conditions. More specifically, implementations of computerized DA in tutorial CALL may provide a more comprehensive view of how use of CF over time relates to more commonly used measures of L2 development, such as accuracy, fluency, or grammatical knowledge. Moreover, Dynamic Assessment may help to circumvent an issue that we experienced in our fourth study, namely that, on the oral production task, learners reverted to test-taking behaviour and to less favourable types of form-meaning-focus switching in what could otherwise have been a meaningful task. Because in DA, learning and assessment are one, such test-taking behaviour may not occur.

Moreover, graduated CF fits in well with game-based, constructivist approaches to language learning (Purushotma et al., 2008). In the graduated prompt approach, the learner is in control of the level of required detail of CF on each occasion where there is a need for external support, and also controls the degree to which attention to linguistic form interferes with meaning focus. For instance, imagine that a young language learner is playing a mini-game on the topic "giving directions". The game involves a worm, a terrain that is in a constant flux of irrigation and drying up, and blackbirds. The objective of the activity is to guide the worm across the terrain from point A to point B by producing spoken utterances, such as "Now you need to go left.", "Move to the right, quickly!", or "Duck!". The activity involves repetition, but learners also need to be creative and focused on the non-linguistic purpose, for they choose the path to follow, and produce the appropriate utterances that correspond to their non-linguistic, meaning-focused intentions.

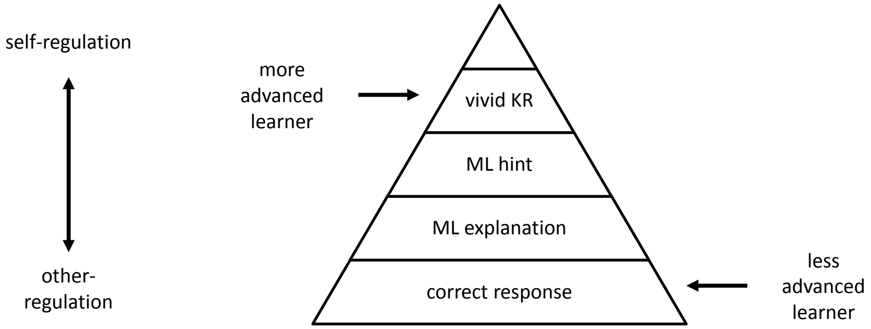


Figure VII-8: example of a graduated CF scale in game-based CALL

If the learner utters the wrong directive or is too slow, the game informs the learners of the result in a way that is consistent with the ontology of the game—in line with the concept described above, the worm suffers from temporary dehydration, or may be eaten by the blackbird. Such feedback comprises the first level in the graduated CF scale, and may be best described as vivid 'knowledge of results' (KR) (see Figure VII-8). It provides minimal learning support, may engage the learner by showing in vivid ways the result of linguistic production in terms of the world represented in the game, and the focus remains on purposeful use of the target language that results in non-linguistic outcomes. If the learner fails time and time again in producing a particular construction, s/he may click on a button to retrieve additional learning support, from metalinguistic and/or lexical hints and explanations to model responses. Hence, graduated CF runs few risks of reducing the learner's agentive capacity and the activity's focus on meaningful and purposeful interaction, while providing maximal learning support.

7.5 Back to the mystery of the blue whale

Our quest for knowledge of the effectiveness of corrective feedback in digital game-based language learning ends here, for now. Let's go back, then, to our learner hero introduced in the first chapter of this dissertation.

Our hero was immersed in the gameful experience of a mystery set in London's Natural History Museum. He was on the quest of explaining why the belly of the blue whale replica, displayed in the museum, had been broken into. He was probably on the point of navigating his player character to the Large Mammals Hall, in search of fingerprints or other pieces of evidence near the whale. But as educators, we expected our hero to use language in order to collect clues about the case and solve it. We also expected our learner hero to make mistakes in the target language, which he did. But we were concerned that focusing on linguistic mistakes and giving a large volume of corrective feedback might interfere with his meaningful and goal-directed use of language. We were not sure that our learner would attend to the feedback. We were not sure whether the red ink would actually benefit his engagement and development in the longer term.

Knowing what we know from this research project, and given limitless technological possibilities, we would advise for the use of red ink at various levels of the adventure of learning a language. As long as our learner hero realizes it is perfectly all right to make mistakes, and is sufficiently supported both in terms of cognition and motivation so that he can improve, corrective feedback provided in digital game-based spaces is bound to make for powerful language learning.

References

Abramson, L. Y., Seligman, M. E., & Teasdale, J. D. (1978). Learned helplessness in humans: critique and reformulation. *Journal of Abnormal Psychology*, *87*(1), 49–74. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/649856

Aldrich, C. (2005). *Learning by doing : a comprehensive guide to simulations, computer games, and pedagogy in e-learning and other educational experiences*. San Francisco: Pfeiffer.

Aleven, V., Stahl, E., Schworm, S., Fischer, F., & Wallace, R. (2003). Help Seeking and Help Design in Interactive Learning Environments. *Review of Educational Research*, *73*(3), 277–320.

Aljaafreh, A., & Lantolf, J. P. (1994). Negative Feedback as Regulation and Second Language Learning in the Zone of Proximal Development. *Modern Language Journal*, *78*(4), 465–483.

Anderson, J. R. (1992). Automaticity and the ACT* theory. *American Journal of Psychology*, *105*(2), 165–180.

Anderson, J. R., Bothell, D., Byrne, M. D., Douglass, S., Lebiere, C., & Qin, Y. (2004). An integrated theory of the mind. *Psychological Review*, *111*(4), 1036–60. doi:10.1037/0033-295X.111.4.1036

Anttila, A., Adams, M., & Speriosu, M. (2010). The role of prosody in the English dative alternation. *Language and Cognitive Processes*, *25*(7-9), 946–981. doi:10.1080/01690960903525481

Baker, R., Walonoski, J., Heffernan, N., Roll, I., Corbett, A., & Koedinger, K. (2008). Why Students Engage in “Gaming the System” Behavior in Interactive Learning Environments. *Journal of Interactive Learning Research*, *19*(2), 185–224.

Baltra, A. (1990). Language Learning through Computer Adventure Games. *Simulation & Gaming*, *21*(4), 445–452.

Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York: Freeman.

Becker, K. (2007). Pedagogy in commercial video games. In M. Prensky, C. Aldrich, & D. Gibson (Eds.), *Games and simulations in online learning : research and development frameworks* (pp. 21–47). Hershey: Information science.

Bedwell, W. L., Pavlas, D., Heyne, K., Lazzara, E. H., & Salas, E. (2012). Toward a Taxonomy Linking Game Attributes to Learning: An Empirical Study. *Simulation & Gaming*, *43*(6), 729–760. doi:10.1177/1046878112439444

Birdsong, D. (1989). *Metalinguistic Performance and Interlinguistic Competence*. Berlin: Springer-V.

Black, P., & Wiliam, D. (1998). Assessment and Classroom Learning. *Assessment in Education: Principles, Policy & Practice*, *5*(1), 7–74. doi:10.1080/0969595980050102

Black, R. W. (2009). Online Fan Fiction, Global Identities, and Imagination. *Research in the Teaching of English*, *43*(4), 397–425.

Boero, R., & Novarese, M. (2012). Feedback and Learning. In (N. Seel, Ed.)*Encyclopedia of the Sciences of Learning*. Springer.

Bogost, I. (2011). *How to Do Things with Videogames*. Minneapolis: University Of Minnesota Press.

Brandl, K. K. (1995). Strong and Weak Students' Preferences for Error Feedback Options and Responses. *The Modern Language Journal*, *79*(2), 194–211.

Bransford, J. D., Brown, A. L., & Cocking, R. R. (2000). *How people learn: Brain, mind, experience, and school*. Washington, DC: National Academy Press.

Bullard, N. (1990). Briefing and debriefing. In D. Crookall & R. L. Oxford (Eds.), *Simulation, gaming and language learning* (pp. 55–66). New York: Newbury House.

Butler, D. L., & Winne, P. H. (1995). Feedback and Self-Regulated Learning: A Theoretical Synthesis. *Review of Educational Research*, *65*(3), 245–281. doi:10.2307/1170684

Butler, R. (1987). Task-involving and ego-involving properties of evaluation: Effects of different feedback conditions on motivational perceptions, interest, and performance. *Journal of Educational Psychology*, *79*(4), 474–482. doi:10.1037//0022-0663.79.4.474

Cannon, M. D., & Edmondson, A. C. (2012). Learning from failure. In (N. Seel, Ed.)*Encyclopedia of the Sciences of Learning*. New York: Springer.

Carroll, S. E. (1995). On the irrelevance of verbal feedback to language learning. In L. Eubank (Ed.), *The current state of interlanguage : studies in honor of William E. Rutherford* (pp. 73–88). Amsterdam: John Benjamins.

Carroll, S. E. (2001). *Input and Evidence. The raw material of second language acquisition*. Amsterdam: John Benjamins.

Carroll, S. E., & Swain, M. (1993). Explicit and implicit negative feedback. An empirical study of the learning of linguistic generalizations. *Studies in Second Language Acquisition*, *15*(3), 357–386.

Cathcart, R., & Olsen, J. (1976). Teachers' and students' preferences for correction of classroom conversation errors. In J. E. Fanselow & R. Crymes (Eds.), *On TESOL '76* (pp. 41–53). Washington, DC: TESOL.

Chapelle, C. (1998). Multimedia call: lessons to be learned from research on instructed SLA. *Language Learning & Technology*, *2*(1), 21–36.

Chenoweth, N. A., Day, R. R., Chun, A. E., & Luppescu, S. (1983). Attitudes and Preferences of ESL Students to Error Correction. *Studies in Second Language Acquisition*, *6*(1), 79–87.

Clarebout, G., & Elen, J. (2006). Tool use in computer-based learning environments: towards a research framework. *Computers in Human Behavior*, *22*(3), 389–411. doi:10.1016/j.chb.2004.09.007

Clarebout, G., & Elen, J. (2009). The complexity of tool use in computer-based learning environments. *Instructional Science*, *37*(5), 475–486. doi:10.1007/s11251-008-9068-3

Cobb, T., & Horst, M. (2011). Does Word Coach Coach Words ? *CALICO Journal*, *28*(3), 639–661.

Colpaert, J. (2004). *Design of online interactive language courseware: Conceptualization, specification and prototyping. Research into the impact of linguistic-didactic functionality on software architecture*. University of Antwerp, Antwerp.

Colpaert, J. (2010). Elicitation of language learners' personal goals as design concepts. *Innovation in Language Learning and Teaching*, *4*(3), 259–274. doi:10.1080/17501229.2010.513447

Cornillie, F., Clarebout, G., & Desmet, P. (2012). Between learning and playing? Exploring learners' perceptions of corrective feedback in an immersive game for English pragmatics. *ReCALL*, *24*(3), 257–278. doi:10.1017/S0958344012000146

Cornillie, F., & Desmet, P. (n.d.). Mini-games for language learning. In L. Murray & F. Farr (Eds.), *Routledge Handbook of Language Learning and Technology*.

Cornillie, F., Thorne, S. L., & Desmet, P. (2012). Digital games for language learning: from hype to insight? *ReCALL*, *24*(3), 243–256. doi:10.1017/S0958344012000134

Council of Europe. (2011). *Common European Framework of Reference for Languages: Learning, Teaching, Assessment (CEFR)*. Retrieved from http://www.coe.int/t/dg4/linguistic/Source/Framework_EN.pdf

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, *16*(3), 297–334.

Csikszentmihalyi, M. (1990). *Flow: The psychology of optimal experience*. New York: Harper Perennial.

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, *13*(3), 319–340.

De Cuypere, L., & Buysse, M. (n.d.). A corpus-based study of the dative alternation in spoken British English. *Journal of English Linguistics*.

De Grove, F., Cornillie, F., Mechant, P., & Van Looy, J. (2013). Tapping into the field of foreign language learning games. *International Journal of Arts and Technology*, *6*(1), 22–43. doi:10.1504/IJART.2013.050690

De Grove, F., Van Looy, J., & Courtois, C. (2010). Towards a Serious Game Experience Model: Validation, Extension and Adaptation of the GEQ for Use in an Educational Context. In L. Calvi, K. C. M. Nuijten, & H. Bouwknegt (Eds.), *Playability and player experience – Proceedings of the Fun and Games 2010 Workshop* (pp. 47–61). Breda: Breda University of Applied Sciences.

De Jong, T. (2005). The Guided Discovery Principle in Multimedia Learning. In R. E. Mayer (Ed.), *The Cambridge Handbook of Multimedia Learning* (pp. 215–228). New York: Cambridge University Press.

De Vries, B. P., Cucchiarini, C., Bodnar, S., Strik, H., & van Hout, R. (2014). Spoken grammar practice and feedback in an ASR-based CALL system. *Computer Assisted Language Learning*, (June 2014), 1–27. doi:10.1080/09588221.2014.889713

Deci, E. L., Koestner, R., & Ryan, R. M. (1999). A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation. *Psychological Bulletin*, *125*(6), 627–668. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10589297

Deci, E. L., & Ryan, R. M. (2000). The "What" and "Why" of Goal Pursuits : Human Needs and the Self-Determination of Behavior. *Psychologial Inquiry*, *11*(4), 227–268.

deHaan, J. (2005). Learning Language through Video Games: A Theoretical Framework, An Evaluation of Game Genres and Questions for Future Research. In S. P. Schaffer & M. L. Price (Eds.), *Interactive convergence: critical issues in multimedia* (pp. 229–239). Oxford: Inter-Disciplinary Press. Retrieved from http://www.inter-disciplinary.net/publishing/id-press/ebooks/interactive-convergence-critical-issues-in-multimedia/

deHaan, J. (2008). *Video games and second language acquisition: The effect of interactivity with a rhythm video game on second language vocabulary recall, cognitive load, and telepresence*. New York University.

deHaan, J., Reed, W. M., & Kuwada, K. (2010). The effect of interactivity with a music video game on second language vocabulary recall. *Language Learning & Technology*, *14*(2), 74–94.

DeKeyser, R. M. (1993). The Effect of Error Correction on L2 Grammar Knowledge and Oral Proficiency. *The Modern Language Journal*, *77*(4), 501–514.

DeKeyser, R. M. (1997). Beyond Explicit Rule Learning: Automatizing Second Language Morphosyntax. *Studies in Second Language Acquisition*, *19*(2), 195–221.

DeKeyser, R. M. (1998). Beyond focus on form. Cognitive perspectives on learning and practicing second language grammar. In C. Doughty & J.

Williams (Eds.), *Focus on Form in Classroom Second Language Acquisition* (pp. 42–63). Cambridge: Cambridge University Press.

DeKeyser, R. M. (2001). Automaticity and automatization. In P. Robinson (Ed.), *Cognition and Second Language Instruction* (pp. 125–151). New York: Cambridge University Press.

DeKeyser, R. M. (2005). Implicit and Explicit Learning. In C. J. Doughty & M. H. Long (Eds.), *The Handbook of Second Language Acquisition* (pp. 313–348). Oxford: Blackwell Publishing Ltd.

DeKeyser, R. M. (2007a). Conclusion: The future of practice. In R. M. DeKeyser (Ed.), *Practice in a Second Language: Perspectives from Applied Linguistics and Cognitive Psychology* (pp. 287–304). New York: Cambridge University Press.

DeKeyser, R. M. (2007b). Introduction: Situating the concept of practice. In R. M. DeKeyser (Ed.), *Practice in a Second Language: Perspectives from Applied Linguistics and Cognitive Psychology* (pp. 1–18). New York: Cambridge University Press.

DeKeyser, R. M. (2008). Skill Acquisition Theory. In B. VanPatten & J. Williams (Eds.), *Theories in Second Language Acquisition* (pp. 97–114). New York: Routledge.

Desmet, P. (2007). L'apport des TIC à la mise en place d'un dispositif d'apprentissage des langues centré sur l'apprenant. *ITL - International Journal of Applied Linguistics*, *154*, 91–110.

Deterding, S., Dixon, D., Khaled, R., & Nacke, L. (2011). From Game Design Elements to Gamefulness : Defining " Gamification ." In *Mindtrek 2011 Proceedings*. Tampere: ACM Press.

Dörnyei, Z. (2003a). Attitudes, Orientations, and Motivations in Language Learning: Advances in Theory, Research, and Applications. *Language Learning*, *53*(May 2003 Supplement 2), 3–32.

Dörnyei, Z. (2003b). *Questionnaires in second language research: Construction, administration, and processing*. Mahwah, New Jersey: Lawrence Erlbaum.

Dörnyei, Z. (2005). *The psychology of the language learner: Individual differences in second language acquisition*. Mahwah, New Jersey: Lawrence Erlbaum.

Dörnyei, Z. (2009). *The Psychology of Second Language Acquisition*. Oxford: Oxford University Press.

Doughty, C., & Williams, J. (Eds.). (1998). *Focus on form in classroom second language acquisition*. Cambridge: Cambridge University Press.

Ehsani, F., & Knodt, E. (1998). Speech technology in computer-aided language learning: strengths and limitations of a new CALL paradigm. *Language Learning & Technology*, *2*(1), 54–73.

Elliot, A. J. (1999). Approach and Avoidance Motivation and Achievement Goals. *Educational Psychologist*, *34*(3), 169–189.

Elliot, A. J., Murayama, K., & Pekrun, R. (2011). A 3 × 2 achievement goal model. *Journal of Educational Psychology*, *103*(3), 632– 648. doi:10.1037/a0023952

Ellis, N. C. (1997). Implicit and Explicit Language Learning - An Overview. In N. C. Ellis (Ed.), *Implicit and Explicit Learning of Languages* (2nd ed., pp. 1–31). San Diego: Academic Press.

Ellis, N. C. (2002). Frequency effects in language processing. A review with implications for theories of implicit and explicit language acquisition. *Studies in Second Language Acquisition*, *24*(2), 143–188.

Ellis, N. C. (2005a). At the Interface: Dynamic Interactions of Explicit and Implicit Language Knowledge. *Studies in Second Language Acquisition*, *27*(20), 305–352.

Ellis, N. C. (2005b). Constructions, Chunking, and Connectionism: The Emerge of Second Language Structure. In C. J. Doughty & M. H. Long (Eds.), *The Handbook of Second Language Acquisition* (pp. 63–103). Oxford: Blackwell Publishing Ltd.

