The Dissertation Committee for Arthur Herman Wendorf II certifies that this is the approved version of the following dissertation:

El Mundo de Comida: The Relative Effectiveness of Digital Game Feedback and Classroom Feedback in Helping Students Learn Spanish Food Vocabulary

Committee:

Dale Koike, Supervisor

Carl Blyth

Orlando Kelm

Julie Sykes

Almeida Jacqueline Toribio

El Mundo de Comida: The Relative Effectiveness of Digital Game Feedback and Classroom Feedback in Helping Students Learn Spanish Food Vocabulary

by

Arthur Herman Wendorf II, B.A., M.A., M.A.

Dissertation

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

The University of Texas at Austin

December 2014

Dedication

Dedicated to everyone who has helped me get to this point.

Acknowledgements

I gratefully acknowledge the help provided to me by so many people during my journey to this point. This dissertation was many things, but it was by no means a one-man effort.

First and foremost, I am extremely grateful for the innumerable hours, the in-depth feedback, and, perhaps most of all, the incalculable patience of my supervisor Dale Koike! Without her there is no way I would have ever completed this dissertation! Thank you Dale!

I am also immeasurably grateful for the other members of my dissertation committee, who took time out of their busy schedules to help, guide and encourage me! Thank you, in alphabetical order because I can't put all your names down at the same time, Carl Blyth, Orlando Kelm, Julie Sykes, and Jacqueline Toribio!

I am also especially thankful for the efforts of all of those who have directly helped me work on this dissertation including (in alphabetical order by last name): Sally Amen (Statistics Counseling), Pablo Duran (Voice Actor), John Edwards (Programming Advice), Erika Hale (Statistics Counseling), Royce Kimmons (Programming Assistance), Michael J. Mahometa (Statistics Counseling), Robert Sauveur (Reader), Jesse Schell (Content Consultation), Jenny Steen (Editing), Steve Thorne (Design Consultation), and Cecilia Tocaimaza-Hatch (Reader).

Finally, I also gratefully acknowledge the indispensable help of the undergraduate instructors who helped with recruitment, the game playtesters who provided feedback on initial versions of the game and, of course, all of the undergraduates who played MuCo.

El Mundo de Comida: The Relative Effectiveness of Digital Game Feedback and Classroom Feedback in Helping Students Learn Spanish Food Vocabulary

Arthur Herman Wendorf II, PhD

The University of Texas at Austin, 2014

Supervisor: Dale Koike

Feedback has been defined as "helpful information or criticism that is given to someone to say what can be done to improve a performance, product, etc." (Merriam-Webster, 2014) Within the field of Second Language Acquisition (SLA) researchers have shown that language learners acquire languages best when they are provided with feedback (Gass & Selinker, 2008; Loewen, 2012). Because of the importance of feedback to the language learning process, there is an ongoing line of investigation that seeks to determine whether differences in how and when feedback is provided lead to different results in acquisition (Loewen, 2012). To date this research has primarily been focused on comparing the effectiveness of the different types of feedback that naturally occur within language classrooms, as identified by such classic studies as Lyster and Ranta (1997; Bargiela, 2003). However, there are other possible approaches to feedback than those that naturally occur within the language classroom. One of these alternatives is the approach to feedback used in digital games. Similar to what is found in the field of SLA, within the field of digital game research it has been established that feedback is important for successful learning (Schell, 2008). Nevertheless, to date no research has been

conducted which compares the SLA approach to feedback and the digital game approach to feedback in order to determine which would lead to better language acquisition within a digital game. Answering this question is the goal of the present dissertation.

In order to answer this question I created two versions of a digital game, called *"Mundo de Comida"* (MuCo) 'World of Food', which is designed to help novice Spanish learners acquire food vocabulary. One version of the game employs feedback strategies based on the most commonly employed feedback used in Spanish language classes, while the other uses feedback designed according to the most commonly used feedback mechanisms in commercial digital games. A comparison of the vocabulary gains according to feedback type allows us to see which type of feedback seems to help learners of Spanish acquire vocabulary within the context of MuCo.

The findings indicate that MuCo does indeed help participants acquire food vocabulary. However, there is no significant difference in the effectiveness of the two different feedback types, which is likely due to the fact that both feedback types have been refined within their respective environments. Nevertheless, there is evidence to suggest that participants found the game that contained the digital game-style feedback to be more game-like than the other version. It was also found that, for several participants, MuCo did motivate them in the sense that they played more of the game than was required. Finally, there was no significant effect found for the participants' self-reported gaming habits, personalities, or motivation. These findings suggest that well-designed digital games can help learners acquire Spanish vocabulary, and that the impact of differences among participants is negligible when the game is well designed.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

In the early twentieth century, within the field of computer science, the term "feedback" emerged as a way to identify the information that a computer receives about the results of the actions that it has taken (Allwood, Nivre, & Ahlsén, 1992; Harper, 2001). By the mid-twentieth century 'feedback' had begun to emerge as a term in other fields as well, including the field of linguistics. Within linguistics, the term 'feedback' is generally used when discussing oral interaction. In this context, feedback occurs when the interlocutor responds to what is being communicated in a way that is detected by the speaker. A simple example is the telling of a joke; when a person tells a joke and the listeners laugh, the laughter is the feedback. The listeners produced the laughter as a direct result of the joke and the laughter provides information to the speaker about how the joke was received by the audience. Because of feedback, the source of any given output is not only able to observe the effect of the output produced, but also to adjust future output in order to obtain better results. For example, if a speaker tells a bad joke, the listeners may respond with feedback in the form of blank stares instead of laughter, and the speaker may thus decide not to tell that particular joke again.

The prototypical feedback that occurs naturally within language classrooms follows certain principles, as detailed in 1.1, regardless of which component of the language (e.g. vocabulary, grammar, pragmatics, etc.) is being taught. In this study vocabulary was chosen because the types of tasks recommended for vocabulary instruction (see 2.2 to follow) are the most easily duplicated within the context of a digital game. However, language learning is not restricted to classrooms, and can even occur in digital games, which are growing in popularity in the classroom teaching

context. Nevertheless, the feedback that naturally occurs in digital games is different from that which we find in language classrooms, as indicated in 1.2. Thus, the question arises regarding which of these two types of feedback in a digital game best helps language learners acquire a second language.

The central focus of this dissertation is the question of feedback within the field of Second Language Acquisition (SLA)[1] when this feedback is administered within a digital game. That is, I ask whether feedback can effectively promote second language vocabulary learning when it is provided in the context of a digital game in the way that classroom feedback is given, or in the form of typical digital game feedback. Accordingly, in this first chapter I establish the foundation for the present dissertation by first closely examining the theories and research related to feedback within SLA. Next, I examine the literature relevant to the use of feedback in digital games. Having established the general principles of the feedback that naturally occurs within both environments, I then compare and contrast the way feedback is most commonly provided within these two environments. Subsequently, I look at the previous literature that has touched on the use of feedback within digital games that are used to teach a second language. Finally, I conclude this chapter by outlining the remainder of the dissertation.

1.1. Feedback in SLA

Within this section we examine the role that feedback plays in the context of SLA. Because the purpose of the present study is largely applied in nature, a particular emphasis is placed on the role of feedback within SLA classroom methodology. We begin by looking at research performed to date on feedback in the second language (L2)

[1] See Appendix C for a complete list of all acronyms used herein.

classroom. The different aspects of this phenomenon that have been researched, as well as the findings resultant from these studies, are expounded upon in detail.

In SLA, feedback occurs in response to a Second Language Learner (SLL) producing some form of linguistic output, whether written or spoken. When interlocutors react to this linguistic production, they are providing the learners with feedback on their production (Cook, 2008). Nevertheless, neither the feedback thus provided nor the circumstances in which it is provided are by any means homogenous, and they may be differentiated in several ways. The purpose of the present section is to identify the principles most common to the majority of L2 feedback as documented by observational studies and as prescribed in theoretical works. In doing so, I identify three principles that are commonly used in language classrooms related to the generation of classroom feedback. These three principles are listed here.

CF 1: Classroom feedback is provided in response to an error.

CF 2: Classroom feedback is linguistic.

CF 3: Class goes on regardless of whether classroom feedback is successful.

In the following subsections I discuss these principles in detail and how they affect the production of feedback within the classroom.

1.1.1. Positive vs. negative feedback in SLA

The most common distinction made in SLA literature about feedback is related to the presence or absence of an error in the original learner production. If the feedback does not react to an error in the learner's production, and more particularly if the feedback affirms that the learner's production was error free, then that feedback is called 'positive feedback' (e.g. *bien* 'good'). On the other hand, feedback can also indicate that the interlocutor perceived a linguistic error in the learner's production. When this happens,

the interlocutor produces what is commonly known as 'negative' or 'corrective' feedback (e.g. *¿cómo?* 'what?'). By far the most commonly researched of these two forms of feedback is the latter (Cornillie, Clarebout, & Desmet, 2012a).

The interest in negative feedback stems from its importance in the language acquisition process. Although research has consistently shown that language learners make greater short-term improvement with language use when negative feedback is provided in laboratory settings (Carroll & Swain, 1993; Michael H Long, Inagaki, & Ortega, 1998; Takimoto, 2006), classroom settings (Havranek, 2002; Loewen & Philp, 2006; Lyster & Ranta, 1997), and even in computer-based language learning environments (Brandl, 1995; Heift, 2004; Nagata, 1993; Pujolà, 2001), there have yet to be any longitudinal studies performed that confirm the long-term benefits of negative feedback. Moreover, depending on a researcher's particular theoretical approach to SLA, they may hold that negative feedback can impede learning because it may contribute to learner anxiety (Truscott, 1996), or have absolutely no bearing on language acquisition because it targets explicit learning, while acquisition is taken to be a separate, implicit process (Krashen, 1982). On the other hand, other researchers point out that explicit knowledge may eventually become implicit knowledge (DeKeyser, 2008), and that language learners report a strong preference for receiving corrective feedback (Cathcart & Olsen, 1976; Chenoweth, Day, Chun, & Luppescu, 1983; Hedgcock & Lefkowitz, 1994; Radecki & Swales, 1988; Schulz, 2001).

The research surrounding Long's (1996) Interaction Hypothesis (IH), in particular, underscored the importance of negative feedback in helping learners acquire a TL. And the central importance of feedback is maintained in the SLA theory of Communicative Language Teaching, which is based on IH and is at the present one of the most widely accepted theoretical perspectives on SLA. I use the IH herein in spite of the

fact that it is somewhat dated because it allows for a clear categorization of feedback types, although it is limited to only certain kinds of feedback and therefore is narrow in scope. Moreover, the focus of the present work is not on which theoretical framework is most accurate, but rather on how feedback can be implemented regardless of which theoretical framework is being used. In the IH, Long submitted that negative feedback from an interlocutor is indispensable because it allows learners to identify errors in their interlanguage. He argued that SLLs could not overcome these errors if the errors were not first brought to their attention by the negative feedback. Indeed, subsequent research supports the idea that negative feedback benefits learners, as Long hypothesized (Gass & Selinker, 2008; Loewen, 2012). One such study was conducted by Dilans (2010), who compared the vocabulary retention for three different groups: one that received no feedback; one that received negative feedback in the form of prompts; and one that received negative feedback in the form of recasts, which are described later in this section. He found that the groups that received negative feedback did significantly better at recalling the target vocabulary than did the no-feedback group. He also found that the group that received prompts did better with longer-term retention than did the group that received recasts.

Nevertheless, regardless of the theoretical perspectives of the researcher, one aspect that no one argues is that negative feedback does indeed naturally occur in the language-learning classroom. In their now classic study, Lyster and Ranta (1997) found that educators provided negative feedback for 62% of the linguistic errors produced by students. Moreover, students responded to 55% of all cases of negative feedback with immediate uptake, or an attempt to right the error that they had made that was brought to their attention by the feedback. Thus, thanks to negative feedback, learners attempted to correct themselves after approximately 34% of the errors that they committed. Having

established the possible theoretical merit of negative feedback, as well as the indisputable presence of negative feedback in the SLA classroom, and having shown that negative feedback is far more heavily emphasized in the pertinent literature, we take as our first defining principle of SLA feedback that it is provided in response to a learner error. That is, prototypically at least, SLA feedback is negative feedback. This first principle of SLA feedback is listed in 1.1 above and is repeated below for convenience.

> CF 1: Classroom feedback is provided in response to an error.

This principle is discussed by Ellis (2009a, p. 3) as follows:

> In pedagogical theory positive feedback is viewed as important because it provides affective support to the learner and fosters motivation to continue learning. In SLA, however, positive feedback (as opposed to negative feedback) has received little attention, in part because discourse analytical studies of classroom interaction have shown that the teacher's positive feedback move is frequently ambiguous (e.g. "Good" or "Yes" do not always signal the learner is correct, for they may merely preface a subsequent correction or modification of the student's utterance).

Thus, not only is positive feedback not studied or emphasized in spite of its recognized importance, it is often poorly administered. Nevertheless, regardless of the question of why, the fact of the matter is that negative feedback is theoretically and practically more important within SLA than positive feedback because it is the form of feedback that is both most studied within the field of SLA and most used within language classrooms. However, while making this distinction between feedback types within the classroom, and in making all the distinctions between feedback types both in the classroom and in digital games to follow, I emphasize that the distinction between feedback types is not simple as such distinctions suggest. Like all communicative tools, feedback is a complex phenomenon, and the different types of feedback lie along a continuum rather than in completely separate categories. Nevertheless, for the purposes of this study I use these

artificial dichotomies because the purpose of the present work is to compare prototypical feedback scenarios rather than examine the reality of different feedback types.

1.1.2. Mode of classroom feedback

Having established when SLA feedback is usually administered, we now turn to look at how it is usually administered. In general, SLA methodology is divided into two different modes: oral and written communication (Lee & VanPatten, 2003). Nevertheless, although work has been done on SLA feedback in writing (e.g. Polio, 2012), the vast majority of the literature has focused on SLA feedback in oral communication (Loewen, 2012). In order to determine how SLA feedback in oral classroom communication is usually administered, we turn to the findings of the feedback study by Lyster and Ranta (1997). In this study Lyster and Ranta observed several different SLA classes and noted each time negative feedback was provided, and how it was provided. Upon analyzing their data they identified six different ways that the different educators provided SLA feedback: recasts, elicitations, clarification requests, metalinguistic feedback, explicit corrections, and repetitions. Below are hypothetical examples of these SLA feedback mechanisms.

Example 1. Recast

Teacher: What is your name?

Student: My name Peggy. [error]

Teacher: Your name is Peggy. [recast]

Example 2. Elicitation

Teacher: What is your name?

Student: My name Peggy. [error]

Teacher: My name Peggy? My name… [elicitation]

Example 3. Clarification request

Teacher: What is your name?

Student: My name Peggy. [error]

Teacher: I'm sorry? [clarification request]

Example 4. Metalinguistic feedback

Teacher: What is your name?

Student: My name Peggy. [error]

Teacher: What word do we need to add to link 'name' with 'Peggy'? [metalinguistic feedback]

Example 5. Explicit correction

Teacher: What is your name?

Student: My name Peggy. [error]

Teacher: Oh, you mean 'My name is Peggy'. [explicit correction]

Example 6. Repetition

Teacher: What is your name?

Student: My name Peggy. [error]

Teacher: My name Peggy? [repetition]

As these examples illustrate, the different types of corrective feedback identified by Lyster and Ranta vary primarily in how directly they address the learner's error. On the one hand we have repetitions and clarification requests that merely point out to the learner than an error has been made. On the other hand, we have explicit corrections and recasts that correct the error for the learner. Another thing that we notice when looking over these types of CF, is that they are all linguistic in nature. That is, in the classroom, there is predominantly only one tool for providing feedback, which is language. Obviously the examples listed above are overly simplified. As has been well established,

language is much more than mere words, and so it is with feedback. Thus, we should not discount the importance of factors such as tone of voice and body language employed by instructors in concert with the actual words they speak when providing negative oral feedback, and future research should examine the effects of such suprasegmental features of feedback both within the classroom and within digital games. Nevertheless, we do see that none of the feedback provided involves any elements beyond those that are commonly associated with normal language use. Thus, for example, the educators do not include grading as part of their feedback, nor do they make any extra-linguistic visual displays, such as jumping up and down. Instead, all of the feedback provided is within the realm of normal oral communication. Hence, we take as our second principle of SLA feedback that such feedback is linguistic, as illustrated below.

CF 2: Classroom feedback is linguistic.

To date there is no research that I am aware of that compares the relative effectiveness of exclusively linguistic feedback with richer forms of feedback when provided in a language learning environment. This gap is likely due to the fact that providing real-time extra-linguistic feedback within an L2 classroom is not normal, as indicated by the findings of Lyster and Ranta. Moreover, attempting to provide this type of feedback would be difficult at best.