Ellis, N. C., & Larsen-Freeman, D. (2006). Language Emergence: Implications for Applied Linguistics--Introduction to the Special Issue. *Applied Linguistics*, *27*(4), 558–589. doi:10.1093/applin/aml028

Ellis, R. (2003). *Task-based language learning and teaching*. Oxford: Oxford University Press.

Ellis, R. (2004). The Definition and Measurement of L2 Explicit Knowledge. *Language Learning*, *54*(2), 227–275.

Ellis, R. (2005). Measuring implicit and explicit knowledge of a second language: A Psychometric Study. *Studies in Second Language Acquisition*, *27*(2), 141–172. doi:10.1017/S0272263105050096

Ellis, R. (2009a). Implicit and Explicit Learning, Knowledge and Instruction. In R. Ellis, S. Loewen, C. Elder, R. Erlam, J. Philp, & H. Reinders (Eds.), *Implicit and Explicit Knowledge in Second Language Learning, Testing and Teaching* (pp. 3–29). Bristol: Multilingual Matters.

Ellis, R. (2009b). Measuring Implicit and Explicit Knowledge of a Second Language. In R. Ellis, S. Loewen, C. Elder, R. Erlam, J. Philp, & H. Reinders (Eds.), *Implicit and Explicit Knowledge in Second Language Learning, Testing and Teaching* (pp. 31–64). Bristol: Multilingual Matters.

Ellis, R., Loewen, S., Elder, C., Erlam, R., Philp, J., & Reinders, H. (2009). *Implicit and explicit knowledge in second language learning, testing and teaching*. Clevedon: Multilingual Matters.

Ellis, R., Loewen, S., & Erlam, R. (2006). Implicit and Explicit Corrective Feedback and the Acquisition of L2 Grammar. *Studies in Second Language Acquisition*, *28*(2), 339–368.

Enginarlar, H. (1993). Student response to teacher feedback in EFL writing. *System*, *21*(2), 193–204.

Engwall, O., & Bälter, O. (2007). Pronunciation feedback from real and virtual language teachers. *Computer-Assisted Language Learning*, *20*(3), 235–262.

Erlam, R. (2009). The Elicited Oral Imitation Test as a Measure of Implicit Knowledge. In R. Ellis, S. Loewen, C. Elder, R. Erlam, J. Philp, & H. Reinders (Eds.), *Implicit and Explicit Knowledge in Second Language Learning, Testing and Teaching2* (pp. 65–93). Bristol: Multilingual Matters.

Farrar, M. J. (1992). Negative Evidence and Grammatical Morpheme Acquisition. *Developmental Psychology*, *28*(1), 90–98.

Fischer, R. (2007). How do we know what students are actually doing? Monitoring students' behavior in CALL. *Computer Assisted Language Learning*, *20*(5), 409–442. doi:10.1080/09588220701746013

Fishbein, M., & Ajzen, I. (1975). *Belief, Attitude, Intention, and Behavior: An Introduction to Theory and Research*. Reading, MA: Addison-Wesley.

Fukuya, Y. J., & Zhang, Y. (2002). Effects of recasts on EFL learners's acquisition of pragmalinguistic conventions of request. *Second Language Studies*, *21*(1), 1–47.

García-Carbonell, A., Rising, B., Montero, B., & Watts, F. (2001). Simulation/Gaming and the Acquisition of Communicative Competence in Another Language. *Simulation & Gaming*, *32*(4), 481–491.

Gass, S. M., & Mackey, A. (2008). Input, Interaction and Output in Second Language Acquisition. In B. VanPatten & J. Williams (Eds.), *Theories in Second Language Acquisition* (pp. 175–200). New York: Routledge.

Gatbonton, E., & Segalowitz, N. (1988). Creative Automatization: Principles for Promoting Fluency within a Communicative Framework. *TESOL Quarterly*, *22*(3), 473–492.

Gatbonton, E., & Segalowitz, N. (2005a). Rethinking Communicative Language Teaching: A Focus on Access to Fluency. *The Canadian Modern Language Review*, *61*(3), 325–353.

Gatbonton, E., & Segalowitz, N. (2005b). Rethinking Communicative Language Teaching: A Focus on Access to Fluency. *The Canadian Modern Language Review / La Revue Canadienne Des Langues Vivantes*, *61*(3), 325–353. doi:10.1353/cml.2005.0016

Gee, J. P. (2003). *What video games have to teach us about learning and literacy.* New York: Palgrave Macmillan.

Gee, J. P. (2007). *Good Video Games and Good Learning: Collected Essays*. New York: Peter Lang.

Goo, J., & Mackey, A. (2013). The Case Against the Case Against Recasts. *Studies in Second Language Acquisition*, *35*(1), 127–165. doi:10.1017/S0272263112000708

Gould, J. D., & Lewis, C. (1985). Designing for Usability : Key Principles and What Designers Think. *Communications of the ACM*, *28*(3), 300–311.

Graaff, R. de, & Housen, A. (2009). Investigating the Effects and Effectiveness of L2 Instruction. In M. H. Long & C. J. Doughty (Eds.), *The Handbook of Language Teaching* (pp. 726–755). Oxford: Wiley-Blackwell.

Gregg, K. R. (2001). Learnability and second language acquisition theory. In P. Robinson (Ed.), *Cognition and Second Language Instruction* (pp. 152–180). New York: Cambridge University Press.

Gropen, J., Pinker, S., Hollander, M., Goldberg, R., & Wilson, R. (1989). The Learnability and Acquisition of the Dative Alternation in English. *Language*, *65*(2), 203–257.

Gutiérrez, X. (2013). The construct validity of grammaticality judgment tests as measures of implicit and explicit knowledge. *Studies in Second Language Acquisition*, *35*(3), 423–449.

Habgood, M. P. J., & Ainsworth, S. E. (2011). Motivating Children to Learn Effectively : Exploring the Value of Intrinsic Integration in Educational Games. *Journal of the Learning Sciences*, *20*(2), 169–206.

Hattie, J. (2009). *Visible Learning: A Synthesis of Over 800 Meta-Analyses Relating to Achievement*. New York: Routledge.

Hattie, J., & Timperley, H. (2007). The Power of Feedback. *Review of Educational Research*, *77*(1), 81–112. doi:10.3102/003465430298487

Hattie, J., & Yates, G. (2014). *Visible Learning and the Science of How We Learn*. Abingdon: Routledge.

Havranek, G. (2002). When is corrective feedback most likely to succeed? *International Journal of Educational Research*, *37*(3-4), 255–270. doi:10.1016/S0883-0355(03)00004-1

Havranek, G., & Cesnik, H. (2001). Factors affecting the success of corrective feedback. *EUROSLA Yearbook*, *1*(1), 99–122.

Hedgcock, J., & Lefkowitz, N. (1994). Feedback on Feedback: Assessing Learning Receptivity to Teacher Response in L2 Composing. *Journal of Second Language Writing*, *3*(2), 141–163.

Heift, T. (2001). Error-specific and individualised feedback in a Web-based language tutoring system: Do they read it? *ReCALL*, *13*(01), 99–109. doi:10.1017/S095834400100091X

Heift, T. (2002). Learner Control and Error Correction in ICALL: Browsers, Peekers, and Adamants. *CALICO Journal*, *19*(2), 295–313.

Heift, T. (2004). Corrective feedback and learner uptake in CALL. *ReCALL*, *16*(02), 416–431. doi:10.1017/S0958344004001120

Heift, T. (2006). Context-sensitive Help in CALL. *Computer Assisted Language Learning*, *19*(2-3), 243–259. doi:10.1080/09588220600821552

Hémard, D. (2003). Language Learning Online: designing towards user acceptability. In U. Felix (Ed.), *Language Learning Online: towards best practice* (pp. 21–43). Lisse: Swets and Zeitlinger.

Hubbard, P. (1991). Evaluating computer games for language learning. *Simulation & Gaming*, *22*(2), 220–223.

Hubbard, P. (2002). Interactive Participatory Dramas for Language Learning. *Simulation & Gaming*, *33*(2), 210–216. Retrieved from http://sag.sagepub.com/cgi/reprint/33/2/210

Hubbard, P., & Bradin Siskin, C. (2004). Another look at tutorial CALL. *ReCALL*, *16*(2), 448–461.

Hulstijn, J. H. (2002). Towards a unified account of the representation, processing and acquisition of second language knowledge. *Second Language Research*, *18*(3), 193–223. doi:10.1191/0267658302sr207oa

Hulstijn, J. H. (2005). Theoretical and empirical issues in the study of implicit and explicit second-language learning. Introduction. *Studies in Second Language Acquisition*, *27*(2), 129–140. doi:10.1017/S0272263105050084

Hulstijn, J. H. (2007). Psycholinguistic perspectives on language and its acquisition. In J. Cummins & C. Davison (Eds.), *International Handbook of English Language Teaching* (Vol. II, pp. 783–796). Norwell, MA: Springer.

Hulstijn, J. H., & de Graaff, R. (1994). Under what conditions does explicit knowledge of a second language facilitate the acquisition of implicit knowledge? A research proposal. *AILA Review*, *11*, 97–112.

Ijsselsteijn, W., Hoogen, W. van den, Klimmt, C., Kort, Y. de, Lindley, C., Mathiak, K., … Vorderer, P. (2008). Measuring the Experience of Digital Game Enjoyment. In A. J. Spink, M. R. Ballintijn, N. D. Bogers, F. Grieco, L. W. S. Loijens, L. P. J. J. Noldus, … P. H. Zimmerman (Eds.), *Proceedings of Measuring Behavior 2008* (pp. 88–89). Maastricht, The Netherlands.

Jager, S. (2009). *Towards ICT-integrated Language Learning. Developing an Implementation Framework in terms of Pedagogy, Technology and Environment*. Rijksuniversiteit Groningen.

Johnson, W. L. (2007). Serious Use of a Serious Game for Language Learning. In R. Luckin, K. R. Koedinger, & J. Greer (Eds.), *Proceeding of the 2007 conference on Artificial Intelligence in Education: Building Technology Rich Learning Contexts That Work* (pp. 67–74). Amsterdam, The Netherlands: IOS Press.

Jordan, G. (1992). Exploiting Computer-Based Simulations for Language-Learning Purposes. *Simulation & Gaming*, *23*(1), 88–98.

Juul, J. (2010). *A casual revolution: reinventing video games and their players*. Cambridge, MA: MIT Press.

Juul, J. (2013). *The Art of Failure. An Essay on the Pain of Playing Video Games*. Cambridge, MA: MIT Press.

Karabenick, S. A. (2011). Classroom and technology-supported help seeking: The need for converging research paradigms. *Learning and Instruction*, *21*(2), 290–296. doi:10.1016/j.learninstruc.2010.07.007

Keuleers, E., Brysbaert, M., & New, B. (2010). SUBTLEX-NL: a new measure for Dutch word frequency based on film subtitles. *Behavior Research Methods*, *42*(3), 643–50. doi:10.3758/BRM.42.3.643

Kickmeier-Rust, M. D., & Albert, D. (2010). Micro-adaptivity: protecting immersion in didactically adaptive digital educational games. *Journal of Computer Assisted Learning*, *26*(2), 95–105. doi:10.1111/j.1365-2729.2009.00332.x

Kiili, K. (2005). Digital game-based learning: Towards an experiential gaming model. *The Internet and Higher Education*, *8*(1), 13–24. doi:10.1016/j.iheduc.2004.12.001

Kim, H., & Mathes, G. (2001). Explicit vs. implicit corrective feedback. *The Korea TESOL Journal*, *4*, 1–15.

Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why Minimal Guidance During Instruction Does Not Work : An Analysis of the Failure of Constructivist , Discovery , Problem-Based , Experiential , and Inquiry-Based Teaching. *Educational Psychologist*, *41*(2), 75–86.

Kluger, A. N., & Denisi, A. (1996). The Effects of Feedback Interventions on Performance: A Historical Review, a Meta-Analysis, and a Preliminary Feedback Intervention Theory. *Psychological Bulletin*, *119*(2), 254–284.

Koike, D. A., & Pearson, L. (2005). The effect of instruction and feedback in the development of pragmatic competence. *System*, *33*(3), 481–501.

Koster, R. (2005). *A Theory of Fun for Game Design*. Scottsdale, Arizona: Paraglyph Press.

Krashen, S. D. (1981). *Second language acquisition and second language learning*. Oxford: Pergamon.

Krashen, S. D. (1998). Comprehensible output? *System*, *26*(2), 175–182.

Kuder, G. F., & Richardson, M. W. (1937). The theory of the estimation of test reliability. *Psychometrika*, *2*(3), 151–160.

Kulhavy, R. W. (1977). Feedback in Written Instruction. *Review of Educational Research*, *47*(2), 211–232. doi:10.3102/00346543047002211

Kuppens, A. H. (2010). Incidental foreign language acquisition from media exposure. *Learning, Media and Technology*, *35*(1), 65–85. doi:10.1080/17439880903561876

Lagatie, R., & De Causmaecker, P. (2010). The Effect of Repetition Feedback on Success Rate and Uptake. In J. Sanchez & K. Zhang (Eds.), *Proceedings of World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education 2010* (pp. 536–542). Chesapeake, VA: AACE.

Lai, C., Fei, F., & Roots, R. (2008). The Contingency of Recasts and Noticing. *CALICO Journal*, *26*(1), 70–90.

Lai, C., & Zhao, Y. (2006). Noticing and text-based chat. *Language Learning & Technology*, *10*(3), 102–120.

Lantolf, J. P., & Poehner, M. E. (2014). *Sociocultural Theory and the Pedagogical Imperative in L2 Education. Vygotskian Praxis and the Research/Practice Divide*. London: Routledge.

Lantolf, J. P., & Thorne, S. L. (2006). *Sociocultural Theory and the Genesis of Second Language Development*. Oxford: Oxford University Press.

Lantolf, J. P., & Thorne, S. L. (2008). Sociocultural Theory and Second Language Learning. In B. VanPatten & J. Williams (Eds.), *Theories in Second Language Acquisition* (pp. 201–224). New York: Routledge.

Larsen-Freeman, D. (2003). *Teaching language. From grammar to grammaring*. Boston: Thomson/Heinle.

Laurel, B. (1993). *Computers as Theatre*. Boston: Addison-Wesley Professional.

Leech, G., & Svartvik, J. (1994). *A communicative grammar of English*. London: Longman.

Leeman, J. (2007). Feedback in L2 learning: Responding to errors during practice. In R. M. DeKeyser (Ed.), *Practice in a Second Language: Perspectives from Applied Linguistics and Cognitive Psychology* (pp. 111–137). New York: Cambridge University Press.

Levin, B. (1993). *English verb classes and alternations. A preliminary investigation*. Chicago: The University of Chicago Press.

Levy, M. (1997). *Computer-Assisted Language Learning: context and conceptualization*. Oxford: Clarendon Press.

Li, S. (2010). The Effectiveness of Corrective Feedback in SLA: A Meta-Analysis. *Language Learning*, *60*(2), 309–365. doi:10.1111/j.1467-9922.2010.00561.x

Lightbown, P. M. (2008). Transfer Appropriate Processing as a Model for Classroom Second Language Acquisition. In Z. Han (Ed.), *Understanding Second Language Process* (pp. 27–44). Clevedon: Multilingual Matters.

Loewen, S. (2009). Grammaticality Judgment Tests and the Measurement of Implicit and Explicit L2 Knowledge. In R. Ellis, S. Loewen, C. Elder, R. Erlam, J. Philp, & H. Reinders (Eds.), *Implicit and Explicit Knowledge in Second Language Learning, Testing and Teaching* (pp. 94–112). Bristol: Multilingual Matters.

Loewen, S., & Philp, J. (2006). Recasts in the Adult English L2 Classroom: Characteristics, Explicitness, and Effectiveness. *The Modern Language Journal*, *90*(4), 536–556. doi:10.1111/j.1540-4781.2006.00465.x

Loewen, S., & Reinders, H. (2011). *Key concepts in second language acquisition*. New York: Palgrave.

Logan, G. D. (1988). Toward an instance theory of automatization. *Psychological Review*, *95*(4), 492–527. doi:10.1037//0033-295X.95.4.492

Long, M. H. (2007). *Problems in SLA*. Mahwah, New Jersey: Lawrence Erlbaum Associates.

Long, M. H., Inagaki, S., & Ortega, L. (1998). The Role of Implicit Negative Feedback in SLA: Models and Recasts in Japanese and Spanish. *The Modern Language Journal*, *82*(3), 357–371.

Lust, G., Elen, J., & Clarebout, G. (2011). Adopting webcasts over time: the influence of perceptions and attitudes. *Journal of Computing in Higher Education*, *24*(1), 40–57. doi:10.1007/s12528-011-9052-9

Luyten, L., Lowyck, J., & Tuerlinckx, F. (2001). Task perception as a mediating variable: a contribution to the validation of instructional knowledge. *The British Journal of Educational Psychology*, *71*(2), 203–223. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11449933

Lyster, R. (1998). Recasts, repetition, and ambiguity in L2 classroom discourse. *Studies in Second Language Acquisition*, *20*(1), 51–81.

Lyster, R., & Ranta, L. (1997). Corrective Feedback and Learner Uptake. Negotiation of Form in Communicative Classrooms. *Studies in Second Language Acquisition*, *19*(01), 37–66. doi:10.1017/S0272263197001034

Lyster, R., & Saito, K. (2010). Oral Feedback in Classroom SLA. A Meta-Analysis. *Studies in Second Language Acquisition*, *32*(02), 265–302. doi:10.1017/S0272263109990520

Mackey, A., Gass, S. M., & McDonough, K. (2000). How do learners perceive interactional feedback? *Studies in Second Language Acquisition*, *44*(2), 471–497.

Mackey, A., & Goo, J. (2007). Interaction research in SLA: a meta-analysis and research synthesis. In A. Mackey (Ed.), *Conversational Interaction in Second Language Acquisition* (pp. 407–452). Oxford: Oxford UP.

Mackey, A., & Philp, J. (1998). Conversational Interaction and Second Language Development: Recasts, Responses, and Red Herrings ? *Modern Language Journal*, *82*(3), 338–356.

Magilow, D. H. (1999). Case Study #2: Error Correction and Classroom Affect. *Die Unterrichtspraxis / Teaching German*, *32*(2), 125–129. doi:10.2307/3531752

Malone, T. W. (1981). Toward a theory of intrinsically motivating instruction. *Cognitive Science*, *5*(4), 333–369.