1.1.3. Uptake of classroom feedback

Moving on from the when and the how of SLA classroom feedback, we now look more closely at the role that it plays practically within the classroom. As discussed earlier, in SLA theory, the purpose of feedback is to aid with the language acquisition process, and it has been shown to play an important part in that process (Cook, 2008; Loewen, 2012). Nevertheless, what we have not looked at it is what happens after

feedback is unsuccessful. This question is important because it reflects the practical integration of feedback within the SLA classroom. That is, the more integrated feedback is practically into the classroom, the more notable should be the result if a learner does not learn from it.

In order to see what happens when feedback is unsuccessful, we must be able to determine when it *is* successful. One gauge that researchers often use in order to determine whether feedback has been successful is known as 'uptake'. Uptake occurs when negative feedback is perceived by the learners, who then attempt to correct their original construction (e.g. *Ella tiene un **novio***. 'She has a **boy**friend.'). The presence or absence of immediate uptake is often identified as one of the signs of how successful any given negative feedback was at aiding the learners identify the error in their interlanguage (Bargiela, 2003; Heift, 2004). If uptake occurs, the feedback was successful; if there is no uptake, then the feedback was unsuccessful. This method of evaluating the efficacy of feedback has several weaknesses (Lyster & Ranta, 2013). First, it does not confirm that the learner actually learned anything. In the case where the feedback contained the corrected learner response, uptake could merely consist of the learner repeating the feedback, without actually having internalized it. Second, looking exclusively at uptake does not account for any learning that occurs that is not immediately vocalized by the learner. Third, uptake can only occur after negative feedback, and therefore cannot be used to measure the effectiveness of positive feedback.

Although looking for uptake is not a perfect way to tell whether feedback has been successful, it is the most commonly employed methodology (Loewen, 2012; Lyster & Ranta, 1997), and is thus the one we use herein. When feedback is found to be successful, instructors often react with what Lyster and Ranta call "reinforcement" (1997). That is, they confirm that the learner has successfully corrected their utterance.

However, educators may also omit reinforcement and simply proceed with class. On the other hand, when a learner does not respond to feedback with uptake, the educators generally follow one of three courses of action. They either repeat the feedback, provide a different form of feedback (e.g. they may provide a prompt after a recast has been unsuccessful), or they may simply ignore the omission of uptake and move on with class (Lyster & Ranta, 1997). To my knowledge, no study to date has looked at the frequency with which any of these strategies is usually employed by educators. Nevertheless, because educators can react the same way to both successful and unsuccessful feedback, by simply ignoring it and moving on with class, we see that, practically speaking, feedback is not fully integrated into a language class. That is, whether or not feedback is provided, and whether or not feedback that is provided is perceived and responded to by the learner, the class continues on. This lack of a practical importance of feedback in the L2 classroom is reflected in our final principle of classroom feedback.

CF 3: Class goes on regardless of whether classroom feedback is successful.

The fact that classroom feedback is not fully integrated on a practical level in class does not mean that it is not important. As stated earlier, feedback in the classroom is considered to be essential because it helps learners progress in their mastery of the language. Nevertheless, it would be logistically difficult, if not impossible, to integrate feedback fully into a language classroom at a practical level. Such integration would require that class not progress until all feedback had been correctly responded to with learner feedback, and would thus make the class slow to a crawl.

1.1.4. Prototypical classroom feedback

As has been discussed, feedback has been found to be an important component of the language learning process (Gass & Selinker, 2008; Loewen, 2012). Research has

shown that language learners need feedback when they attempt to communicate in a new language so that they can identify weaknesses and errors in their own command of the language and can thus be enabled to work to overcome these weaknesses and errors (Gass & Selinker, 2008; Loewen, 2012). Because of the importance of feedback to the language learning process, there is an ongoing line of investigation that seeks to determine which, if any, factors can influence the effectiveness of feedback. One factor that continues to be investigated is the nature of the feedback itself. Nevertheless, in this line of research the different types of feedback that are compared are all forms of feedback that naturally occur within the language learning classroom, as documented in observational studies such as Lyster and Ranta (1997).

Most research that compares the effectiveness of different types of feedback is probably limited to comparing the types of feedback that naturally occur within the language learning classroom because most of this research has either been conducted in the classroom environment, or in laboratory settings that are designed to mimic the classroom environment as closely as possible. However, language learning is not a process that only happens in the classroom. In fact, with the advances of modern technology, language learning is a process that is increasingly occurring within virtual spaces, or in places that only exist digitally (Thorne, Black, & Sykes, 2009). Some examples of virtual spaces include chat rooms, course management systems, and digital games.

Because these environments are digital, they have their own types of naturally occurring feedback, which do not necessarily correspond with the types of feedback that naturally occur within the language classroom. Because language learning is occurring within digital spaces, and because the feedback that occurs naturally within these environments can differ from the feedback that naturally occurs within the language

learning classroom, research is needed that compares the feedback that naturally occurs within the language learning classroom and the feedback that naturally occurs within digital spaces in order to determine which type of feedback is most effective in aiding the language acquisition process within these digital spaces. The purpose of the present study is to establish a foundation for this line of research by comparing the effectiveness of these two types of feedback within a digital game that is designed to teach Spanish food vocabulary to novice language learners.

In order to achieve this goal I have identified three principles that can be used to identify the prototypical feedback that is provided in language classes. These principles are: (1) classroom feedback is provided in response to an error, (2) classroom feedback is linguistic, and (3) class goes on regardless of whether classroom feedback is successful. These are the principles that were used to develop the feedback for the classroom feedback (CF) version of the game used in the present study. We now follow a similar method of investigation in order to determine the corresponding features behind the development of digital game feedback (DGF) version of the game.

1.2. Feedback in digital games

As is the case with CF, DGF is generally broken into two super-categories. In the field of digital game design, these are known as 'rewards' and 'punishments' (Schell, 2008). The classifications of rewards and punishments correspond to the CF categories of positive and negative feedback in turn. Thus, we can see that both environs provide feedback to identify both desired and undesired behavior on the part of the participants. However, although both environments use the same superordinate categories of feedback, there are several notable differences between the features used in developing the

feedback provided in these two settings. I now discuss the nature of the principles behind DGF. Then I compare the principles of CF with those of DGF.

Researchers have already identified several common principles behind how feedback should be designed and used within digital games (McGonigal, 2011; Purushotma, Thorne, & Wheatley, 2008; Schell, 2008). What follows is a synthesized list of some of the more common principles. These include:

1. Punishments and rewards are treated as equally valuable.
2. All feedback is playful.
3. Metadata are provided at the player's behest, and are not part of the core feedback.
4. Feedback shows that the player's choices have made a real difference.
5. Just-in-time information is provided as needed by the player.
6. Feedback is provided continuously.
7. Feedback should be 'juicy' (rewarding to players in multiple ways at once).
8. Feedback should have endogenous value (importance for the game itself).
9. Most feedback should be in the form of as many rewards as possible.
10. As little feedback in the form of punishments as possible is provided.

Each of these is addressed below. The analysis of this list of ten principles allows us to generate a list of three principles behind the creation of digital game feedback that correspond with the three principles of classroom feedback identified in the previous section. These three principles of digital game feedback are listed here.

DGF 1: Digital Game Feedback is provided for each and every player action.

DGF 2: Digital Game Feedback is 'juicy'.

DGF 3: Digital Game Feedback has endogenous value.

The first principle for both CF and DGF relates to the frequency with which the feedback is provided. In CF, feedback is provided each time a player makes an error, while in DGF, feedback is provided each time the player performs any action. Each of the second principles is related to the mode in which the feedback is provided. In CF, feedback is linguistic while in DGF feedback is 'juicy', making use of as many modes as possible in the provision of feedback. Finally, the final principle for each environment is connected to the practical integration of the feedback into that environment. In the classroom, the feedback is not practically integrated, and therefore the class moves on regardless of whether the feedback has resulted in uptake on the part of the learner. On the other hand, in DGF, the feedback is fully integrated into the game environment in such a way that the feedback directly affects the player's success in the game, and therefore has endogenous value. In the following subsections I discuss the DGF principles in greater detail and focus on how they affect the production of feedback within digital games.

1.2.1. Positive vs. negative DGF

One point that is instantly clear when looking at items 1, 6, 9, and 10 is that feedback in digital games is not restricted to being provided only when the player makes a mistake. On the contrary, feedback in digital games is provided incessantly, and is especially emphasized when the player has performed well. The guideline for putting emphasis on positive feedback is that feedback should be provided as much as possible when the player does well. Even so, negative feedback is just as important as positive feedback in terms of quality. That is, both positive and negative feedback should be well designed and thought out, even if there is more positive feedback. These observations lead us to our first principle for DGF.

DGF 1: Digital Game Feedback is provided for each and every player action.

The consistency with which feedback is applied within digital games is one of the primary arguments in favor of the use of digital games as language teaching tools (Gee, 2003, 2005, 2007, 2012; Jackson, Dempsey, & McNamara, 2012; Johnson, Adams, & Cummins, 2012; Prensky, 2001; Sykes & Reinhardt, 2012). When feedback is provided consistently after every player action, then the players always know exactly how they are doing. If we apply this to language learning, then dependable feedback allows SLLs to know continuously how well they are using the language, which in turn permits SLLs to detect and overcome weaknesses in their interlanguage. By extension, dependable feedback helps learners to know clearly what they are doing correctly with their use of the language.

1.2.2. Mode of DGF

For our next principle we look specifically at point 7 stated above and repeated here for convenience: "DGF should be 'juicy'." The term 'juicy' in this context is used to contrast with the more generally used term of something being 'dry' (Schell, 2008). When we speak of something as being 'dry', we usually mean that it is uninteresting, bland, flat, or boring. Thus the term 'juicy' can be taken in this context to refer to something that is interesting, flavorful, multifaceted, and exciting. The key to making 'juicy' feedback is to use as many diverse tools as possible for all feedback (Schell, 2008). The task of looking at the possible feedback tools that are used in digital games is somewhat challenging because there are many digital game genres, all with differing types of play and feedback (Schell, 2008; Sykes & Reinhardt, 2012). To date, no one has conducted an empirical study to compare the efficacy of the different feedback mechanisms provided in digital games, as doing so would, according to Schell, be "sort of like Van Gogh looking for … evidence that sunflowers are the best flowers to paint"

(personal communication). Nonetheless, Sykes and Reinhardt (2012) do provide a list of some of the most commonly used feedback mechanisms in digital games:

1. Leveling: Indicating how far a player has advanced in the game.
2. Points: Keeping track of how well the player is doing.
3. Asset building: Acquiring resources needed for certain game mechanics.
4. Skill building: Increasing the number of actions a player can perform.
5. Tips and hints: Providing key information on an as-needed basis.
6. Real-time progress bar: Tracking a player's short-term progress in real time.
7. Sound effects: Providing auditory cues used for a vast variety of purposes.
8. Active and inactive game elements: Restricting the game elements that players can utilize until they have completed certain prerequisites.

As the preceding list illustrates, these tools are quite varied in several respects. One difference is the medium in which they are used. The media include sound, visual stimuli, numbers, and language. For example, in the game "Final Fantasy", written text is used to provide tips and hints, while in the game "New Super Mario Brothers Wii", ancillary videos are used to provide the tips and hints. Another way these tools vary is the type of player behavior to which they respond. While some provide feedback on short-term player performance, others are used to provide feedback for longer-term behavior. For example, in the game "Just Dance 3", language, numbers and color are used to provide short-term feedback on every move performed by the player. On the other hand, sound and special effects are used to provide feedback on overall, longer-term progress during a song. This discussion leads us to the second principle for the design of DGF.

DGF 2: Digital Game Feedback is 'juicy'.

Digital game designers argue in favor of juicy feedback on the grounds that it helps both immerse the player within the game and retain them there (Schell, 2008). This

engagement occurs because when everything you do has several very noticeable effects, you are provided with a feeling of power and control (Schell, 2008). Another potential argument in favor of juiciness, specifically for games that are designed to be educational, known as 'serious games', may be the phenomena of differing learning styles (Coffield, Moseley, Hall, & Ecclestone, 2004; Reid, 1995). If feedback is provided for as many different learning styles as possible, then the serious game has the prospective to be more beneficial for a wider variety of SLLs.

At this point I would like to state that the term 'serious games' is rather unfortunate because it suggests that these games are distinct from what might be called 'normal games'. As discussed in chapter 2, the arguments in favor of 'serious games' rest on the fact that they are exactly like 'normal games'; thus forcing a distinction between the two in effect calls into question their utility as pedagogical tools. Nevertheless, for the sake of simplicity I adhere to the tradition of calling games that are designed with pedagogical aims 'serious games', while keeping in mind that this term is not ideal.

1.2.3. Uptake of DGF

For the final principle we look at the integration of feedback within the game. To address this point, we re-examine the 8th point in the principles of DGF design listed above: 'Feedback should have endogenous value (importance for the game itself)'. This means that the feedback provided should be critical for player success. In order to address the endogenous value of feedback, I first distinguish between what I call 'feedback elements' and 'feedback events'. I use the term 'feedback elements' to refer to the various possible varieties of feedback, such as points, leveling, skill building, etc., listed in the preceding discussion. I use the term 'feedback events' to refer to each time feedback is provided. Thus, in accordance with the first principle of DGF, each player

action is followed by a feedback event. And in accordance with the second principle of DGF, each feedback event is composed of several different feedback elements. For example, in the game "Minecraft", when one mines a block of stone, one receives a feedback event. This event consists of several different feedback elements: a sound effect (a small popping sound), a visual effect (the block shrinks and moves toward the player's avatar), and an increase in resources (the number of stone blocks increases by one).

The importance of a feedback event is determined by the combined importance of each distinct feedback element. Thus, in our example, even though the sound effect and the visual effect do not affect player success, which means that they have no endogenous value, the increased number of resources will affect player success, which would mean that it does have endogenous value, and thus the feedback event likewise has endogenous value. Therefore I propose that the importance of the integration of feedback within a game is at the feedback event level. That is to say, even if each feedback element in a given feedback event does not have endogenous value, at least one feedback element in every feedback event should have endogenous value. This leads to our final principle of DGF.

DGF 3: Digital Game Feedback has endogenous value.

The concept of endogenous value may seem counterintuitive at first, in the sense that it is assumed that digital games are played for entertainment. Why would the players care whether the feedback provided has an important role to play in the game as long as it is entertaining? To answer these questions we first look at what games are, and why they are entertaining. Although there is no single vetted definition of what a game is, almost every definition of 'game' includes a reference to the fact that a game involves a struggle, challenge, obstacle, or conflict (Costikyan, 2002; Schell, 2008). This focus on achievement of a specific goal also relates to Weigand's (2010) Mixed Game Model,

which is based on the premise that dialogue itself is a game in which the interlocutors attempt to achieve the goal of communicating their message. Thus, the fact that games present players with something to overcome is one of the main reasons why they are successful at being entertaining. Therefore, anything within the game that does not help the player overcome the challenge is bereft of endogenous meaning, and consequently loses its entertainment value.

1.2.4. Prototypical DGF

To recapitulate, we have identified three principles for DGF that correlate with the three principles identified earlier for CF that are relevant to the present study. These principles are illustrated in Table 1.1 below.

Table 1.1. Principles of CF and DGF

	CF	DGF
Frequency of Feedback	Provided in response to an error.	Provided for each and every player action.
Mode of Feedback	Linguistic.	'Juicy'.
Practical Integration of Feedback	Not integral to what happens in class.	Has endogenous value.

The objective of this dissertation is to compare the relative effectiveness of these feedback principles within the environment of a serious (educational) Second Language Vocabulary Acquisition (SLVA) game. In order to do this, I built two versions of a video game that differ only in the principles found in the feedback provided by the game. In

order to do so, I first needed to analyze these differences, which is presented in the following section.

1.3. Comparing CF and DGF

Before we begin with an analysis of the differences between the principles behind prototypical classroom feedback and prototypical digital game feedback, a brief discussion of the reason behind making this dichotomy is warranted. Obviously, classroom feedback and digital game feedback do not exist as a strict dichotomy. Rather, the types of feedback that occur within each environment are variable and exist on a continuum of feedback. Thus, for example, within some digital games, feedback may only be provided after errors, and within some classrooms, feedback may be 'juicy'. Nevertheless, such cases are deviations from the norm, as discussed above. As the purpose of the present research is to establish a baseline comparison between classroom feedback and digital game feedback that most naturally occur within their respective environments, it was necessary to disregard the variation that equally occurs naturally within the feedback in these two environments. Instead, I compare the prototypical feedback that occurs within these environments, as determined by the principles analyzed in the previous sections.