Mantovani, F., & Castelnuovo, G. (2003). Sense of Presence in Virtual Training : Enhancing Skills Acquisition and Transfer of Knowledge through Learning Experience in Virtual Environments. In G. Riva, F. Davide, & W. A. Ijsselsteijn (Eds.), *Being There: Sense of presence in virtual training* (pp. 167–182). Amsterdam: IOS Press.

Matterson, C. (2013). Scientists' public engagement work should be generously funded. *The Guardian*.

Mawer, K., & Stanley, G. (2011). *Digital Play. Computer games and language aims*. Peaslake: Delta Publishing.

Mayer, R. E. (2001). *Multimedia learning*. New York: Cambridge University Press.

Mazurkewich, I., & White, L. (1984). The acquisition of the dative alternation: unlearning overgeneralizations. *Cognition*, *16*(3), 261–283. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/6541106

McDonough, K. (2007). Interactional feedback and the emergence of simple past activity verbs in L2 English. In A. Mackey (Ed.), *Conversational Interaction in Second Language Acquisition* (pp. 323–338). Oxford: Oxford UP.

McGonigal, J. (2011). *Reality Is Broken: Why Games Make Us Better and How They Can Change the World*. New York: The Penguin Press.

Mikulincer, M., Kedem, P., & Zilkha-Segal, H. (1989). Learned helplessness, reactance, and cue utilization. *Journal of Research in Personality*, *23*(2), 235–247. doi:10.1016/0092-6566(89)90026-3

Miller, M., & Hegelheimer, V. (2006). The SIMs meet ESL. Incorporating authentic computer simulation games into the language classroom. *Interactive Technology and Smart Education*, *3*(4), 311–328.

Morton, H., & Jack, M. A. (2005). Scenario-Based Spoken Interaction with Virtual Agents. *Computer-Assisted Language Learning*, *18*(3), 171–191.

Nagata, N. (1993). Intelligent Computer Feedback for Second Language Instruction. *Modern Language Journal*, *77*(3), 330–339.

Neville, D. O., Shelton, B. E., & McInnis, B. (2009). Cybertext redux: Using digital game-based learning to teach L2 vocabulary, reading, and culture. *Computer Assisted Language Learning*, *22*(5), 409–424.

Nicholas, H., Lightbown, P. M., & Spada, N. (2001). Recasts as feedback to language learners. *Language Learning*, *51*(4), 719–758.

Noels, K. A. (2001). Learning Spanish as a Second Language: Learners' Orientations and Perceptions of Their Teachers' Communication Style. *Language Learning*, *51*(1), 107–144. doi:10.1111/0023-8333.00149

Noels, K. A., Pelletier, L. G., Clément, R., & Vallerand, R. J. (2000). Why Are You Learning a Second Language? Motivational Orientations and Self-Determination Theory. *Language Learning*, *50*(1), 57–85. doi:10.1111/1467-9922.53223

Norman, G. (2010). Likert scales, levels of measurement and the "laws" of statistics. *Advances in Health Sciences Education : Theory and Practice*, *15*(5), 625–32. doi:10.1007/s10459-010-9222-y

Norris, J. M., & Ortega, L. (2000). Effectiveness of L2 Instruction: A Research Synthesis and Quantitative Meta-analysis. *Language Learning*, *50*(3), 417–528.

Örnberg Berglund, T. (2012). Corrective Feedback and Noticing in Text-Based Second Language Interaction. In L. Bradley & S. Thouësny (Eds.), *CALL:*

Using, Learning, Knowing. Proceedings of EUROCALL 2012 Conference, Gothenburg, Sweden, 22-25 August 2012 (Vol. 2, pp. 234–239). Research-publishing.net. doi:10.14705/rpnet.2012.000058

Ortega, L. (2007). *Understanding second language acquisition*. London: Hodder.

Papert, S. (1998). Does Easy Do It? Children, Games, and Learning. *Game Developer*, 88. Retrieved from http://www.papert.org/articles/Doeseasydoit.html

Paulston, C. B., & Bruder, M. N. (1976). *Teaching English as a second language: Techniques and procedures*. Cambridge, MA: Winthrop.

Pendergrast, M. (1997). *For God, country and Coca-Cola: the unauthorized history of the Great American soft drink and the company that makes it*. New York: Touchstone.

Petersen, K. A. (2010). *Implicit corrective feedback in computer-guided interaction: does mode matter*. Georgetown University.

Phillips, M. K. (1986). *Communicative language learning and the microcomputer*. London: British Council.

Phillips, M. K. (1987). Potential Paradigms and Possible Problems for CALL. *System*, *15*(3), 275–287.

Philp, J. (2003). Constraints on "Noticing the Gap". Nonnative Speakers' Noticing of Recasts in NS-NNS Interaction. *Studies in Second Language Acquisition*, *25*(01). doi:10.1017/S0272263103000044

Pica, T. (1994). Questions from the Language Classroom: Research Perspectives. *TESOL Quarterly*, *28*(1), 49–79.

Pinker, S. (1989). *Learnability and Cognition*. Cambridge, MA: MIT Press.

Pintrich, P. R., Smith, D. A. F., Garcia, T., & Mckeachie, W. J. (1993). Reliability and Predictive Validity of the Motivated Strategies for Learning Questionnaire (MSLQ). *Educational and Psychological Measurement*, *53*(3), 801–813. doi:10.1177/0013164493053003024

Plant, R. W., & Ryan, R. M. (1985). Intrinsic motivation and the effects of self-involvement : An investigation of internally controlling styles. *Journal of Personality*, *53*(3), 435–449.

Prensky, M. (2001). *Digital game-based learning*. St. Paul: Paragon House.

Przybylski, A. K., Rigby, C. S., & Ryan, R. M. (2010). A motivational model of video game engagement. *Review of General Psychology*, *14*(2), 154–166. doi:10.1037/a0019440

Pujolà, J.-T. (2001). Did CALL feedback feed back ? Researching learners' use of feedback. *ReCALL*, *13*(1), 79–98.

Purushotma, R., Thorne, S. L., & Wheatley, J. (2008). *10 key principles for designing video games for foreign language learning*. Retrieved from http://knol.google.com/k/ravi-purushotma/10-key-principles-for-designing-video/27mkxqba7b13d/2#done

Radecki, P. M., & Swales, J. M. (1988). ESL Student Reaction to Written Comments on Their Written Work. *System*, *16*(3), 355–365.

Raimes, A. (1991). Errors: Windows into the mind. *College ESL*, *1*(2), 55–64.

Ranalli, J. (2008). Learning English with The Sims: exploiting authentic computer simulation games for L2 learning. *Computer Assisted Language Learning*, *21*(5), 441–455. doi:10.1080/09588220802447859

Ranta, L., & Lyster, R. (2007). A cognitive approach to improving immersion students' oral language abilities: The Awareness-Practice-Feedback sequence. In R. M. DeKeyser (Ed.), *Practice in a Second Language: Perspectives from Applied Linguistics and Cognitive Psychology* (pp. 141–160). New York: Cambridge University Press.

Ravaja, N., Saari, T., Salminen, M., Laarni, J., & Kallinen, K. (2006). Phasic Emotional Reactions to Video Game Events: A Psychophysiological Investigation. *Media Psychology*, *8*(4), 343–367. doi:10.1207/s1532785xmep0804_2

Reber, A. S. (1989). Implicit learning and tacit knowledge. *Journal of Experimental Psychology: General*, *118*(3), 219–235. doi:10.1037//0096-3445.118.3.219

Reichle, R. V. (2012). Sprites and Rules: What ERPs and Procedural Memory Can Tell Us about Video Games and Language Learning. In H. Reinders (Ed.), *Digital Games in Language Learning and Teaching* (pp. 139–155). Basingstoke: Palgrave Macmillan.

Reinhardt, J., & Sykes, J. M. (2012). Conceptualizing Digital Game-Mediated L2 Learning and Pedagogy: Game-Enhanced and Game-Based Research and Practice. In H. Reinders (Ed.), *Digital Games in Language Learning and Teaching* (pp. 32–49). Basingstoke: Palgrave Macmillan.

Rigby, C. S., & Przybylski, A. K. (2009). Virtual worlds and the learner hero: How today's video games can inform tomorrow's digital learning environments. *Theory and Research in Education*, *7*(2), 214–223. doi:10.1177/1477878509104326

Rigby, C. S., & Ryan, R. M. (2011). *Glued to Games. How Video Games Draw Us In and Hold Us Spellbound*. Santa Barbara: Praeger.

Robinson, G. L. (1991). Effective feedback strategies in CALL: learning theory and empirical research. In P. A. Dunkel (Ed.), *Computer-Assisted Language Learning and Testing: Research Issues and Practice* (pp. 155–167). New Jersey: Newbury House.

Robinson, P. (1997). Generalizability and automaticity of second language learning under implicit, incidental, enhanced, and instructed conditions. *Studies in Second Language Acquisition*, *19*(2), 223–247.

Robinson, P., & Ha, M. A. (1993). Instance theory and second language rule learning under explicit conditions. *Studies in Second Language Acquisition*, *15*(4), 413–438.

Russell, J., & Spada, N. (2006). The effectiveness of corrective feedback for the acquisition of L2 grammar: A meta-analysis of the research. In J. M. Norris & L. Ortega (Eds.), *Synthesizing Research on Language Learning and Teaching* (pp. 133–164). Amsterdam: John Benjamins.

Ryan, R. M., & Deci, E. L. (2000). Intrinsic and Extrinsic Motivations: Classic Definitions and New Directions. *Contemporary Educational Psychology*, *25*(1), 54–67. doi:10.1006/ceps.1999.1020

Ryan, R. M., Rigby, C. S., & Przybylski, A. K. (2006). The Motivational Pull of Video Games: A Self-Determination Theory Approach. *Motivation and Emotion*, *30*(4), 344–360. doi:10.1007/s11031-006-9051-8

Sachs, R., & Suh, B.-R. (2007). Textually enhanced recasts, learner awareness, and L2 outcomes in synchronous computer-mediated interaction. In A. Mackey (Ed.), *Conversational Interaction in Second Language Acquisition* (pp. 197–227). Oxford: Oxford UP.

Saito, H. (1994). Teachers' Practices and Students' Preferences for Feedback on Second Language Writing : A Case Study of Adult ESL Learners. *TESL Canada Journal*, *11*(2), 46–70.

Salen, K., & Zimmerman, E. (2004). *Rules of Play: Game Design Fundamentals*. Cambridge: MIT Press.

Sanders, R. H., & Sanders, A. F. (1995). History of an AI spy game: SPION. *CALICO Journal*, *12*(4), 114–127.

Schachter, J. (1991). Corrective feedback in historical perspective. *Second Language Research*, *7*(2), 89–102. doi:10.1177/026765839100700202

Schmidt, R. W. (1990). The role of consciousness in second language learning. *Applied Linguistics*, *11*(2), 17–46.

Schulz, R. A. (2001). Cultural Differences in Student and Teacher Perceptions Concerning the Role of Grammar Instruction and Corrective Feedback: USA-Colombia. *The Modern Language Journal*, *85*(2), 244–258. doi:10.1111/0026-7902.00107

Schulze, M. (2003). Grammatical Errors and Feedback: Some Theoretical Insights. *CALICO Journal*, *20*(3), 437–450.

Schulze, M. (2010). Taking Intelligent CALL to Task. In M. Thomas & H. Reinders (Eds.), *Task-Based Language Learning and Teaching with Technology* (pp. 63–82). London: Continuum.

Schwartz, B. D. (1993). On explicit and negative data effecting and affecting competence and "linguistic behavior." *Studies in Second Language Acquisition*, *15*(2), 147–163.

Schwienhorst, K. (2002). Why Virtual, Why Environments? Implementing Virtual Reality Concepts in Computer-Assisted Language Learning. *Simulation & Gaming*, *33*(2), 196–209. Retrieved from http://sag.sagepub.com/cgi/reprint/33/2/196

Segalowitz, N. (2005). Automaticity and Second Languages. In C. J. Doughty & M. H. Long (Eds.), *The Handbook of Second Language Acquisition* (pp. 383–408). Oxford: Blackwell Publishing Ltd.

Segalowitz, N., & Hulstijn, J. H. (2005). Automaticity in bilingualism and second language learning. In J. F. Kroll & A. M. . B. De Groot (Eds.), *Handbook of bilingualism: Psycholinguistic approaches* (pp. 371–388). Oxford: Oxford University Press.

Segalowitz, N., & Segalowitz, S. (1993). Skilled Performance, Practice, and the Differentiation of Speed-Up from Automatization Effects: Evidence from Second Language Word Recognition. *Applied Psycholinguistics*, *14*(3), 369–85.

Sheen, Y. (2006). Exploring the relationship between characteristics of recasts and learner uptake. *Language Teaching Research*, *10*(4), 361–392. doi:10.1191/1362168806lr203oa

Sheen, Y. (2007). The effect of corrective feedback, language aptitude, and learner attitudes on the acquisition of English articles. In A. Mackey (Ed.), *Conversational Interaction in Second Language Acquisition* (pp. 301–322). Oxford: Oxford UP.

Sheen, Y. (2008). Recasts, Language Anxiety, Modified Output, and L2 Learning. *Language Learning*, *58*(4), 835–874. doi:10.1111/j.1467-9922.2008.00480.x

Sheen, Y. (2010). Introduction. The Role of Oral and Written Corrective Feedback in SLA. *Studies in Second Language Acquisition*, *32*(02), 169–179. doi:10.1017/S0272263109990489

Sheen, Y. (2011). *Corrective Feedback, Individual Differences and Second Language Learning*. London: Springer.

Shute, V. J. (2008). Focus on Formative Feedback. *Review of Educational Research*, *78*(1), 153–189. doi:10.3102/0034654307313795

Shute, V. J., Masduki, I., & Donmez, O. (2010). Conceptual Framework for Modeling, Assessing and Supporting Competencies within Game Environments. *Technology, Instruction, Cognition and Learning*, *8*(2), 137–161.

Sims, R. (1997). Interactivity: A forgotten art? *Computers in Human Behavior*, *13*(2), 157–180. doi:10.1016/S0747-5632(97)00004-6

Smith, B. (2012). Eye Tracking as a Measure of Noticing: A Study of Explicit Recasts in SCMC. *Language Learning & Technology*, *16*(3), 53–81.

Steinkuehler, C., & Duncan, S. (2008). Scientific Habits of Mind in Virtual Worlds. *Journal of Science Education and Technology*, *17*(6), 530–543. doi:10.1007/s10956-008-9120-8

Steuer, J. (1992). Defining Virtual Reality: Dimensions Determining Telepresence. *Journal of Communication*, *42*(4), 73–93.

Stouten, F., Duchateau, J., Martens, J.-P., & Wambacq, P. (2006). Coping with disfluencies in spontaneous speech recognition: Acoustic detection and linguistic context manipulation. *Speech Communication*, *48*(11), 1590–1606. doi:10.1016/j.specom.2006.04.004

Strik, H., Cornillie, F., Colpaert, J., van Doremalen, J., & Cucchiarini, C. (2009). Developing a CALL System for Practicing Oral Proficiency: How to Design for Speech Technology, Pedagogy and Learners. In *Proceedings of the SLaTE-2009 workshop on Speech and Language Technology in Education*. Warwickshire (England).

Strik, H., Drozdova, P., & Cucchiarini, C. (2013). GOBL : Games Online for Basic Language Learning. In P. Badin, T. Hueber, G. Bailly, D. Demolin, & F. Raby (Eds.), *Proceedings of the SLaTE-2013 workshop on Speech and Language Technology in Education* (Vol. 7, pp. 48–53). Grenoble, France.

Swink, S. (2006). What is Virtual Sensation ? *Gamasutra. The art & Business of Making Games*. Retrieved January 10, 2012, from http://www.gamasutra.com/view/feature/1781/principles_of_virtual_sensation.php

Sykes, J. M. (2009). Learner requests in Spanish: Examining the potential of multiuser virtual environments for L2 pragmatic acquisition. In L. Lomicka & G. Lord (Eds.), *The Next Generation: Social Networking and Online Collaboration in Foreign Language Learning*. Durham: Computer Assisted Language Instruction Consortium.

Takimoto, M. (2006a). The effects of explicit feedback and form–meaning processing on the development of pragmatic proficiency in consciousness-raising tasks. *System*, *34*(4), 601–614. doi:10.1016/j.system.2006.09.003

Takimoto, M. (2006b). The effects of explicit feedback on the development of pragmatic proficiency. *Language Teaching Research*, *10*(4), 393–417. doi:10.1191/1362168806lr198oa

Taylor, M. (1990). Simulations and Adventure Games in CALL. *Simulation & Gaming*, *21*(4), 461–466.

Thorne, S. L. (2008). Transcultural Communication in Open Internet Environments and Massively Multiplayer Online Games. In S. S. Magnan (Ed.), *Mediating Discourse Online* (pp. 305–327). Amsterdam: John Benjamins.

Thorne, S. L., Black, R. W., & Sykes, J. M. (2009). Second Language Use, Socialization, and Learning in Internet Interest Communities and Online Gaming. *Modern Language Journal*, *93*(0), 802–821.

Thorne, S. L., Fischer, I., & Lu, X. (2012). The semiotic ecology and linguistic complexity of an online game world. *ReCALL*, *24*(03), 279–301. doi:10.1017/S0958344012000158

Tobias, S., Fletcher, J. D., Dai, D. Y., & Wind, A. P. (2011). Review of Research on Computer Games. In S. Tobias & J. D. Fletcher (Eds.), *Computer Games and Instruction* (pp. 127–222). Charlotte: Information Age Publishing.

Tops, G. A. J., Dekeyser, X., Devriendt, B., & Geukens, S. (2001). Dutch speakers. In M. Swan & B. Smith (Eds.), *Learner English. A teacher's guide to interference and other problems* (2nd ed., pp. 1–20). Cambridge: Cambridge University Press.

Truscott, J. (1996). The Case Against Grammar Correction in L2 Writing Classes. *Language Learning*, *46*(2), 327–369.

Ushioda, E., & Dörnyei, Z. (2009). Motivation, Language Identities and the L2 Self: A Theoretical Overview. In Z. Dörnyei & E. Ushioda (Eds.), *Motivation, language identity and the L2 self* (pp. 1–8). Bristol: Multilingual Matters.

Vallerand, R. J., & Reid, G. (1984). On the Causal Effects of Perceived Competence on Intrinsic Motivation: A Test of Cognitive Evaluation Theory. *Journal of Sport & Exercise Psychology*, *6*(1), 94–102.

Van den Branden, K. (2007). Second language education: Practice in perfect learning conditions? In R. M. DeKeyser (Ed.), *Practice in a Second Language: Perspectives from Applied Linguistics and Cognitive Psychology* (pp. 161–179). New York: Cambridge University Press.

Van den Branden, K., Bygate, M., & Norris, J. M. (2009). *Task-Based Language Teaching. A reader*. Amsterdam: John Benjamins.

Van Eck, R. (2006). Digital Game-Based Learning: It's Not Just the Digital Natives Who Are Restless. *Educause Review*, *41*(2), 17–30.

Van Eck, R. (2009). A guide to integrating COTS games into your classroom. In R. E. Ferdig (Ed.), *Handbook of research on effective electronic gaming in education* (pp. 179–199). Hershey, PA: Information Science.

Van Merriënboer, J. J. G., & Kirschner, P. A. (2007). *Ten steps to complex learning. A systematic approach to Four-Component Instructional Design.* Mahwah, New Jersey: Lawrence Erlbaum Associates.

Vandercruysse, S., Vandewaetere, M., & Clarebout, G. (2012). Game-Based Learning: A Review on the Effectiveness of Educational Games. In M. M. Cruz-Cunha (Ed.), *Handbook of Research on Serious Games as Educational, Business, and Research Tools*. Hershey, PA: IGI Global.

Vandercruysse, S., Vandewaetere, M., Cornillie, F., & Clarebout, G. (2013). Competition and students' perceptions in a game-based language learning environment. *Educational Technology Research and Development*, *61*(6), 927–950.

Vandewaetere, M., Cornillie, F., Clarebout, G., & Desmet, P. (2013). Adaptivity in Educational Games: Including Player and Gameplay Characteristics. *International Journal of Higher Education*, *2*(2), 106–114. doi:10.5430/ijhe.v2n2p106

Vandewaetere, M., Desmet, P., & Clarebout, G. (2011). The contribution of learner characteristics in the development of computer-based adaptive learning environments. *Computers in Human Behavior*, *27*(1), 118–130. doi:10.1016/j.chb.2010.07.038

Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User Acceptance of Information Technology: Toward a Unified View. *MIS Quarterly*, *27*(3), 425–478.

Vygotsky, L. S. (1978). *Mind in Society. The Development of Higher Psychological Processes*. (M. Cole, V. John-Steiner, S. Scribner, & E. Souberman, Eds.). Cambridge, MA: Harvard University Press.

Walz, S. P., & Deterding, S. (2014). *The Gameful World: Approaches, Issues, Applications*. Cambridge, MA: MIT Press.

Waring, R., & Nation, P. (1997). Vocabulary size, text coverage, and word lists. In N. Schmitt & M. McCarthy (Eds.), *Vocabulary: Description, Acquisition and Pedagogy* (pp. 6–19). Cambridge: Cambridge University Press.

White, L. (2008). Linguistic Theory, Universal Grammar, and Second Language Acquisition. In B. VanPatten & J. Williams (Eds.), *Theories in Second Language Acquisition* (pp. 37–56). New York: Routledge.

Willis, D., & Willis, J. (2007). *Doing task-based teaching*. Oxford: Oxford University Press.

Winne, P. H. (1987). Why process-product research cannot explain process-product findings and a proposed remedy: The cognitive mediational paradigm. *Teaching and Teacher Education*, *3*(4), 333–356. doi:10.1016/0742-051X(87)90025-4

Winne, P. H. (2004). Students' calibration of knowledge and learning processes: Implications for designing powerful software learning environments. *International Journal of Educational Research*, *41*(6), 466–488. doi:10.1016/j.ijer.2005.08.012

Wong, W., & VanPatten, B. (2003). The Evidence is IN: Drills are OUT. *Foreign Language Annals*, *36*(3), 403–423.

Wood, C., Kemp, N., & Waldron, S. (2014). Exploring the longitudinal relationships between the use of grammar in text messaging and performance on grammatical tasks. *The British Journal of Developmental Psychology*, 1–15. doi:10.1111/bjdp.12049

Wright, W. (2003). Lessons from Game Design. Conversations Network. Retrieved from http://itc.conversationsnetwork.org/shows/detail195.html

Wu, X. (2003). Intrinsic motivation and young language learners: the impact of the classroom environment. *System*, *31*(4), 501–517. doi:10.1016/j.system.2003.04.001

Wu, X., Lowyck, J., Sercu, L., & Elen, J. (2012). Self-efficacy, task complexity and task performance: Exploring interactions in two versions of vocabulary learning tasks. *Learning Environments Research*, *15*(1), 17–35. doi:10.1007/s10984-012-9098-2

Wylin, B., & Desmet, P. (2005). A Typology of Educational Game Use : How to Integrate Games or Game Elements in Educational Multimedia ? *Proceedings of the World Conference on Educational Multimedia, Hypermedia and Telecommunications 2005*, 1194–1199.

Zacharski, R. (2003). A Discourse System for Conversational Characters. In *Proceedings of the Fourth International Conference on Intelligent Text Processing and Computational Linguistics, ed. by Alexander Gelbukh* (pp. 492–495). Heidelberg: Springer-Verlag.

Zaman, B., Poels, Y., Sulmon, N., Annema, H., Verstraete, M., Cornillie, F., … Desmet, P. (2012). Concepts and Mechanics for Educational Mini-Games. A Human-Centred Conceptual Design Approach involving Adolescent Learners and Domain Experts. *International Journal On Advances in Intelligent Systems*, *5*(3-4), 567–576.

Zheng, D., Young, M. F., Brewer, R. A., & Wagner, M. (2009). Attitude and Self-Efficacy Change: English Language Learning in Virtual Worlds. *CALICO Journal*, *27*(1), 205–231.

Zobl, H. (1995). Converging Evidence for the “Acquisition-Learning” Distinction. *Applied Linguistics*, *16*(1), 35–56.

Appendices

Appendix 1

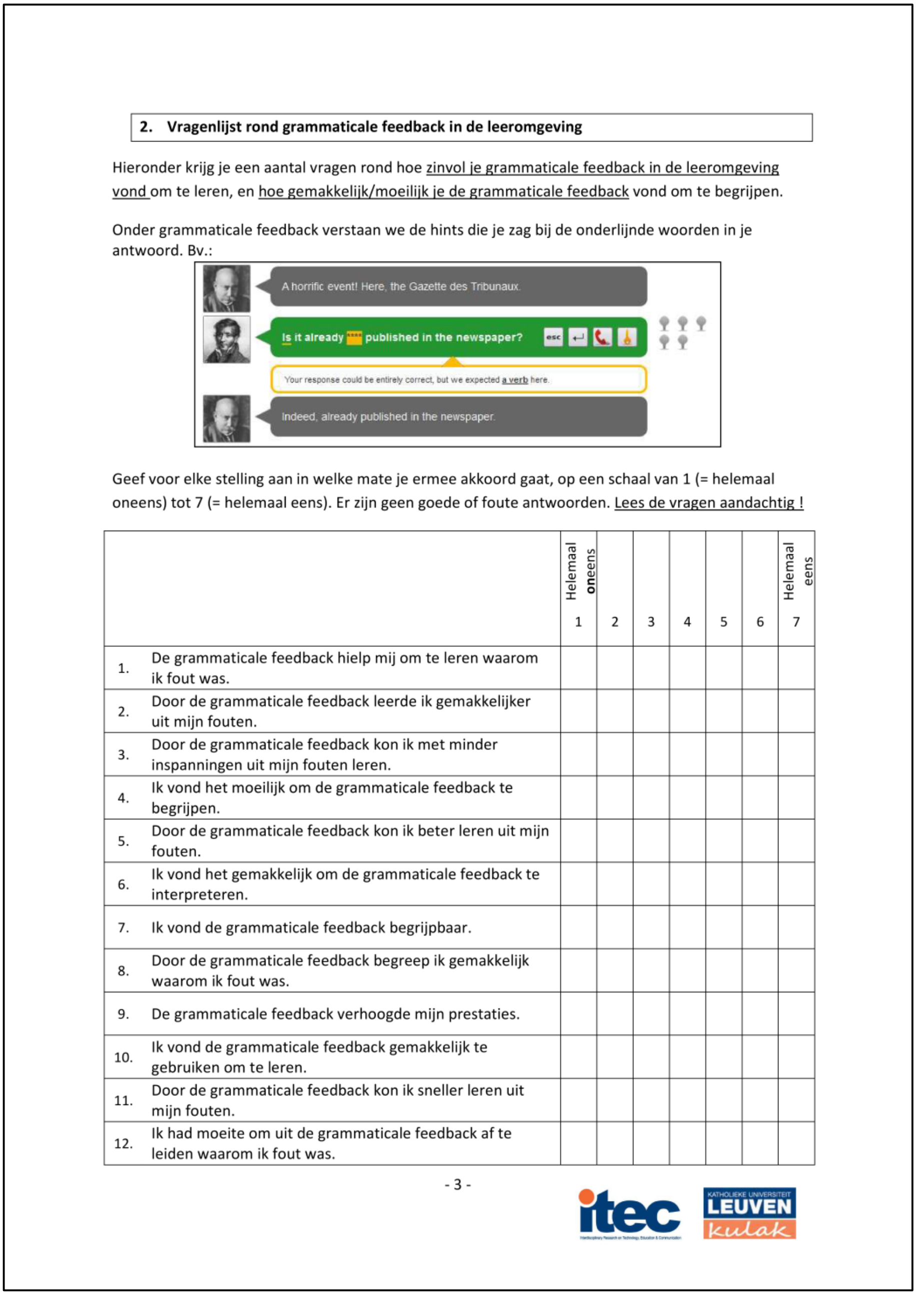
This appendix concerns the questionnaire on CF used in empirical study 1.

#	statement	scale	*M*	*SD*
1	I had the impression that I learned more from my mistakes when I could discover the rule myself (i.e. when it wasn't shown automatically).	I	4.1	1.8
2	I had the impression that I learned more when I could review the options after giving a response (i.e. if the conversation did not go on immediately).	E	5.3	1.1
3	The characters' reactions in the conversations helped me to learn from my mistakes.	I	5.1	1.3
4	I had the feeling that I learned from the rules that were shown in combination with incorrect responses.	E	5.0	1.4
5	Comparing incorrect responses to correct answers helped me to learn.	E	5.3	1.2
6	If I make a mistake in such an environment, I prefer that one of the characters indicates through his/her reaction that I was wrong.	I	5.0	1.5
7	If I make a mistake in such an environment, I prefer that the system lets me discover myself what the mistake was.	I	4.0	1.6
8	If I make a mistake in such an environment, I prefer that I can request a rule that explains my mistake.	E	5.0	1.4
9	If I make a mistake in such an environment, I prefer that the system automatically shows me a rule that explains my mistake.	E	4.8	1.6
10	If I make a mistake in such an environment, I prefer that my mistakes are only shown after the conversation.	I, -E	2.3	1.4
11	If I make a mistake in such an environment, I prefer that the rules are only shown after the conversation.	I, -E	2.4	1.4

scale: I = implicit, E = explicit, - = reverse scored

Appendix 2

This appendix concerns the post-questionnaire on CF used in empirical study 2.



2. Vragenlijst rond grammaticale feedback in de leeromgeving

Hieronder krijg je een aantal vragen rond hoe zinvol je grammaticale feedback in de leeromgeving vond om te leren, en hoe gemakkelijk/moeilijk je de grammaticale feedback vond om te begrijpen.

Onder grammaticale feedback verstaan we de hints die je zag bij de onderlijnde woorden in je antwoord. Bv.:

A horrific event! Here, the Gazette des Tribunaux.

Is it already **** published in the newspaper?

Your response could be entirely correct, but we expected a verb here.

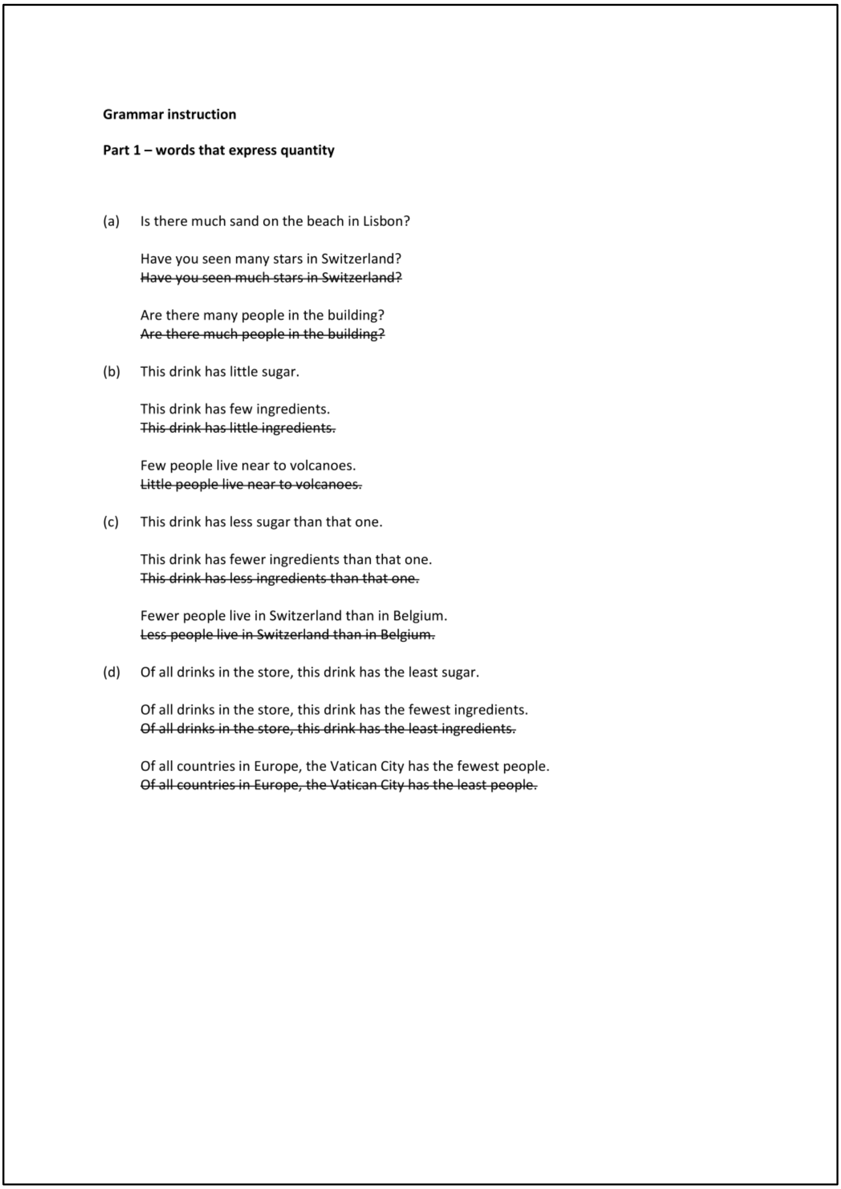
Indeed, already published in the newspaper.

Geef voor elke stelling aan in welke mate je ermee akkoord gaat, op een schaal van 1 (= helemaal oneens) tot 7 (= helemaal eens). Er zijn geen goede of foute antwoorden. Lees de vragen aandachtig !

		Helemaal **oneens** 1	2	3	4	5	6	Helemaal eens 7
1.	De grammaticale feedback hielp mij om te leren waarom ik fout was.							
2.	Door de grammaticale feedback leerde ik gemakkelijker uit mijn fouten.							
3.	Door de grammaticale feedback kon ik met minder inspanningen uit mijn fouten leren.							
4.	Ik vond het moeilijk om de grammaticale feedback te begrijpen.							
5.	Door de grammaticale feedback kon ik beter leren uit mijn fouten.							
6.	Ik vond het gemakkelijk om de grammaticale feedback te interpreteren.							
7.	Ik vond de grammaticale feedback begrijpbaar.							
8.	Door de grammaticale feedback begreep ik gemakkelijk waarom ik fout was.							
9.	De grammaticale feedback verhoogde mijn prestaties.							
10.	Ik vond de grammaticale feedback gemakkelijk te gebruiken om te leren.							
11.	Door de grammaticale feedback kon ik sneller leren uit mijn fouten.							
12.	Ik had moeite om uit de grammaticale feedback af te leiden waarom ik fout was.							

Appendix 3

This appendix contains the sheets with exemplars used for inductive rule instruction in empirical study 4.



Grammar instruction

Part 1 – words that express quantity

(a) Is there much sand on the beach in Lisbon?

Have you seen many stars in Switzerland?
~~Have you seen much stars in Switzerland?~~

Are there many people in the building?
~~Are there much people in the building?~~

(b) This drink has little sugar.

This drink has few ingredients.
~~This drink has little ingredients.~~

Few people live near to volcanoes.
~~Little people live near to volcanoes.~~

(c) This drink has less sugar than that one.