In order to compare the first principle "Frequency of Feedback" for each environment, we need to compare the goals of feedback within each environment. As discussed in detail above, within current SLA theories, the primary purpose of feedback is to help the learners identify anomalies in their interlanguage so that they can work to overcome them. On the other hand, there are several other purposes for feedback within digital games. One purpose is comparable to that of the purpose of feedback in the language classroom, and consists of pointing out to the players where they need to

improve in order to have more success. However, other purposes consist of entertaining the player, encouraging experimentation, and rewarding the player for doing the right thing (Schell, 2008). The variance in the goals of the feedback within these environments likely contributes to the difference in the frequency with which feedback is delivered. That is, because the goal of feedback in the classroom is to point out errors, the principle implies that feedback should be used only after errors. However, because the goal of feedback within digital games is multifaceted and relates to all player actions, the principle in this context dictates that feedback should be provided all of the time for all player actions.

It is unclear which approach would be better within a serious SLA game. On the one hand, providing feedback only after errors, in accordance with the frequency principle of CF design, could help the learner-players to focus on the material they are studying, and especially on the material that they need to work on. On the other hand, providing feedback after each player action, in accordance with the frequency principle of DGF design, could help the learner-players to be more immersed in the game. This deeper level of absorption could help with language learning in several ways. First, it could increase the importance of winning the game for the player. Second, it could increase the amount the player will play the game. And finally, it could help the players feel more at ease, thus lowering their affective filter.

The second pair of principles "Mode of Feedback" is related to the nature of the feedback provided in each environment. The difference here is in part due to the nature of these environments. In the classroom, the instructors, who are human, can do only one thing at a time, which limits the amount of feedback they can provide. In addition, the only activity that can be performed in real time is oral interaction, which means that the most natural and easily provided form of feedback in the classroom is in the form of

linguistic feedback. It should be pointed out that this oral feedback is likely to be frequently enriched with gestures, tone of voice, etc., but to my knowledge there is no study that has looked at the importance of these variables for the effectiveness of classroom feedback; therefore I do not consider them in this study. On the other hand, a digital game can do many different tasks of many different types at the same time. This means that it can easily provide feedback in several different ways and forms that are not necessarily limited to the type of action performed by the player. For example, if the player is responding to textual prompts within the game, the game can provide feedback in the form of sound effects, visual effects, score and resource changes, etc., in addition to providing feedback in the form of oral utterances or written text. The difference in the nature of the environment, therefore, is what causes the difference in the second set of principles. In the classroom, feedback needs to be administered quickly and easily by a human within a spoken conversation, and therefore it is only spoken; although, as mentioned in 1.1.2, it can also be accompanied by pragmatic and suprasegmental cues (e.g. changes in tone and body language). On the other hand, because a computer can easily perform many actions at the same time and one of the goals of digital games is to entertain the player, feedback in digital games is designed to be 'juicy'.

In this case again it is unclear which approach will work better within serious SLA games. It could be again that the classroom approach will help learner-players to focus on the pedagogical tasks within the game, which could lead to improved learning. But it could also be that the digital game approach will help them to become more immersed within the game, which may lend communicative strength to what is presented in the game, and thus lead to greater acquisition.

The final set of principles, "Practical Integration of Feedback", relates to the integration of the feedback into the environment in which it is delivered. Like the second

principle, this one is also related to the nature of the two environments. In particular, the principle of feedback integration indexes both how the feedback contributes to success within each environment from the perspective of the participant, and how the feedback affects what happens in the environment.

In the classroom, students consider their grades to be the mark of success. This means that everything that does not directly affect their grade is of a secondary importance. From the student's perspective, things that happen in class that do not directly affect their grades, at best, help prepare them for other activities that will have a direct impact on their grades. An example of this point of view is the well-established grades-ratings correlation in student evaluations (Greenwald & Gillmore, 1997). Thus, the significance of feedback for the students is that it will help them learn the language so that they can be better prepared to take tests, quizzes, etc. It is not possible to have feedback directly affect grades because feedback within the classroom is generally administered to a single student at a time. Hence, if feedback increased the score of a single student who had either volunteered before anyone else had a chance or had been called on, the other students would feel that an injustice had been served because they had not had the opportunity to receive the feedback and the accompanying improvement in their grade. On the other hand, if the feedback provided to a single student was accompanied by a lowering of their grade, then that student would feel that they had been treated unjustly as no one else had to run the risk of receiving the feedback and damage to their grade. These concerns could be ignored if it were always possible to provide systematic feedback for all students all the time and if it were always possible for everyone who wanted to volunteer to do so, but this is not possible for a single educator teaching multiple students.

Another result of full integration of feedback into the classroom would be that class would not move forward until all feedback had been correctly responded to with appropriate uptake. However, it would be logistically prohibitive to stop class each time a learner does not respond to feedback with immediate, appropriate uptake.

Nonetheless, both results of practical integration of feedback are exactly what we find within digital games. In digital games the feedback provided is determined by the computer in the same way for all players, all the time. Because the feedback provided in digital games is 'fair', it is acceptable to the players for the feedback to have a direct impact on their success within the game. Both negative and positive feedback can hinder and facilitate success, though normally it is negative feedback that hinders (e.g. in the form of resource depletion), and in extreme cases even prevents (e.g. in the form of ending the game) success, while positive feedback most commonly facilitates success (e.g. an increase of resources). For example, in World of Warcraft one type of success that can be achieved is obtaining the "Exalted" status with different organizations within the game. One form of feedback that can be obtained from killing certain 'creatures' within the game is either an increase or decrease of one's reputation with a specific group. If players kill a member or an ally of a group that they are trying to befriend, their reputation score will drop, making it more difficult to achieve "Exalted" status. On the other hand, if they kill an enemy of the group, their reputation will increase, drawing them closer to high status. This direct effect of the feedback on the player's reputation is accepted by the players because they know that the feedback will be provided in the same way for all players each time they destroy members of either group. Moreover, players enjoy the fact that all actions receive feedback that has an important, direct impact on success within a game, because it means that everything they do is relevant and important. They are never wasting time.

Yet again, it is not clear which approach will perform better within a serious SLA game. On the one hand, the classroom approach may help learner-players to feel more at ease, since mistakes will not hinder their ability to win the game. On the other hand, the digital game approach may create choices that the player makes that are more meaningful.

To illustrate these differences in the principles of feedback within these two settings, we look at a hypothetical example of the task of picking a named food from an assortment on a table. In a classroom we could expect something like example 7 where a learner gives an incorrect response, and something like example 8 where the student produces the correct response.

Example 7. Typical negative classroom feedback

Teacher: Where is the tomato?

Student (Pointing at the cherry): Here. [error]

Teacher (Pointing at the tomato): The tomato is here. [negative feedback: recast]

Student: Okay.

Teacher: So where is the tomato?

Student (Pointing at the tomato): Here. [uptake]

Teacher: Very good! (moving on to the next student) [positive feedback: transition]

Example 8. Typical positive classroom feedback

Teacher: Where is the tomato?

Student (Pointing at the tomato): Here. [correct response]

Teacher: Very good! (moving on to the next student.) [positive feedback: transition]

In examples 7 and 8, the student is provided with both negative and positive feedback. Nevertheless, in accordance with the frequency principle of CF, the negative feedback is emphasized more, with the positive feedback serving as a transition to indicate to the learner that the turn is over and that it is time to move on to the next student or activity. In accordance with the second principle of mode, the feedback provided is mostly spoken. The visual cue of pointing is provided only because without it, the spoken feedback would be meaningless. Finally, in accordance with the third principle of practical integration, neither the positive nor the negative feedback has any direct effect on the student's grade.

In a digital game we expect feedback like example 9 after an incorrect reaction, and feedback like example 10 after a correct response. Importantly, in the digital game imagined for examples 9 and 10, dollars are used to purchase additional food, which is a necessary resource for winning the game, and the student's current dollar total is continuously displayed in the top, right-hand corner of the screen.

Example 9. Typical negative DGF

> Cook: Put the tomato in the pot.
>
> Student (Clicks on the cherry). [error]
>
> Cook (The image of the cherry vanishes in a puff of smoke. There is the sound of a faint explosion. The student loses ten dollars and the dollar display flashes red. Their turn is over and two buttons appear. One button allows the student to continue on to the computer's turn, while the other will identify the tomato for them.): No, not the cherry! [negative feedback]

Example 10. Typical positive DGF

> Cook: Put the tomato in the pot.