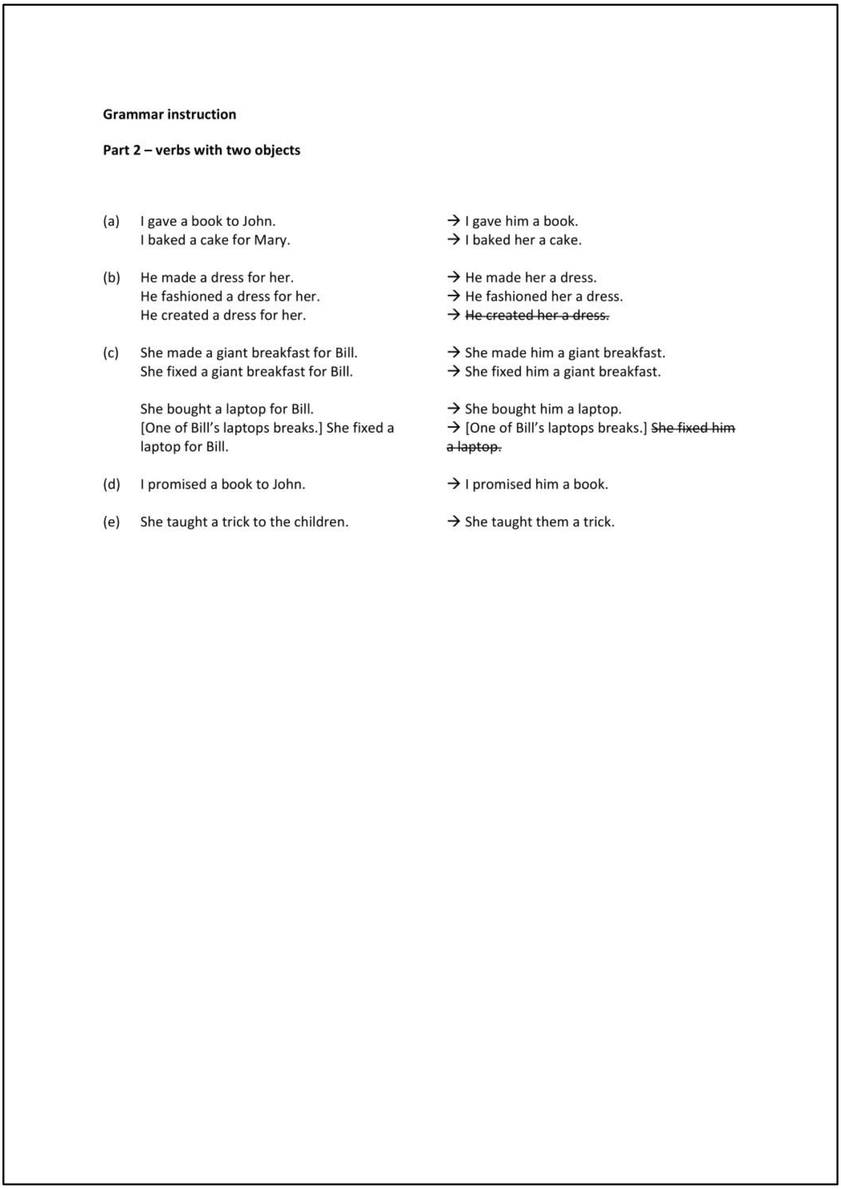
This drink has fewer ingredients than that one.
~~This drink has less ingredients than that one.~~

Fewer people live in Switzerland than in Belgium.
~~Less people live in Switzerland than in Belgium.~~

(d) Of all drinks in the store, this drink has the least sugar.

Of all drinks in the store, this drink has the fewest ingredients.
~~Of all drinks in the store, this drink has the least ingredients.~~

Of all countries in Europe, the Vatican City has the fewest people.
~~Of all countries in Europe, the Vatican City has the least people.~~



Grammar instruction

Part 2 – verbs with two objects

(a) I gave a book to John. → I gave him a book.
I baked a cake for Mary. → I baked her a cake.

(b) He made a dress for her. → He made her a dress.
He fashioned a dress for her. → He fashioned her a dress.
He created a dress for her. → ~~He created her a dress.~~

(c) She made a giant breakfast for Bill. → She made him a giant breakfast.
She fixed a giant breakfast for Bill. → She fixed him a giant breakfast.

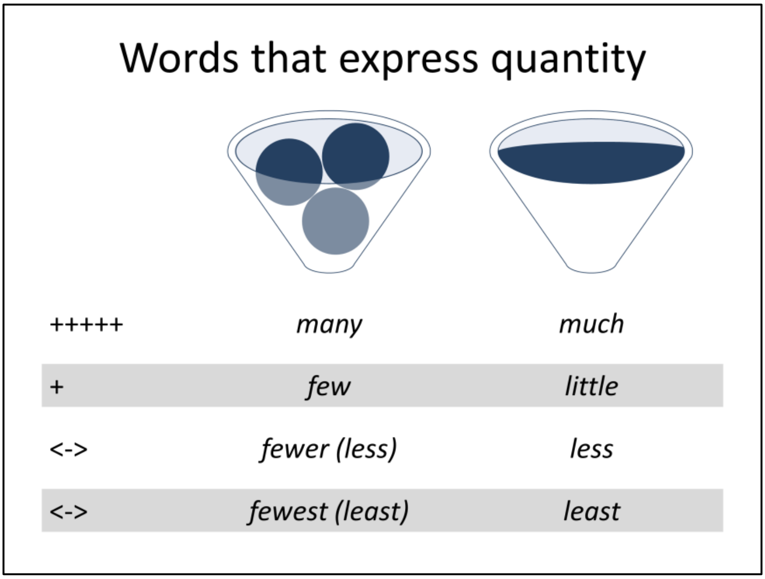
She bought a laptop for Bill. → She bought him a laptop.
[One of Bill's laptops breaks.] She fixed a laptop for Bill. → [One of Bill's laptops breaks.] ~~She fixed him a laptop.~~

(d) I promised a book to John. → I promised him a book.

(e) She taught a trick to the children. → She taught them a trick.

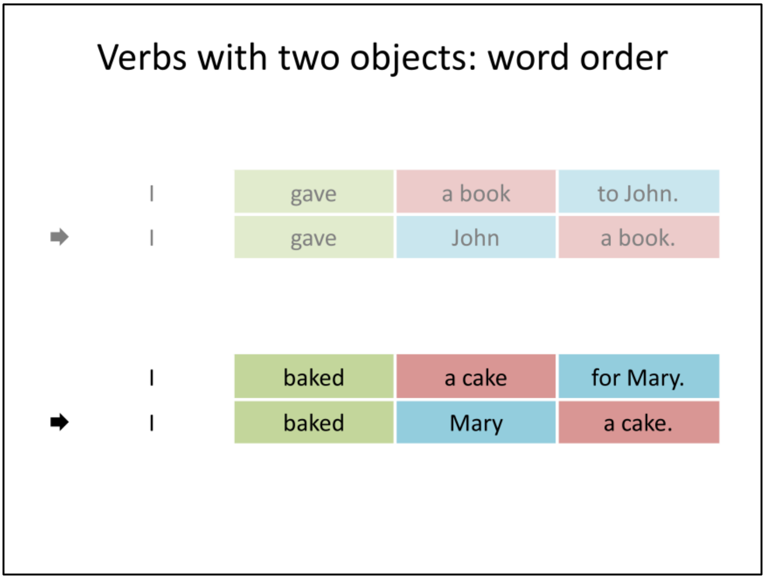
Appendix 4

This appendix contains the slides supporting the rule instruction in study 4.



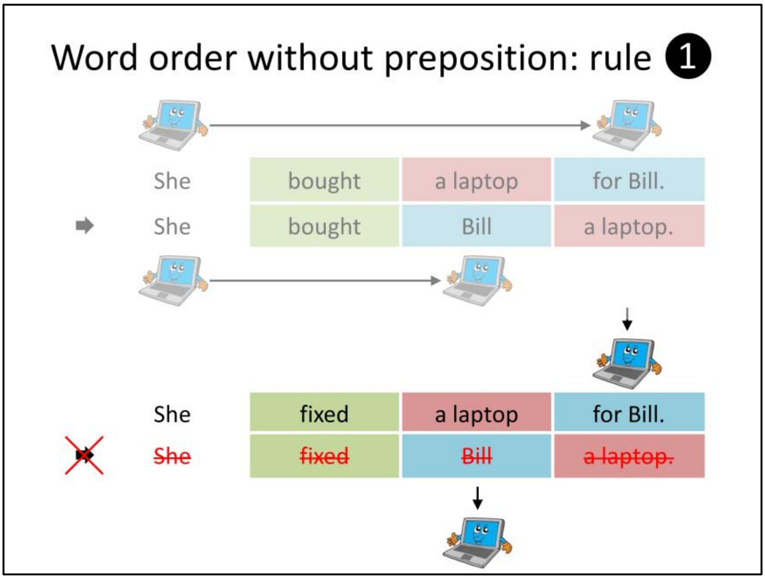
Words that express quantity

+++++	*many*	*much*
+	*few*	*little*
<->	*fewer (less)*	*less*
<->	*fewest (least)*	*least*



Verbs with two objects: word order

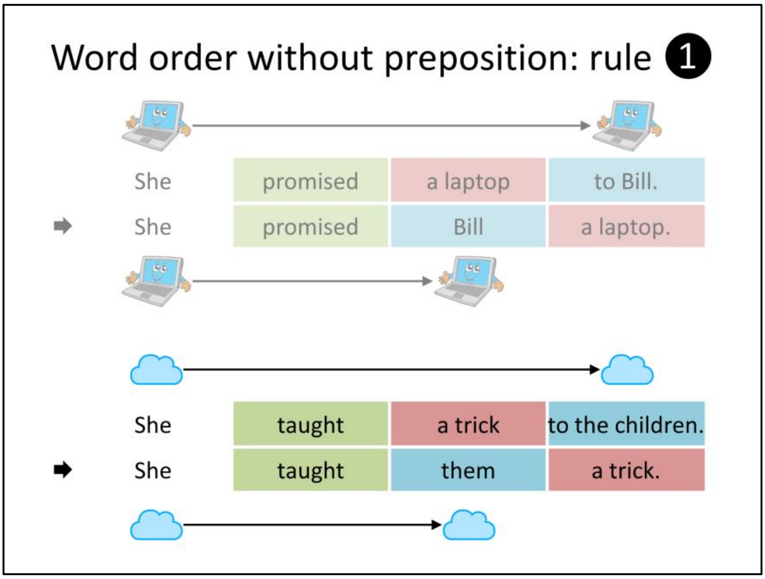
	I	gave	a book	to John.
➡	I	gave	John	a book.
	I	baked	a cake	for Mary.
➡	I	baked	Mary	a cake.



Word order without preposition: rule 1

	She	bought	a laptop	for Bill.
➡	She	bought	Bill	a laptop.

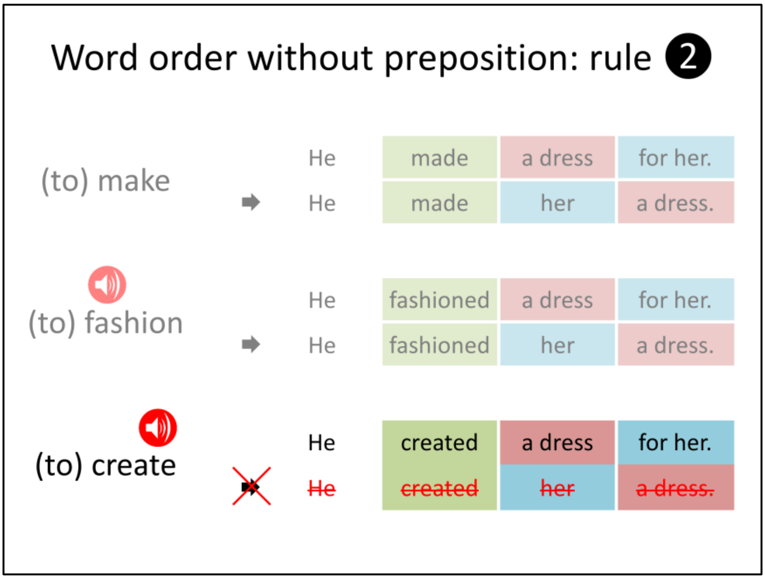
	She	fixed	a laptop	for Bill.
☒	~~She~~	~~fixed~~	~~Bill~~	~~a laptop.~~



Word order without preposition: rule 1

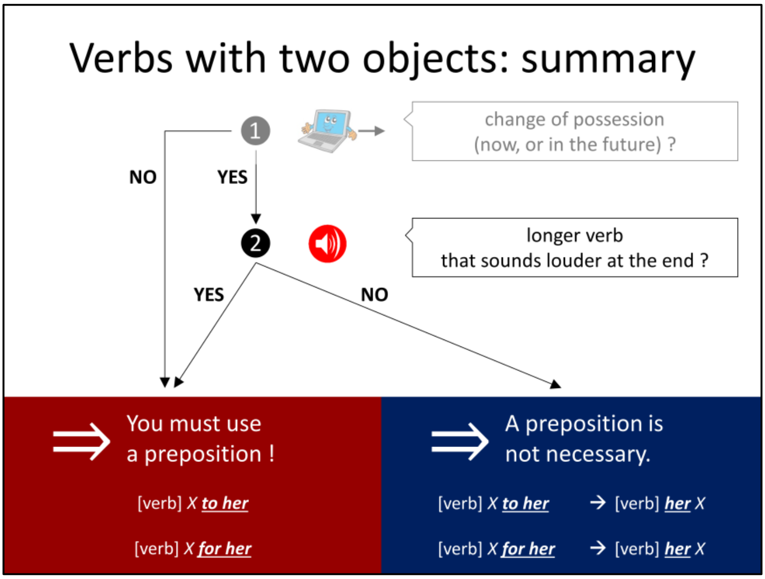
	She	promised	a laptop	to Bill.
➡	She	promised	Bill	a laptop.

	She	taught	a trick	to the children.
➡	She	taught	them	a trick.



Word order without preposition: rule ❷

(to) make		He	made	a dress	for her.
	➡	He	made	her	a dress.
(to) fashion		He	fashioned	a dress	for her.
	➡	He	fashioned	her	a dress.
(to) create		He	created	a dress	for her.
	~~➡~~	~~He~~	~~created~~	~~her~~	~~a dress.~~



Verbs with two objects: summary

1 change of possession (now, or in the future) ?

NO

YES

2 longer verb that sounds louder at the end ?

YES

NO

⇒ You must use a preposition !

[verb] *X* ***to her***

[verb] *X* ***for her***

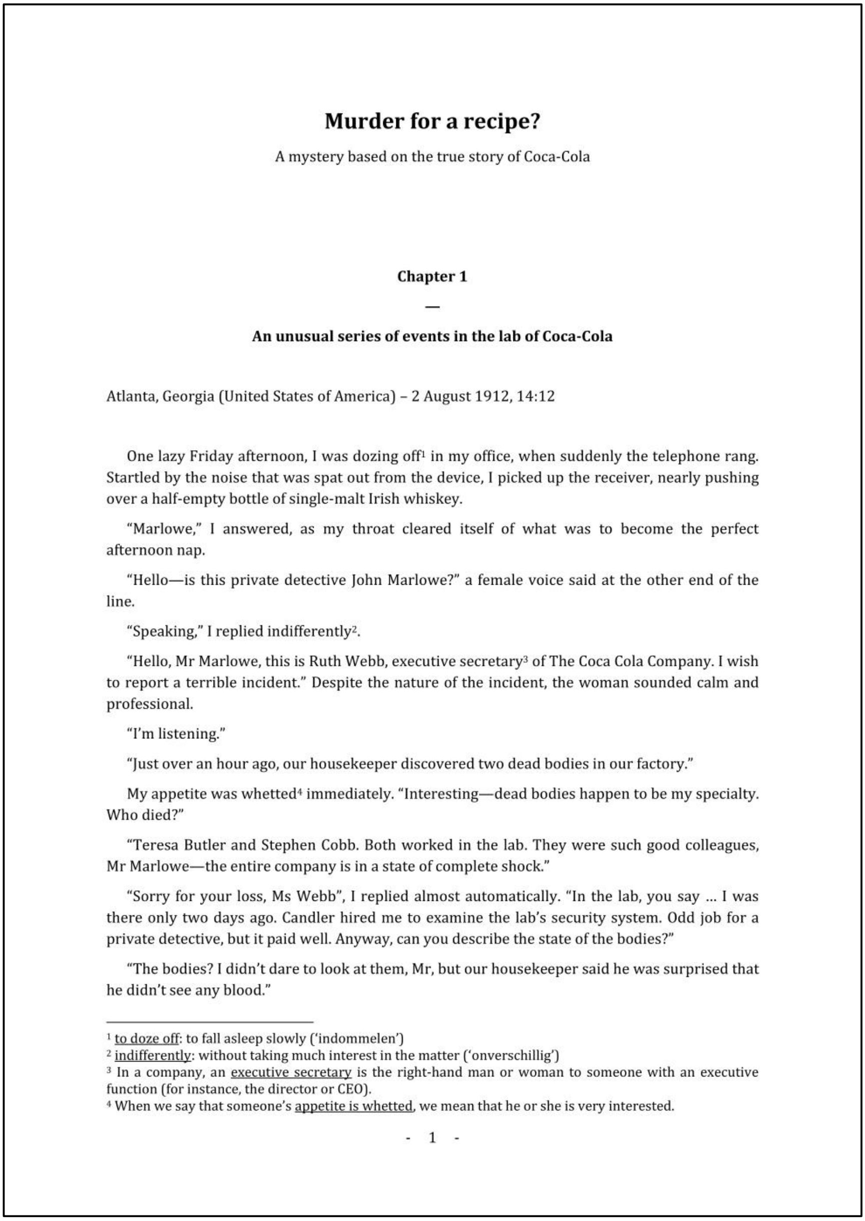
⇒ A preposition is not necessary.

[verb] *X* ***to her*** → [verb] ***her*** *X*

[verb] *X* ***for her*** → [verb] ***her*** *X*

Appendix 5

This appendix contains the mystery text read and discussed in class in study 4. Note: In this representation of the text, instances of the double object construction are underlined. This was not so in the learners' version of the text.



Murder for a recipe?

A mystery based on the true story of Coca-Cola

Chapter 1

—

An unusual series of events in the lab of Coca-Cola

Atlanta, Georgia (United States of America) – 2 August 1912, 14:12

One lazy Friday afternoon, I was dozing off[1] in my office, when suddenly the telephone rang. Startled by the noise that was spat out from the device, I picked up the receiver, nearly pushing over a half-empty bottle of single-malt Irish whiskey.

"Marlowe," I answered, as my throat cleared itself of what was to become the perfect afternoon nap.

"Hello—is this private detective John Marlowe?" a female voice said at the other end of the line.

"Speaking," I replied indifferently[2].

"Hello, Mr Marlowe, this is Ruth Webb, executive secretary[3] of The Coca Cola Company. I wish to report a terrible incident." Despite the nature of the incident, the woman sounded calm and professional.

"I'm listening."

"Just over an hour ago, our housekeeper discovered two dead bodies in our factory."

My appetite was whetted[4] immediately. "Interesting—dead bodies happen to be my specialty. Who died?"