Student (Clicks on the tomato). [correct response]

Cook (The image of the tomato flashes cheerfully. There is a bright ringing sound. The student receives ten dollars and their score is highlighted with a burst of sparkles. A single button appears. They are allowed to continue cooking by pressing the button.): Ah, what a nice tomato! [positive feedback]

In examples 9 and 10, we see that the player is also provided with both negative and positive feedback. However, in accordance with the first principle of frequency, feedback, and not merely a transition, is provided after both correct and incorrect choices. Moreover, when they make a choice, juicy feedback is provided in the form of oral cues, sound effects, visual effects, and a change in resources in accordance with the principle of mode. The monetary feedback has endogenous value for the players because they need dollars in order to succeed at the game in accordance with the principle of practical integration. These examples of classroom and digital game feedback in examples 7 to 10 are fairly demonstrative of what is typically found both in real-life classrooms and in digital games, and illustrate clearly the application of the six principles discussed.

As the preceding discussion illustrates, both of these sets of principles have been developed according to the needs and limitations of their respective environments. However, recently these environments have become blended within digital games designed with pedagogical ends, also known as ‘serious games’. Moreover, the goals of both environments have become blended within these serious games. We now have a new situation in which the goals include both instruction and entertainment. Because the settings and their goals have become blended, it is unclear which, if either, of these sets of features would do a better job in achieving the learning objectives of serious games. Hubbard (2002) noticed this gap over a decade ago and called for studies that

investigated how best to design digital language learning games. Since Hubbard (2002), little work has been done to answer his call (Cornillie et al., 2012a). The small body of work that has been done with regards to this gap and feedback, as well as the other calls that have been made for research that compares the effectiveness of different feedback types, are the topic of the subsequent section.

1.4. Calls for work on feedback within serious games

The need for research on how feedback within Computer-Assisted Language Learning (CALL), and within digital games in particular, should be designed has not gone unnoticed. One early call for research that investigates how best to design feedback for CALL activities related to vocabulary instruction originates in Van Bussel (1994), who discusses the fact that one of the issues hindering the progress of CALL research is the lack of foundational work establishing guidelines for the design of effective CALL tools. Van Bussel attempts to help remedy this problem by proposing four steps that CALL designers should follow with each CALL activity. These steps are:

1. The division of the learning task in units.
2. Posing the right questions and providing feedback and hints.
3. The stimulation of elaboration.
4. The arrangement of training to "automate" acquired skills.

In his discussion of the second step, Van Bussel concludes that all CALL feedback should provide the learners with the correct response to each error that they commit, and that feedback should be tailored as much as possible to each individual learner's learning style. However, he acknowledges that he has based these guidelines on insufficient evidence and therefore calls for further research to investigate how best to design feedback within CALL activities aimed at vocabulary acquisition.

A few years later, Nelson (1998) looked specifically at web-based CALL activities targeted for students in Japan learning English vocabulary. Nelson begins by arguing that web-based CALL activities have much promise for vocabulary instruction chiefly because they can be individualized, they allow for independent learning, and they provide consistent, applicable feedback. He then proceeds to discuss the different web-based vocabulary instruction activities that he had used at the University of Aizu. One aspect that all of these activities have in common is that they deliver consistent, albeit very simplistic, feedback. Unfortunately the only results that he provided were informal comments made by some of the students after having used the materials. One of Nelson's conclusions is that more research is needed in order determine how the feedback could be improved through elaboration.

Heift (2001) takes a step in this direction with a web-based CALL program designed to help learners of German acquire grammar principles. In the program that she used in her study, Heift set up the feedback so that it would be more elaborate in two specific ways: it provided metalinguistic feedback, and it was customized automatically by the program to work with each individual student's personal language learning strengths. Heift's main interest in this study was whether or not the learners would actually pay attention to such elaborate, personalized feedback. Her results showed that learners paid attention to nearly 80% of the feedback provided by the program. However, she recognized that she had not distinguished very much between different learner levels. She thus called for further research to look at the relative effectiveness of different feedback types with different levels of learners.

In a recent edition of the Routledge *Handbook of Second Language Acquisition*, Heift and Chapelle (2012) addressed the value of technology in aiding language instruction in general. In this chapter they conduct a brief review of CALL literature to

that date, and then discuss the implications for the future. They focus a large portion of their discussion on the feedback provided by CALL technologies. One of the things that they point out is that, thanks to technology, we now have a completely new SLA environment. That is, in addition to the traditional environments of students interacting with teachers and with each other, we now have students interacting with technology. Heift and Chappelle point out that this new situation requires its own SLA research, because research conducted in the more traditional environments may not be applicable in this new setting. By extension we see that with respect to feedback they are calling for research that examines the use and merit of traditional classroom feedback techniques and tools within CALL activities.

Recently, one study has examined the role of feedback design in digital language learning games (Cornillie et al., 2012a). The game used in this study was designed to aid in the acquisition of pragmatics. The authors identified three differences between classroom and game feedback: CF is generally very explicit, while games rarely give anything away; learning in games is based on experimentation while learning in the classroom is based on rules; and feedback in games is dependent on the content while feedback in the classroom is not. In their study, 83 participants played their game outside of the classroom setting, even though they were recruited through the classroom. The game used in this study was a 3D role-playing game in which the player needed to use English pragmatics correctly in order to complete quests. Multiple versions of the game were prepared that varied with respect to the explicitness of the corrective feedback provided when the learners made pragmatic mistakes. The only data collected during the gaming phase of the study was how often each participant was exposed to corrective feedback. The analysis was based instead on comments from the participants provided in interviews after they had finished playing the game. Cornillie et al. found that, in general,

the participants rated the more explicit, classroom-style, feedback as more useful and enjoyable than the more implicit, digital game-style, feedback. The researchers concluded that it may be the case 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". Cornillie et al. conclude their paper with a call for future research related to the use of feedback in digital language learning games:

> 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). (2012a, p. 274)

In this study I consider "corrective feedback intended to increase a learner's understanding" as CF, and "more 'game-like' feedback elements that can contribute to intrinsic motivation" as DGF. The purpose of the present study is to address the call made to distinguish between these two varieties of feedback within the context of a serious game designed to aid in the acquisition of Spanish food vocabulary.

1.5. Outline of dissertation

In this chapter I have reviewed the nature of feedback in the SLA classroom and in digital games. In the next chapter I discuss the literature relevant to the other aspects of the current study. In chapter 3, I present the methodology used in this study and provide a detailed description of the game and questionnaire used. The fourth chapter presents the findings obtained from carrying out the present study. Finally, the fifth chapter discusses the implications of the findings, and suggests future avenues for related research.

Chapter 2: Literature Review

The previous chapter presented the topic of feedback, and discussed its importance both within the language learning environment and within commercial digital games. The current chapter establishes the remaining pieces of the foundation needed for the present study. These include the theory behind, and methodological approach to, the SLA framework used herein with a specific emphasis on the methodologies related to the instruction of vocabulary in a second language, the research related to the commercial digital games that was used to guide the development of the digital game used in this study, and a discussion of the history to date of educational digital games and its effect on the current perception and acceptance of educational digital games by educators and learners.

2.1. Second language acquisition (SLA)

A retelling of the history of modern SLA theory and related methodologies is beyond the scope of the present work. Those interested in this subject should look to one of the many different works that go into detail on this subject (Cook, 2008; Hilgendorf, 2012; Larsen-Freeman, 2011; Richards & Rodgers, 2002). Instead, I simply indicate that at the time of this writing, the most commonly accepted theory related to SLA is known as the Communicative Approach, or Communicative Language Learning (CLL; Larsen-Freeman, 2011). Within the theory of the CLL, language instruction should, as the name implies, involve the use of language for the purpose of communicating, and feedback provided when there is a breakdown in communication can be argued to be a critical component of CLL. This theory of SLA originally emerged in the 1970s (Larsen-Freeman, 2011), and has since evolved into two recognized forms. The weak form can be described as an approach that aims at "learning to use" the language, while the strong

form can be described as "'using [the language] to learn it" (Howatt & Widdowson, 1984).

Several different methodological approaches have been proposed based on CLL. One of the most commonly accepted and used is that known as Task-Based Language Teaching (TBLT). TBLT is based on the strong form of CLL and was first used in the late 80's with the oft cited Bangalore Project (Prabhu, 1987). In TBLT the goal is to provide the learner with meaningful tasks that they can accomplish only through communicating in the target language. This approach was developed when researchers began to discover the tendency of successful learners to acquire a language more efficiently when they are using the language to accomplish something meaningful to them (Van den Branden, 2012). Thus, in TBLT, the tasks performed during the learning process are supposedly perceived as valid, authentic, meaningful tasks by the learners themselves. One of the strengths of TBLT is that, because it specifies that the tasks should be meaningful to the learners, it may very well contribute to increased learner motivation (Espinar & Baxter, 2012).

One of the major weaknesses of TBLT, however, is the absence of any single example of a good TBLT task (Van den Branden, Bygate, & Norris, 2009). This ambiguity continues to lead many researchers and practitioners to develop divergent task designs, all under the umbrella of TBLT tasks. In order to clarify what is meant by a 'good' TBLT task, several researchers have proposed criteria to be used to identify such a task (Ellis, 2003). Nevertheless, the specific list of criteria also varies somewhat from person to person. One of the most cited such lists was produced by Ellis (2003, p. 9), who identified six "critical features" of TBLT tasks:

1. A task is a workplan.

2. A task involves primary focus on meaning.

3. A task involves real-world processes of language use.

4. A task can involve any of the four language skills.

5. A task engages cognitive processes.

6. A task has a clearly defined communicative outcome.

Thanks to the vagueness of the items in this list, even these constraints are not very restrictive. Therefore, if two different tasks meet these requirements, they may still be dissimilar in many ways. Some researchers have proposed different ways of categorizing TBLT tasks in order to facilitate comparison between different tasks in different studies. Again Ellis (2009b, p. 491) is one of the most well-known authorities in this matter. He proposes these criteria for distinguishing between tasks:

1. The learners: Second vs. foreign

2. The setting: Classroom vs. laboratory

3. The tasks: Interactive vs. monologic; simple vs. complex

4. Planning: Length; guided vs. unguided; form vs. meaning focused

Again the differences made between different types of tasks are coarse and allow for a large range of variation. Much research is needed in order to determine whether meeting the aforementioned requirements generates a valid pedagogical tool. In the present study, the game that was created is based on the use of tasks for language learning, and this study will thus help to answer this question. Moreover, considerable research is also needed to compare the different types of TLBT tasks in order to determine if they are all equally effective. In this current study, a step is taken even further to compare two TBLT tasks that fall into the same category of L2 learners in a classroom setting performing a simple monologic task with guided, meaning-focused planning. The results will allow us to see whether a difference in the pedagogical effectiveness, specifically of short-term vocabulary acquisition, within a given category exists.

2.2. Second language vocabulary acquisition

In order to make a valid comparison between different tasks, we first need to make sure that they are designed in accordance with currently accepted best practices in language instruction methodology. Languages are large, complex entities, which makes it difficult, if not impossible, to teach a language as a whole. The traditional way in which this difficulty is overcome within SLL pedagogy is by breaking a language down into several primary constituents, and then teaching each of these as independent elements. The categories most often used include pronunciation, vocabulary, grammar, culture, strategies, and pragmatics (Nation, 2001). This subdivision of the language learning process is mirrored in the research literature. Granted, there are some empirical studies in the field that do look at overall learner mastery of a language. Most studies, however, center on a single specific area within the language learning process. The present study maintains this tradition by examining Second Language Vocabulary Acquisition (SLVA) as an isolated key component of SLL (Nation, 2001; Sundqvist & Kerstin Sylvén, 2012; Sylvén & Sundqvist, 2012). The study of vocabulary was chosen for the present work because the types of tasks recommended for vocabulary instruction are the most easily duplicated within the medium of a digital game.

Almost every study to date on SLVA cites the seminal work by Nation (2001), which we also take as our starting point. Nation argues in favor of approaching SLVA in two different ways. When the relative frequency of the vocabulary item is high, he argues in favor of the use of 'rich instruction'. Nation defines rich instruction as taking time to focus on each lemma, its uses and nuances. On the other hand, with low-frequency lemmas he advocates a policy of shallow, frequent processing. With this approach, learners should be exposed to as many low-frequency words as often as possible. Folse (2006) found evidence to support this latter proposal in comparing the relative effect of

three different written task types on vocabulary retention. The different tasks were one fill-in-the-blank sheet, three fill-in-the-blank sheets, and the production of original sentences. Each of the 154 ESL students who participated in his study performed each of these tasks in a random order with a randomly-selected subset of the target vocabulary. Folse found significantly greater gains in vocabulary retention when participants were doing the three worksheets, as opposed to when they performed either of the other tasks. No statistically significant difference between the results for these other two groups was found. Thus his findings support the proposal of emphasizing time on task with low-frequency lemmas because, regardless of how deep the processing was, time on task tended to be the best predictor of vocabulary acquisition.

Nation also hypothesizes that digital games and other computer software in particular may be useful with low-frequency vocabulary instruction because they can provide this type of repetitious, shallow practice. Neville, Shelton, and McInnis (2009) found evidence to support this hypothesis. In their study, they compared the vocabulary retention of a group of students who completed normal homework assignments with that of a group of students who played a digital role-playing game built around the same material. They found that the game group retained significantly more vocabulary than their counterparts. They also found that this was true in spite of the fact that the participants themselves perceived the digital game as less useful than the worksheet. Further research is needed in order to confirm the validity of digital games as a medium for SLVA.

2.3. Introduction to (educational) digital games

Given that researchers such as Nation have argued in favor of the use of digital games as language instruction tools, one may ask why digital games are not already

commonly used in this manner. We will address this question in this section by briefly reviewing the history of digital games, and of educational digital games in particular as discussed by Novak (2012). Digital games were first made available to the general public in the 1970s in arcades In spite of the inconvenience in their location and cost, however, in less than a decade, digital games became a popular form of entertainment. In the beginning, most digital games were designed based on the preferences of each individual game designer. As the industry matured, though, more thought was put into improving the process of game development. A good example is the popular game *Mrs. Pac-Man*, which had the specific goal of appealing to a wider audience: boys *and* girls. This new industry continued to flourish into the beginning of the 80s, when it hit a major slump, from which arcades never recovered. However, the introduction of the original 8-bit Nintendo Entertainment System (NES) in 1983 led to a new era of success for console games, and eventually for personal computer games as well. Much of this success can be traced back to the stiff competition between game developers. The competitiveness within the market caused many advances to occur quickly in the design of digital games. These include rapidly improving graphics and sound systems, more compelling storylines, more developed characters, and more opportunities for social interaction within the games.

Soon after the renewed interest in digital games took hold, game developers began to team up with educators to develop what came to be known as "edutainment", the history of which is discussed by Shuler (2012) and summarized here. Edutainment was a general term employed to refer to any digital game based on a pedagogical purpose. Initially this new genre of digital game was both popular and successful. Edutainment boasted such successful blockbusters as *The Oregon Trail* and *Where in the World is Carmen Sandiego?* These early edutainment games proved the possibility of

designing digital games that were successful both at being entertaining and educational. However, within a few years a quick series of large mergers resulted in the existence of only two edutainment producers. Most digital games were also now being sold at large retail stores that had limited shelf space, reducing the space available for less competitive games. The mergers and limited shelf space led to a price war between the two large edutainment conglomerates.

As Shuler (2012) continues to explain, this price war had two devastating effects on edutainment. First, less money was available to be spent on the development of new games, which led to the stagnation and the abandonment of innovation in the edutainment industry. In contrast, many other game genres were still investing huge amounts of resources into the development of newer, better, often more expensive, games. The second effect was caused by the rapid decrease in prices of edutainment, which led many consumers to assume that edutainment games were of much lower quality than digital games from other genres. The decrease in innovation and the lack of confidence on the part of the consumer led to a quick decline in edutainment, which has never fully recovered from this crash in its market. A testament to the fate of the edutainment industry is the fact that the term 'edutainment' is now employed largely as a pejorative label among game designers and gamers alike. Another witness of this fact is the lack of any successful modern commercial edutainment games. One final bit of evidence of this decline in appreciation is the replacement of the term "edutainment" with "serious games".

2.4. Digital game design principles

Because non-educational digital games have been so successful and serious games have not been very successful, I decided to design a game for this study based on the

principles behind the development of commercial entertainment games, in combination with the principles behind successful vocabulary instruction discussed above. In doing so, I attempted to integrate these two different sets of principles with each other as seamlessly as possible. In this section I review the design principles behind digital games that I employed.

2.4.1. The players

One of the primary factors professional game designers consider is their target audience (Schell, 2008). Not everyone likes to play digital games, and not everyone who likes to play digital games wants to play for the same reason. Thus it behooves game designers to design their games in such a way that they will be gratifying for as many different players as possible. Although different taxonomies have been proposed, one of the oldest and most commonly used is that created by Bartle (1996), which divides gamers into four categories:

1. Socializers: those who enjoy the social interaction;
2. Achievers: those who enjoy dominating the game itself;
3. Explorers: those who enjoy learning everything there is to know about a game;
4. Killers: those enjoy exercising power and control over other players.

Novak (2012) proposes that "If the game satisfies a player's particular motivation, it is more likely to be fun, engrossing, and worthwhile to that player".