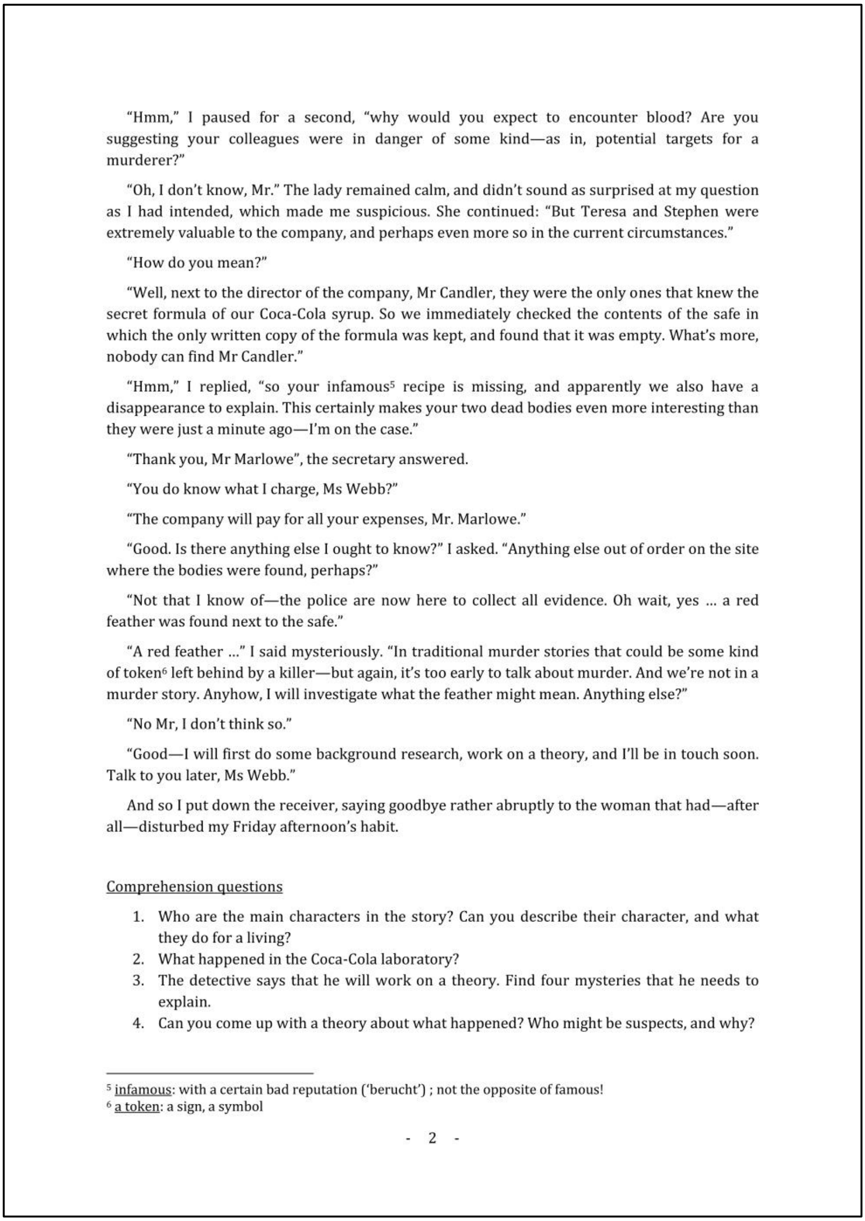
"Teresa Butler and Stephen Cobb. Both worked in the lab. They were such good colleagues, Mr Marlowe—the entire company is in a state of complete shock."

"Sorry for your loss, Ms Webb", I replied almost automatically. "In the lab, you say ... I was there only two days ago. Candler hired me to examine the lab's security system. Odd job for a private detective, but it paid well. Anyway, can you describe the state of the bodies?"

"The bodies? I didn't dare to look at them, Mr, but our housekeeper said he was surprised that he didn't see any blood."

[1] to doze off: to fall asleep slowly ('indommelen')
[2] indifferently: without taking much interest in the matter ('onverschillig')
[3] In a company, an executive secretary is the right-hand man or woman to someone with an executive function (for instance, the director or CEO).
[4] When we say that someone's appetite is whetted, we mean that he or she is very interested.

- 1 -



"Hmm," I paused for a second, "why would you expect to encounter blood? Are you suggesting your colleagues were in danger of some kind—as in, potential targets for a murderer?"

"Oh, I don't know, Mr." The lady remained calm, and didn't sound as surprised at my question as I had intended, which made me suspicious. She continued: "But Teresa and Stephen were extremely valuable to the company, and perhaps even more so in the current circumstances."

"How do you mean?"

"Well, next to the director of the company, Mr Candler, they were the only ones that knew the secret formula of our Coca-Cola syrup. So we immediately checked the contents of the safe in which the only written copy of the formula was kept, and found that it was empty. What's more, nobody can find Mr Candler."

"Hmm," I replied, "so your infamous[5] recipe is missing, and apparently we also have a disappearance to explain. This certainly makes your two dead bodies even more interesting than they were just a minute ago—I'm on the case."

"Thank you, Mr Marlowe", the secretary answered.

"You do know what I charge, Ms Webb?"

"The company will pay for all your expenses, Mr. Marlowe."

"Good. Is there anything else I ought to know?" I asked. "Anything else out of order on the site where the bodies were found, perhaps?"

"Not that I know of—the police are now here to collect all evidence. Oh wait, yes ... a red feather was found next to the safe."

"A red feather ..." I said mysteriously. "In traditional murder stories that could be some kind of token[6] left behind by a killer—but again, it's too early to talk about murder. And we're not in a murder story. Anyhow, I will investigate what the feather might mean. Anything else?"

"No Mr, I don't think so."

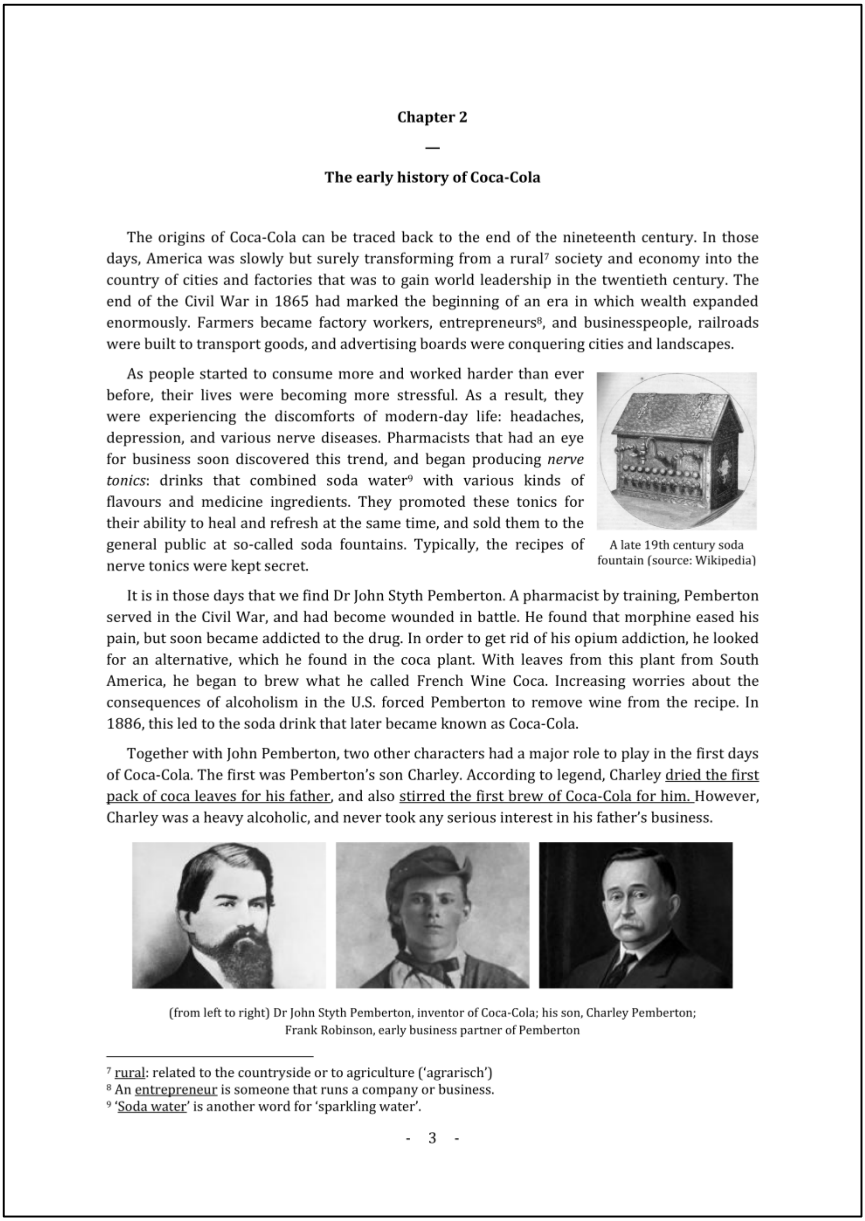
"Good—I will first do some background research, work on a theory, and I'll be in touch soon. Talk to you later, Ms Webb."

And so I put down the receiver, saying goodbye rather abruptly to the woman that had—after all—disturbed my Friday afternoon's habit.

Comprehension questions

1. Who are the main characters in the story? Can you describe their character, and what they do for a living?
2. What happened in the Coca-Cola laboratory?
3. The detective says that he will work on a theory. Find four mysteries that he needs to explain.
4. Can you come up with a theory about what happened? Who might be suspects, and why?

[5] infamous: with a certain bad reputation ('berucht') ; not the opposite of famous!
[6] a token: a sign, a symbol



Chapter 2

—

The early history of Coca-Cola

The origins of Coca-Cola can be traced back to the end of the nineteenth century. In those days, America was slowly but surely transforming from a rural[7] society and economy into the country of cities and factories that was to gain world leadership in the twentieth century. The end of the Civil War in 1865 had marked the beginning of an era in which wealth expanded enormously. Farmers became factory workers, entrepreneurs[8], and businesspeople, railroads were built to transport goods, and advertising boards were conquering cities and landscapes.

As people started to consume more and worked harder than ever before, their lives were becoming more stressful. As a result, they were experiencing the discomforts of modern-day life: headaches, depression, and various nerve diseases. Pharmacists that had an eye for business soon discovered this trend, and began producing *nerve tonics*: drinks that combined soda water[9] with various kinds of flavours and medicine ingredients. They promoted these tonics for their ability to heal and refresh at the same time, and sold them to the general public at so-called soda fountains. Typically, the recipes of nerve tonics were kept secret.

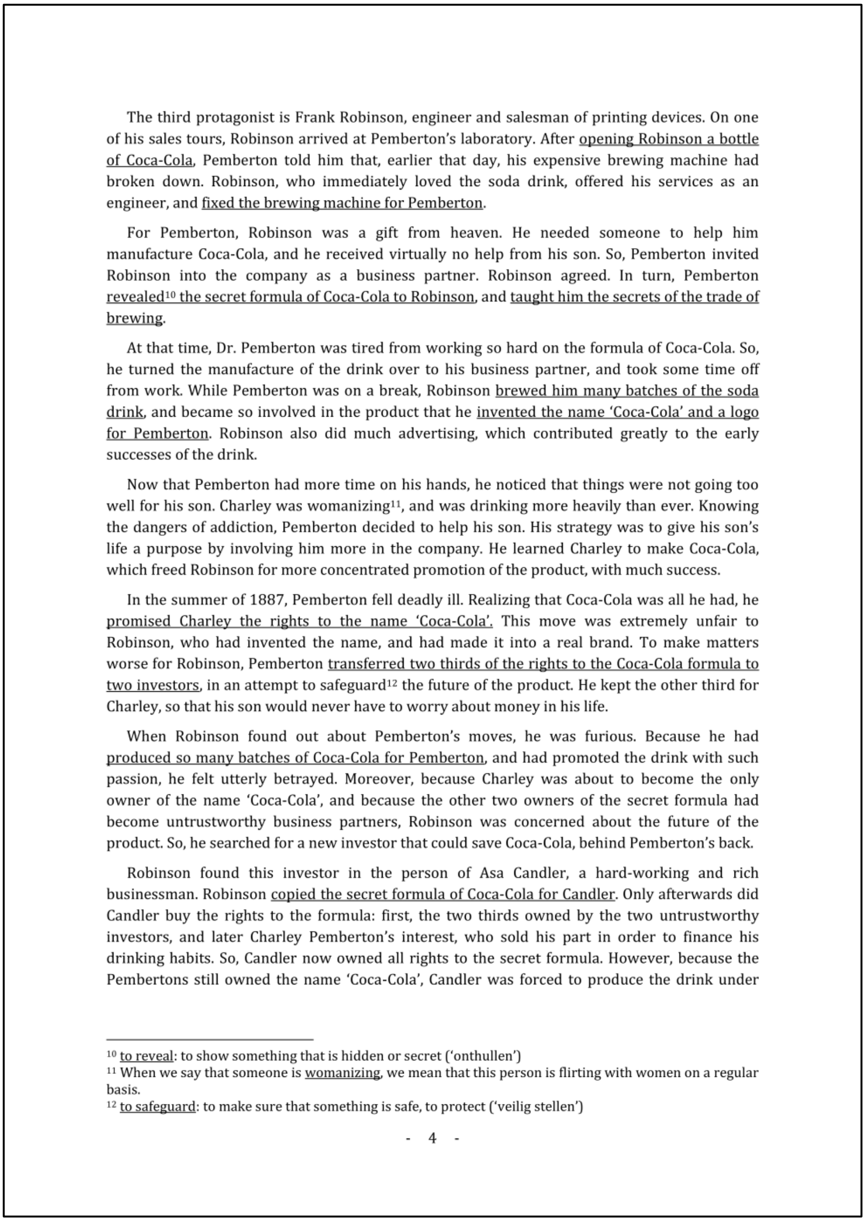
A late 19th century soda fountain (source: Wikipedia)

It is in those days that we find Dr John Styth Pemberton. A pharmacist by training, Pemberton served in the Civil War, and had become wounded in battle. He found that morphine eased his pain, but soon became addicted to the drug. In order to get rid of his opium addiction, he looked for an alternative, which he found in the coca plant. With leaves from this plant from South America, he began to brew what he called French Wine Coca. Increasing worries about the consequences of alcoholism in the U.S. forced Pemberton to remove wine from the recipe. In 1886, this led to the soda drink that later became known as Coca-Cola.

Together with John Pemberton, two other characters had a major role to play in the first days of Coca-Cola. The first was Pemberton's son Charley. According to legend, Charley dried the first pack of coca leaves for his father, and also stirred the first brew of Coca-Cola for him. However, Charley was a heavy alcoholic, and never took any serious interest in his father's business.

(from left to right) Dr John Styth Pemberton, inventor of Coca-Cola; his son, Charley Pemberton; Frank Robinson, early business partner of Pemberton

[7] rural: related to the countryside or to agriculture ('agrarisch')
[8] An entrepreneur is someone that runs a company or business.
[9] 'Soda water' is another word for 'sparkling water'.



The third protagonist is Frank Robinson, engineer and salesman of printing devices. On one of his sales tours, Robinson arrived at Pemberton's laboratory. After opening Robinson a bottle of Coca-Cola, Pemberton told him that, earlier that day, his expensive brewing machine had broken down. Robinson, who immediately loved the soda drink, offered his services as an engineer, and fixed the brewing machine for Pemberton.

For Pemberton, Robinson was a gift from heaven. He needed someone to help him manufacture Coca-Cola, and he received virtually no help from his son. So, Pemberton invited Robinson into the company as a business partner. Robinson agreed. In turn, Pemberton revealed[10] the secret formula of Coca-Cola to Robinson, and taught him the secrets of the trade of brewing.

At that time, Dr. Pemberton was tired from working so hard on the formula of Coca-Cola. So, he turned the manufacture of the drink over to his business partner, and took some time off from work. While Pemberton was on a break, Robinson brewed him many batches of the soda drink, and became so involved in the product that he invented the name 'Coca-Cola' and a logo for Pemberton. Robinson also did much advertising, which contributed greatly to the early successes of the drink.

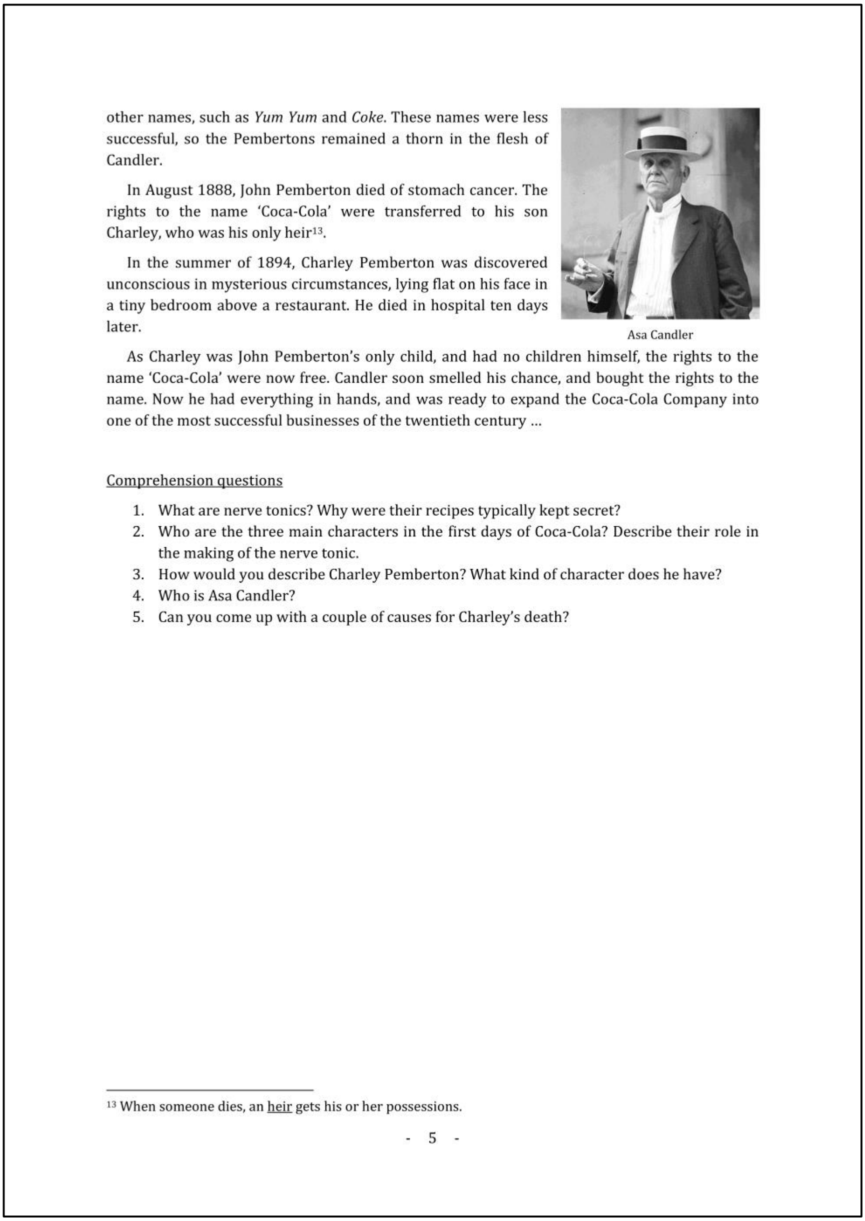
Now that Pemberton had more time on his hands, he noticed that things were not going too well for his son. Charley was womanizing[11], and was drinking more heavily than ever. Knowing the dangers of addiction, Pemberton decided to help his son. His strategy was to give his son's life a purpose by involving him more in the company. He learned Charley to make Coca-Cola, which freed Robinson for more concentrated promotion of the product, with much success.

In the summer of 1887, Pemberton fell deadly ill. Realizing that Coca-Cola was all he had, he promised Charley the rights to the name 'Coca-Cola'. This move was extremely unfair to Robinson, who had invented the name, and had made it into a real brand. To make matters worse for Robinson, Pemberton transferred two thirds of the rights to the Coca-Cola formula to two investors, in an attempt to safeguard[12] the future of the product. He kept the other third for Charley, so that his son would never have to worry about money in his life.

When Robinson found out about Pemberton's moves, he was furious. Because he had produced so many batches of Coca-Cola for Pemberton, and had promoted the drink with such passion, he felt utterly betrayed. Moreover, because Charley was about to become the only owner of the name 'Coca-Cola', and because the other two owners of the secret formula had become untrustworthy business partners, Robinson was concerned about the future of the product. So, he searched for a new investor that could save Coca-Cola, behind Pemberton's back.

Robinson found this investor in the person of Asa Candler, a hard-working and rich businessman. Robinson copied the secret formula of Coca-Cola for Candler. Only afterwards did Candler buy the rights to the formula: first, the two thirds owned by the two untrustworthy investors, and later Charley Pemberton's interest, who sold his part in order to finance his drinking habits. So, Candler now owned all rights to the secret formula. However, because the Pembertons still owned the name 'Coca-Cola', Candler was forced to produce the drink under

[10] to reveal: to show something that is hidden or secret ('onthullen')
[11] When we say that someone is womanizing, we mean that this person is flirting with women on a regular basis.
[12] to safeguard: to make sure that something is safe, to protect ('veilig stellen')



other names, such as *Yum Yum* and *Coke*. These names were less successful, so the Pembertons remained a thorn in the flesh of Candler.

In August 1888, John Pemberton died of stomach cancer. The rights to the name 'Coca-Cola' were transferred to his son Charley, who was his only heir[13].

In the summer of 1894, Charley Pemberton was discovered unconscious in mysterious circumstances, lying flat on his face in a tiny bedroom above a restaurant. He died in hospital ten days later.

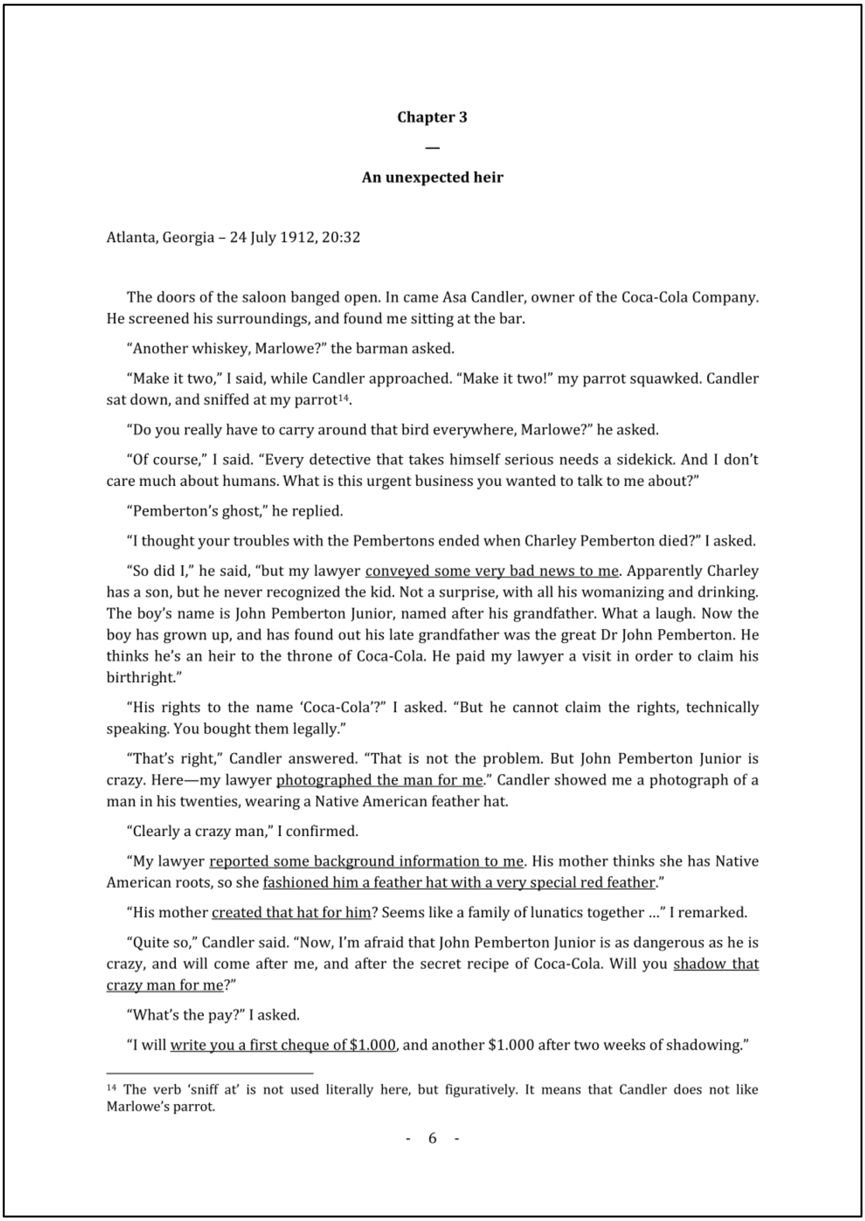
Asa Candler

As Charley was John Pemberton's only child, and had no children himself, the rights to the name 'Coca-Cola' were now free. Candler soon smelled his chance, and bought the rights to the name. Now he had everything in hands, and was ready to expand the Coca-Cola Company into one of the most successful businesses of the twentieth century ...

Comprehension questions

1. What are nerve tonics? Why were their recipes typically kept secret?
2. Who are the three main characters in the first days of Coca-Cola? Describe their role in the making of the nerve tonic.
3. How would you describe Charley Pemberton? What kind of character does he have?
4. Who is Asa Candler?
5. Can you come up with a couple of causes for Charley's death?

[13] When someone dies, an heir gets his or her possessions.



Chapter 3

—

An unexpected heir

Atlanta, Georgia – 24 July 1912, 20:32

The doors of the saloon banged open. In came Asa Candler, owner of the Coca-Cola Company. He screened his surroundings, and found me sitting at the bar.

"Another whiskey, Marlowe?" the barman asked.

"Make it two," I said, while Candler approached. "Make it two!" my parrot squawked. Candler sat down, and sniffed at my parrot[14].

"Do you really have to carry around that bird everywhere, Marlowe?" he asked.

"Of course," I said. "Every detective that takes himself serious needs a sidekick. And I don't care much about humans. What is this urgent business you wanted to talk to me about?"

"Pemberton's ghost," he replied.

"I thought your troubles with the Pembertons ended when Charley Pemberton died?" I asked.

"So did I," he said, "but my lawyer conveyed some very bad news to me. Apparently Charley has a son, but he never recognized the kid. Not a surprise, with all his womanizing and drinking. The boy's name is John Pemberton Junior, named after his grandfather. What a laugh. Now the boy has grown up, and has found out his late grandfather was the great Dr John Pemberton. He thinks he's an heir to the throne of Coca-Cola. He paid my lawyer a visit in order to claim his birthright."

"His rights to the name 'Coca-Cola'?" I asked. "But he cannot claim the rights, technically speaking. You bought them legally."

"That's right," Candler answered. "That is not the problem. But John Pemberton Junior is crazy. Here—my lawyer photographed the man for me." Candler showed me a photograph of a man in his twenties, wearing a Native American feather hat.

"Clearly a crazy man," I confirmed.

"My lawyer reported some background information to me. His mother thinks she has Native American roots, so she fashioned him a feather hat with a very special red feather."

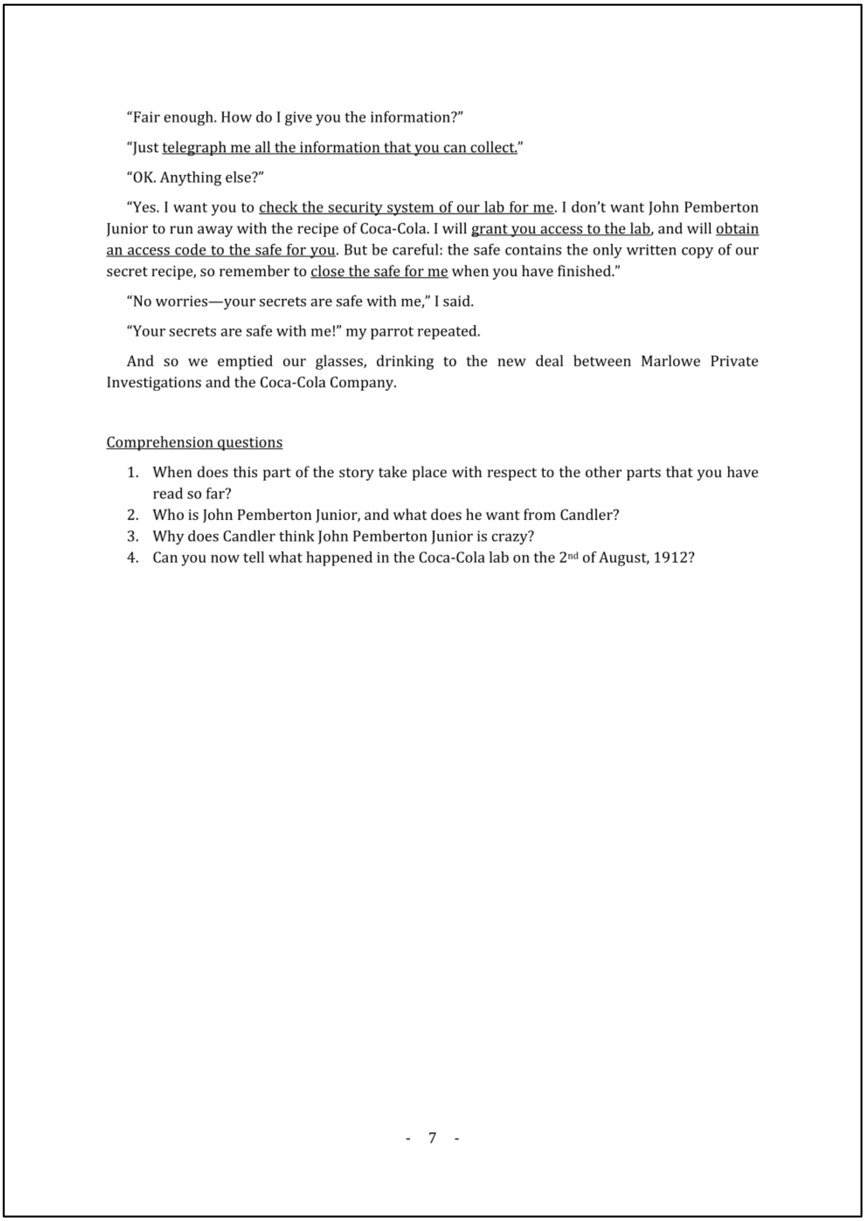
"His mother created that hat for him? Seems like a family of lunatics together …" I remarked.

"Quite so," Candler said. "Now, I'm afraid that John Pemberton Junior is as dangerous as he is crazy, and will come after me, and after the secret recipe of Coca-Cola. Will you shadow that crazy man for me?"

"What's the pay?" I asked.

"I will write you a first cheque of $1.000, and another $1.000 after two weeks of shadowing."

[14] The verb 'sniff at' is not used literally here, but figuratively. It means that Candler does not like Marlowe's parrot.



"Fair enough. How do I give you the information?"

"Just telegraph me all the information that you can collect."

"OK. Anything else?"

"Yes. I want you to check the security system of our lab for me. I don't want John Pemberton Junior to run away with the recipe of Coca-Cola. I will grant you access to the lab, and will obtain an access code to the safe for you. But be careful: the safe contains the only written copy of our secret recipe, so remember to close the safe for me when you have finished."

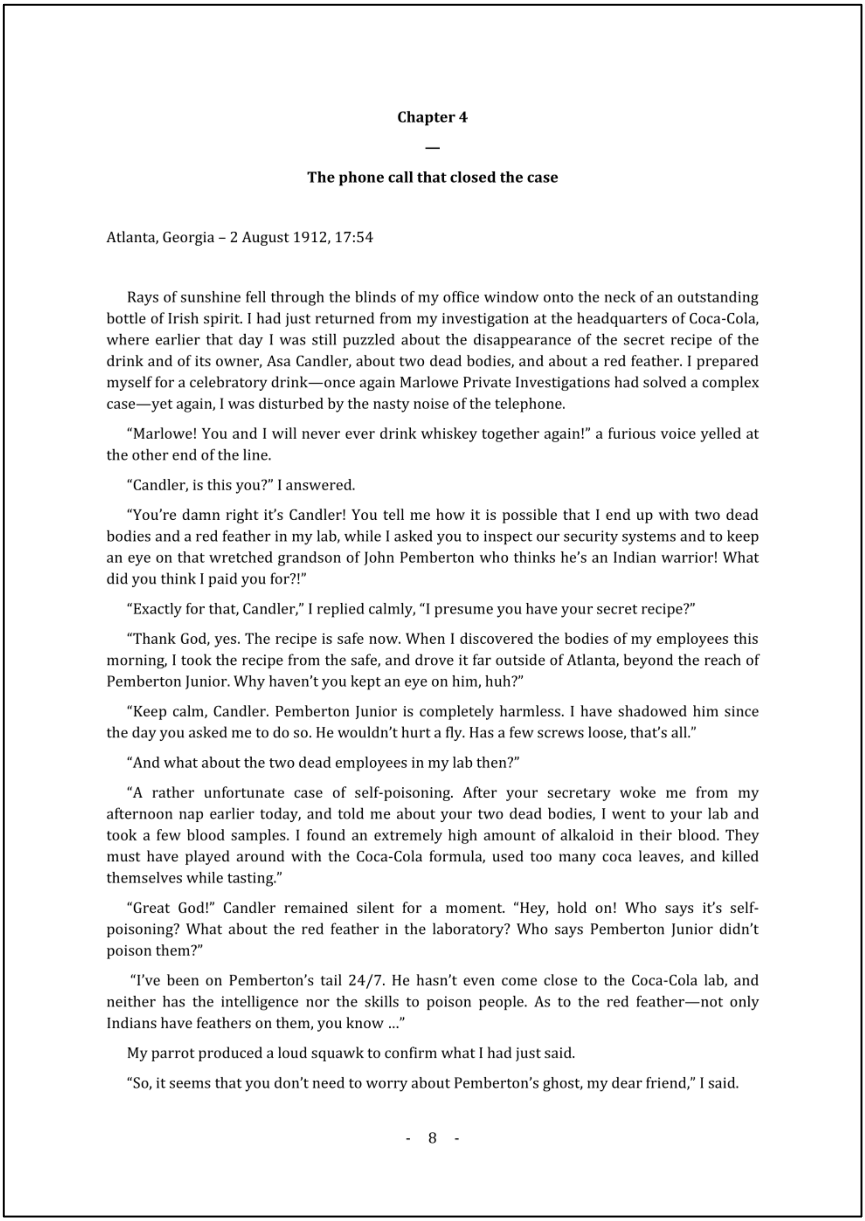
"No worries—your secrets are safe with me," I said.

"Your secrets are safe with me!" my parrot repeated.

And so we emptied our glasses, drinking to the new deal between Marlowe Private Investigations and the Coca-Cola Company.

Comprehension questions

1. When does this part of the story take place with respect to the other parts that you have read so far?
2. Who is John Pemberton Junior, and what does he want from Candler?
3. Why does Candler think John Pemberton Junior is crazy?
4. Can you now tell what happened in the Coca-Cola lab on the 2nd of August, 1912?



Chapter 4

—

The phone call that closed the case

Atlanta, Georgia – 2 August 1912, 17:54

Rays of sunshine fell through the blinds of my office window onto the neck of an outstanding bottle of Irish spirit. I had just returned from my investigation at the headquarters of Coca-Cola, where earlier that day I was still puzzled about the disappearance of the secret recipe of the drink and of its owner, Asa Candler, about two dead bodies, and about a red feather. I prepared myself for a celebratory drink—once again Marlowe Private Investigations had solved a complex case—yet again, I was disturbed by the nasty noise of the telephone.

“Marlowe! You and I will never ever drink whiskey together again!” a furious voice yelled at the other end of the line.

“Candler, is this you?” I answered.

“You’re damn right it’s Candler! You tell me how it is possible that I end up with two dead bodies and a red feather in my lab, while I asked you to inspect our security systems and to keep an eye on that wretched grandson of John Pemberton who thinks he’s an Indian warrior! What did you think I paid you for?!”

“Exactly for that, Candler,” I replied calmly, “I presume you have your secret recipe?”

“Thank God, yes. The recipe is safe now. When I discovered the bodies of my employees this morning, I took the recipe from the safe, and drove it far outside of Atlanta, beyond the reach of Pemberton Junior. Why haven’t you kept an eye on him, huh?”

“Keep calm, Candler. Pemberton Junior is completely harmless. I have shadowed him since the day you asked me to do so. He wouldn’t hurt a fly. Has a few screws loose, that’s all.”

“And what about the two dead employees in my lab then?”

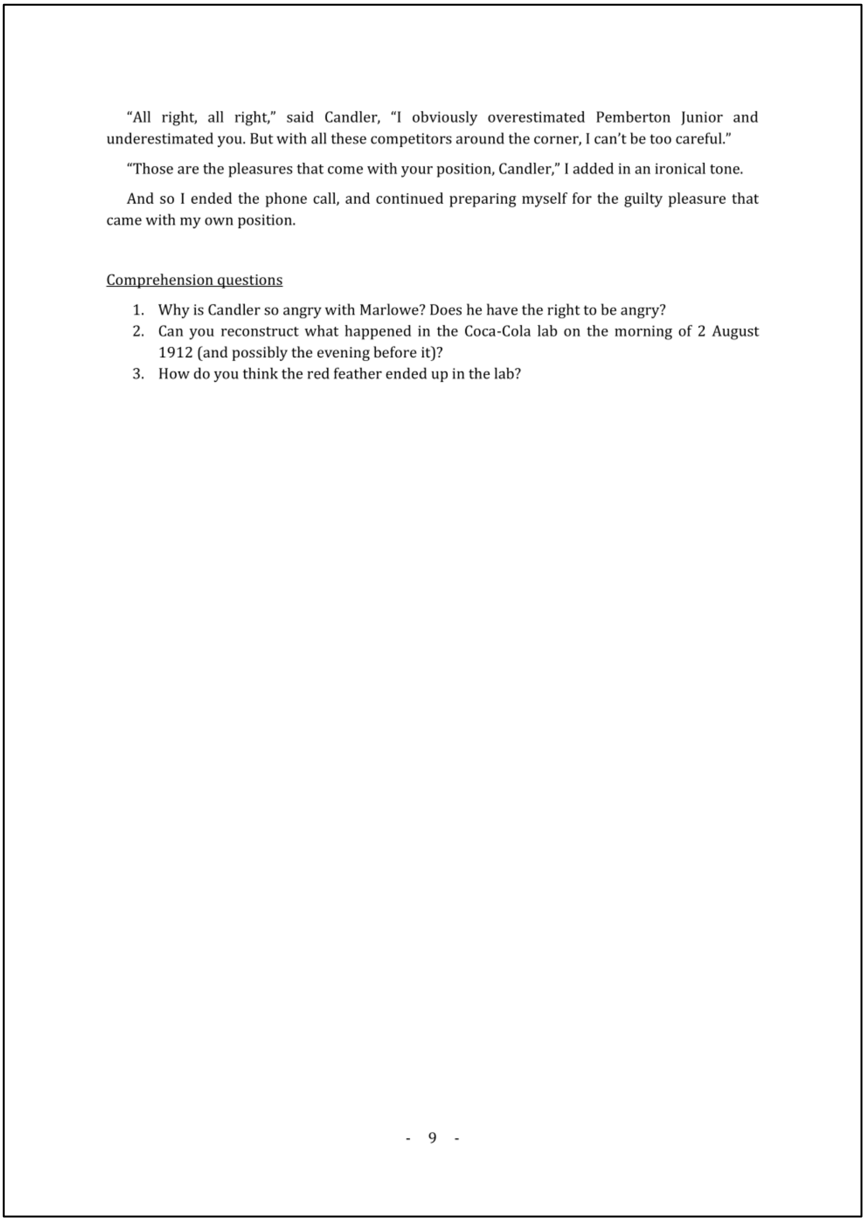
“A rather unfortunate case of self-poisoning. After your secretary woke me from my afternoon nap earlier today, and told me about your two dead bodies, I went to your lab and took a few blood samples. I found an extremely high amount of alkaloid in their blood. They must have played around with the Coca-Cola formula, used too many coca leaves, and killed themselves while tasting.”

“Great God!” Candler remained silent for a moment. “Hey, hold on! Who says it’s self-poisoning? What about the red feather in the laboratory? Who says Pemberton Junior didn’t poison them?”

“I’ve been on Pemberton’s tail 24/7. He hasn’t even come close to the Coca-Cola lab, and neither has the intelligence nor the skills to poison people. As to the red feather—not only Indians have feathers on them, you know …”

My parrot produced a loud squawk to confirm what I had just said.

“So, it seems that you don’t need to worry about Pemberton’s ghost, my dear friend,” I said.



“All right, all right,” said Candler, “I obviously overestimated Pemberton Junior and underestimated you. But with all these competitors around the corner, I can’t be too careful.”

“Those are the pleasures that come with your position, Candler,” I added in an ironical tone.

And so I ended the phone call, and continued preparing myself for the guilty pleasure that came with my own position.

Comprehension questions

1. Why is Candler so angry with Marlowe? Does he have the right to be angry?
2. Can you reconstruct what happened in the Coca-Cola lab on the morning of 2 August 1912 (and possibly the evening before it)?
3. How do you think the red feather ended up in the lab?

Appendix 6

This appendix contains the items for the TGJT used in study 4.

set	item	problem	construction	item type
A	**Charley brewed the least kettles of Coca-Cola.*	quantifiers	**least* + countable	practice1
A	**Yes, I stirred father the first brew of Coca-Cola.*	double object construction	*V no transfer of possession (for) + NP + NP	practice1
A	*She offered him a job in the accountancy department.*	double object construction	V polysyllabic, initial stress (to) + NP + NP	transfer
A	**Bad luck. This area has only little hotels.*	quantifiers	**little* + countable	transfer
A	**The company announced us the name of their new smartphone product.*	double object construction	*V polysyllabic, final stress (to) + NP + NP	transfer
A	**Pemberton revealed me the secret formula.*	double object construction	*V polysyllabic, final stress (to) + NP + NP	practice1
A	*We don't have many books.*	quantifiers	*many* + countable	transfer
A	**My car is more faster and more powerful than your car.*	comparatives	N/A	distractor
A	*Charley has fewer shares in the company.*	quantifiers	*fewer* + countable	practice1
B	**Simply out of love, Forrest Gump ran Jenny a thousand miles.*	double object construction	*V no transfer of possession (for) + NP + NP	transfer
B	**Robinson copied Candler the secret recipe.*	double object construction	*V no transfer of possession (for) + NP + NP	practice1
B	**Tomorrow will be warm with less showers.*	quantifiers	**less* + countable	transfer
B	**He didn't get much presents.*	quantifiers	**much* + countable	transfer
B	**The boys went to bed late last night, is it?*	question tags	N/A	distractor
B	**The teacher explained Michael the answer.*	double object construction	*V polysyllabic, final stress (to) + NP + NP	transfer
B	**His mother created him a special Indian hat with a red feather.*	double object construction	*V polysyllabic, final stress (for) + NP + NP	practice2
B	**Coca-Cola has only little secrets for me.*	quantifiers	**little* + countable	practice1
B	*Unfortunately, we have few brewing machines similar to this one.*	quantifiers	*few* + countable	practice1
C	**He doesn't have much rights.*	quantifiers	**much* + countable	practice1
C	*Last month, I saw few accidents on the M5.*	quantifiers	*few* + countable	transfer
C	**Will you shadow me that dangerous man?*	double object construction	*V no transfer of possession (for) + NP + NP	practice2
C	**I have had less jobs than you have.*	quantifiers	**less* + countable	transfer
C	**Mother cleaned Thomas one of the bedroom windows.*	double object construction	*V no transfer of possession (for) + NP + NP	transfer

C	_*The engineer constructed them a better engine._	double object construction	*V polysyllabic, final stress (for) + NP + NP	transfer
C	_*Our lab only has little security problems._	quantifiers	*_little_ + countable	practice2
C	_I will write you a first cheque of $1.000._	double object construction	V monosyllabic (for) + NP + NP	practice2
C	_*Did Martin visited his father?_	yes/no questions	N/A	distractor
D	_*I've seen too much dead bodies in one day!_	quantifiers	*_much_ + countable	practice2
D	_*The lawyer described him the problem._	double object construction	*V polysyllabic, final stress (to) + NP + NP	transfer
D	_*Even smoking little cigarettes is bad for your health._	quantifiers	*_little_ + countable	transfer
D	_*Robinson fixed me the brewing machine._	double object construction	*V no transfer of possession (for) + NP + NP	practice1
D	_We send fewer planes to Bagdad these days._	quantifiers	_fewer_ + countable	transfer
D	_*Young Indians have the least feathers._	quantifiers	*_least_ + countable	practice2
D	_I will grant you access to the lab._	double object construction	V monosyllabic (to) + NP + NP	practice2
D	_*If he hadn't come to New Zealand, he will stay in Japan._	unreal conditionals	N/A	distractor
D	_*Because Sara was pregnant, John carried her one of the bags._	double object construction	*V no transfer of possession (for) + NP + NP	transfer
E	_*Will you check me the security system of our lab?_	double object construction	*V no transfer of possession (for) + NP + NP	practice2
E	_*She designed him a very fancy new coat._	double object construction	*V polysyllabic, final stress (for) + NP + NP	transfer
E	_If he had bought a ticket, he might have won the prize._	unreal conditionals	N/A	distractor
E	_*How much loaves of bread do you want?_	quantifiers	*_much_ + countable	transfer
E	_Father promised me the rights to the name 'Coca-Cola'._	double object construction	V polysyllabic, initial stress (to) + NP + NP	practice1
E	_I hope to make fewer phone calls to you in the future._	quantifiers	_fewer_ + countable	practice2
E	_*To scare his daughter, he pricked her a balloon._	double object construction	*V no transfer of possession (for) + NP + NP	transfer
E	_*I want to be the one who makes the least mistakes._	quantifiers	*_least_ + countable	transfer
E	_*Copies of Coca-Cola use less ingredients._	quantifiers	*_less_ + countable	practice1
F	_*Because the engine of Sarah's Toyota had broken down, the mechanic towed her the car._	double object construction	*V no transfer of possession (for) + NP + NP	transfer
F	_*She wanted to know why had he studied German._	embedded questions	N/A	distractor
F	_He sold the fewest bottles._	quantifiers	_fewest_ + countable	practice1
F	_*I don't have much screws loose._	quantifiers	*_much_ + countable	practice2
F	_*Brasil missed less penalties than Argentina._	quantifiers	*_less_ + countable	transfer
F	_*James selected her an 18-carat gold watch._	double object construction	*V polysyllabic, final stress (for) + NP + NP	transfer
F	_*My lawyer reported me some background information._	double object construction	*V polysyllabic, final stress (to) + NP + NP	practice2

F	**Please close me the safe when you have finished.*	double object construction	*V no transfer of possession (for) + NP + NP	practice2
F	**This garden only has very little flowers left.*	quantifiers	**little* + countable	transfer

Appendix 7

This appendix contains the items for the WDCT used in study 4.

construction	**items with nouns from practice sessions**	**transfer items**
few + countable	Your friend has a rough time in his relationship with a girl, and asks you for advice. You pretend to be someone who knows everything about women, and say: *You've come to the right man. For me, women have very* ___________. ! You must use the word 'secrets'.	You visit a friend, who has a garden full of flowers. You have recently moved to a flat, and your friend doesn't know this yet. You say to your friend: *My balcony is my garden. In comparison with you, I have very* _______________________. ! You must use the underlined words.
fewer + countable	You work as a pharmacist for Johnson & Johnson, and have just improved the recipe of a painkiller. You are in your boss's office, and want to convince him to produce your improved version of the painkiller. You think that you can convince him by saying that the quantity of ingredients needed to make the painkiller is much smaller now. You say: *My new recipe is better than the old one, because it uses* __________. ! You must use the underlined words.	-
fewest + countable	-.	You present a quiz on national television. The show is coming to an end, and the jury has just counted the mistakes that each team made. You announce the winning team: *Team B wins, because they made the* _________________________ . ! You must use the underlined words.

V$^{\text{polysyllabic, initial stress}}$ + NP + NP	-	You are the boss of an important publishing company of pop music. One of your employees tells you about a very promising band in Moscow. You want to take a plane to Moscow to speak to the band. You say: *Brilliant, I want to see them. Can you charter* ____________ *?* ! You must use the underlined words. ! You have to use the word 'me', and may only use a preposition when you really think it is necessary.
V$^{\text{polysyllabic, final stress}}$ + NP + PP	You work as a telephone receptionist for traffic assistance. Someone calls you to say that he has been part of a car accident. You want to know whether the police already know about the car accident. You say: *What about the police? Have you reported* __________*?* ! You must use the underlined words. ! You have to use the word 'them', and may only use a preposition when you really think it is necessary.	Your karate trainer has just taught you a trick on how to escape when your opponent has forced you on your back. You don't understand, and ask him to repeat the escape trick. You say: *I'm sorry, I don't understand. Can you explain* _____________ *?* ! You must use the underlined words. ! You have to use the word 'me', and may only use a preposition when you really think it is necessary.
V$^{\text{no transfer of possession}}$ + NP + PP	You talk to a friend who is really good at computers. He tells you that he has a new hobby: he buys broken tablet computers on eBay at a very low price, fixes them, and then sells them to people for half the price of new tablets. You tell your friend that recently, your nephew broke your tablet. You ask him whether he wants to repair that broken tablet computer: *Could you fix* ___________________ *?* ! You must use the underlined words. ! You have to use the word 'me', and may only use a preposition when you really think it is necessary.	-

Appendix 8

This appendix contains the items for the OEIT used in study 4

Practice items:

Dr. John Pemberton was a better businessman than his son Charley.

**In the 19th century, Dr. John Pemberton has invented the name Coca-Cola.*

On the morning of August 2nd, the secretary killed her colleagues.

**Charley Pemberton was more smarter than his son Junior.*

Test items:

Asa Candler thinks he has too many competitors.

**Charley created his son a special Indian hat with a red feather.*

**Charley had less shares of Coca-Cola than Frank Robinson.*

Dr. John Pemberton promised his son Charley the rights to the name Coca-Cola.

**Stephen Cobb died because he used too much coca leaves.*

**Candler stirred Pemberton the first brew of Coca-Cola.*

**Charley died because he drank too much nerve tonics.*

Candler has granted me access to the lab.

**Charley revealed Candler the secret recipe of Coca-Cola.*

**Pemberton Junior has less screws loose than me.*

**Robinson fixed Pemberton the brewing machine.*

If I stop drinking, I will have fewer headaches.

Extended table of contents

Thank you/dank jullie wel

Piet—dank je voor je onvoorwaardelijke steun en oneindige enthousiasme; voor het toepassen van je mensenkennis in moeilijkere tijden; om me te omringen met een team en omgeving waarvoor elke interdisciplinaire denker onmiddellijk tekent; om zaadjes te planten en water te geven, elke dag opnieuw.

Geraldine—dank je om mijn project inhoudelijk en methodologisch mee op poten te zetten in de periode dat jij mijn copromotor was; voor de stimulerende dialogen volgens de Socratische methode; om me bij de les te houden.

Kris—dank je om je in het laatste anderhalve jaar van mijn project zo snel in te werken als copromotor; voor je onbevangen en altijd heel ter zake commentaren; voor je positieve feedback.

Steve—thank you for diving into our special issue adventure; for your relentless enthusiasm; for good times at the Graslei in Gent; for falling asleep at the keyboard with me in the small hours.

Jozef—dank je voor de zovele kansen die je me gaf om te groeien op de projecten die mee aan de grondslag liggen van dit doctoraat; om mij de passie bij te brengen voor (én kritische houding tegenover) onze discipline; voor je stimulerende feedback op mijn project.

Elke—dank je voor je constructieve commentaren tijdens één van onze ITEC-seminaries; om deel te willen uitmaken van mijn jury.

Kurt—dank je om de verdediging van mijn proefschrift te willen voorzitten; voor je woorden van aanmoediging in de Spina en daarbuiten.

Mieke, Wilfried, en Wim—dank je om me de weg te wijzen en bij te staan in de wondere wereld van de methodologie en statistiek.

Brecht M, Brecht S, Jan, Joepie, Piet B—dank je om me te laten spelen in en met de IT-infrastructuur van Kulak.

David, Erik, Geraldine, Hans, Ignace, Igor, Jeroen, Kasper, Kathleen, Ken, Koen, Kristof, Lode, Martin, Nathalie, Sem, Swen, Sylke, Thierry—dank je voor de aangename en inspirerende samenwerking binnen het kader van het LLINGO project.

Hans, Ruben L, en Ruben T—dank je voor jullie frisse ideeën en handen-uit-de-mouwen aanpak op ons 'Guided Chat' project.

Andrew, Ann-Sophie, Catia, Febe, Helmer, Ilana, Marijn, Polina, Sake, Scott, Tom, Ton, en Vanja—thank you for the inspiring collaboration on the GOBL project.

Bieke, Jan-Henk, en Yorick—dank je voor de aangename en inspirerende samenwerking binnen het kader van het MIGAME project.

Johan—dank je voor de kostbare tijd die ik van je lessen mocht afsnoepen; voor je tips in verband met het onderwijzen van Engelse grammatica; voor je schitterende acteursprestatie als *Adam Albright*.

Björn, Dieter, Hannes, Joke, Natascha, en Sarah—dank je om mij in alle vertrouwen met jullie leerlingen te laten werken.

Stefan—dank je om steeds in de bres te springen op de vele projecten; voor je *speelvogelarij*, die van hard werken een echt plezier maakte; je was mijn spitsbroeder; Kulak zal nooit hetzelfde zijn zonder jou.

Ann, Ann-Sophie, Antoine, Barbara, Carmen, Caroline, Fien, Hans, Igor, Maribel, Marie, Mark, Martin, Mieke, Patrick, Tommy, Ruben L, Sylke, Wilfried, en Wim—dank jullie om van ITEC een fijne plek te maken om te werken.

Mama en papa—van jullie heb ik alles meegekregen om dit project tot een goed einde te brengen: mijn liefde voor de Engelse taal, literatuur, en cultuur, wat leerkrachtenbloed, mijn eerste cursus programmeren op negenjarige (?) leeftijd, doorzettingsvermogen, de *human resources* in de vorm van jullie leerlingen en jullie eigen expertise; dank voor dit alles en voor jullie niet aflatende steun.

Katrien, Marieke, en Tiele—dit project doorkruiste ons leven niet op een ideaal moment; dank jullie om er begrip voor op te brengen dat ik er soms niet helemaal was; dank voor jullie geduld, zorgzaamheid en liefde.

Katholieke Universiteit Leuven
Faculteit Letteren
Onderzoekseenheid Taalkunde

KU Leuven KULAK
Subfaculteit Letteren
iMinds - ITEC (Interactive Technologies)