2.4.2. The genre

In any case it is not possible to appeal to all player types and satisfy all motivations within a single game (Novak, 2012). This may be one reason why so many different game genres have developed over the years. Novak (2012) identifies nine basic genres, including Role-Playing games, Puzzle games, and Simulation games, among

others. Recently these three genres in particular have been the most commonly used in the development of serious games. In this study I diverged from this recent trend and instead designed my game as a Strategy game. This was a conscious choice based on the fact that early, commercially successful, edutainment games tended to be either strategy games (e.g. *Where in the World is Carmen Sandiego?*) or simulation games with strong strategy components (e.g. *The Oregon Trail*).

2.4.3. The experience

Within any given genre, there is a wide range of very different games, largely due to the fact that each successful game is designed to provide its players with a unique experience (Schell, 2008). The exact nature of the desired experience is dependent on the target audience for the game. Thus there are strategy games that strive to give players divergent experiences, ranging from commanding large, futuristic armies against hordes of aliens in bloody battles (e.g. *StarCraft*), to directing 'cute', one-inch tall aliens in rebuilding their crashed spaceship (e.g. *Pikmin*). Because I wanted to run a study in a classroom environment using low-frequency food vocabulary, I decided to try to give the participants the experience of competing in establishing a food kiosk franchise by matching their cooking skills against those of their opponents.

2.4.3. The mechanics and feedback

The decisions made regarding the players, the genre, and the experience are used as guides in making all other decisions related to the development of the game. That is, each game design decision from this point should be made to help the players have the desired experience within the genre (Schell, 2008). These decisions are related to a variety of aspects of the game known as the game's mechanics. In addition to appealing to the player type, keeping within the genre, and providing the desired experience, the

mechanics should be designed in such a way that they keep a player 'in flow' (Schell, 2008). A game keeps a player in flow when the game is neither too easy nor too difficult for the player. It is important to note that if a game does not keep a player in flow then the feedback will lose its value. That is, if a game is too hard, then the feedback will be unable to help the player progress, and if the game is too easy, then the feedback will serve no purpose because the player will always achieve success. Thus, in order for feedback to be valuable, the game must maintain the player in a state of flow.

An important part of balancing the game is defining the feedback a player will receive for each action they perform. In game design, this feedback is generally broken into rewards and punishments. Schell identifies nine different reward types:

1. Praise: the player's actions and prowess are complimented;
2. Points: there is an increase in the player's score;
3. Prolonged play: the player is allowed to continue to play the game;
4. Gateway: the player is given access to new parts of the game;
5. Spectacle: a visual and/or auditory presentation, such as a short video, is given;
6. Expression: the players are allowed to customize non-essential parts of the game, such as the appearance of their avatar;
7. Powers: the player is given the ability to perform new actions;
8. Resources: the player is provided with needed assets, such as money;
9. Completion: complete and total closure is provided upon winning the game.

He also identifies seven different punishment types:

1. Shaming: the game itself ridicules the player;
2. Loss of points: the player's score is decreased;
3. Shortened play: the number of chances a player has to win a game is limited;
4. Terminated play: the game ends before the player can win it;

5. Setback: the player returns to an earlier point in the game;

6. Removal of powers: the number of actions available to a player is reduced;

7. Resource depletion: needed assets are withdrawn from the player.

Schell (2008) states that games should not make use of each of these feedback options within each and every game, or they risk confusing and frustrating the player. Instead, the feedback used should be chosen based on how well it will help the target player have the desired experience. This was the approach used in designing the feedback for the digital game version of the game used in this study. Exactly how this approach was applied is discussed in greater detail in the chapter 3.

2.5. Language learning and digital gaming

Now that we have looked at the history behind serious games and the pertinent details in how they are designed, we examine more closely the current state of serious language learning games. Gee (2003) hypothesized that digital games have the potential for pedagogical applications. As discussed in the preceding section, the use of digital games for pedagogical purposes was not new to Gee, but it had fallen by the wayside in the field of SLA by the time he made this hypothesis. Since then, and often citing Gee, there has been an explosion of renewed interest in the creation and use of digital games with the explicit purpose of formal instruction (deHaan, 2011; Gee, 2007; Johnson et al., 2012; Prensky, 2001; Reinders, 2012; Squire, 2011). Many arguments have been put forward as to why digital games could be beneficial specifically as language instruction tools themselves including that they emphasize in-context activities, attention to the narrative, and goals; thus encouraging immersion in, and enjoyment of, the game (Gee, 2003, 2005, 2007, 2012; Jackson et al., 2012; Peterson, 2012; Prensky, 2001; Rama, et al, 2012; Reinders, 2012; Reinders & Wattana, 2012; Squire, 2011; Thomas, 2012; Thorne,

2008; Thorne et al., 2009). They encourage interaction between a player and the game and, often, between various players (Gee, 2005, 2007, 2012; Johnson et al., 2012; Peterson, 2012; Prensky, 2001; Rama et al., 2012; Reinders, 2012; Reinders & Wattana, 2012; Sundqvist & Sylvén, 2012; Sylvén & Sundqvist, 2012; Thomas, 2012; Thorne et al., 2009). They provide all of the tools and information learners need in order to succeed while at the same time remaining challenging (Gee, 2003, 2005, 2007, 2012; Reinders & Wattana, 2012). They are highly motivational and encourage experimentation and perseverance (Gee, 2003, 2005, 2007, 2012; Johnson et al., 2012; Rama et al., 2012; Reinders, 2012; Reinders & Wattana, 2012; Sylvén & Sundqvist, 2012). They provide continuous, helpful feedback (Gee, 2003, 2005, 2007, 2012; Jackson et al., 2012; Johnson et al., 2012; Prensky, 2001; Sykes & Reinhardt, 2012). They always involve learning on the part of the gamer and can thus be used to directly and naturally link in-class and out-of-class learning (Gee, 2003, 2005, 2007, 2012; Reinders, 2012; Squire, 2011). They adapt to the needs and skills of each individual (Gee, 2003, 2005, 2007, 2012; Jackson et al., 2012; Sundqvist & Sylvén, 2012; Thorne et al., 2009). Finally, they put the player in control (Gee, 2003, 2005, 2007, 2012; Peterson, 2012; Thomas, 2012).

These arguments are based largely on anecdotal evidence of language learning occurring when a SLL is playing a commercial digital game in either their native language or in the target language (Cornillie, Thorne, & Desmet, 2012b; deHaan, 2011; Thorne et al., 2009). Nevertheless, there is a growing number of empirical studies that test many of these theories, as discussed below. Other researchers argue instead in favor of the use of digital games not as the means of instruction, but rather as an excellent source of material for students to discuss in groups in an SLL class (Arnseth, 2006; Carrier, 1991; Coleman, 2002; deHaan, 2011; Jordan, 1992; Reinders, 2009; Squire,

2002). This latter line of thought will not be explored further herein as we are interested rather in the potential use of digital games as stand-alone language instruction tools.

Each of the potential benefits of digital games are now discussed with reference to one of the most successful and commonly referenced digital games of all time, *World of Warcraft* (WoW) (Thorne, Fischer, & Lu, 2012). WoW encourages players to become immersed in the game world by heavily emphasizing in-game narratives and goals. An example of one of the more predominant in-game narratives is the fate of the main character Prince Arthas, which is referenced often throughout much of the game. Players often become so involved in the story that they will go out of their way to learn more about what eventually happened to the prince. One of the more commonly identified goals in WoW is to capture the enemy's flag three times before they capture yours three times. This is extremely engaging because it elicits a strong sense of competition, which is further strengthened by a storyline that divides the two primary factions. It is also important to point out that the narratives to which a player is exposed and the goals that are placed before them are constantly changing and always under the control of the player. Interaction is strongly encouraged in WoW in many ways, including most notably through the formation of 'Guilds'. Guilds are groups of players who meet together within the game on a regular basis to accomplish predetermined tasks, such as killing the enemy faction's leader (Rama et al., 2012). In such large-scale guild activities, team coordination is essential, facilitated by multiple chat, trade, and messaging functionalities. One way motivation is encouraged in WoW is by having higher-level characters look much more impressive than lower-level characters. This visual reward for progressing in the game is perhaps the predominant reward within WoW. The more a player advances, the larger their weapons and armor become. Moreover their weapons and armor often gain other special visual aspects such as a glowing aura or sparkles.

These visual changes are extremely rewarding because all of the hundreds of players in a WoW world can see how fearsome one's player looks thanks to the progress one is making. Other benefits of making progress include becoming exponentially more powerful, and obtaining achievements and rewards for completing specific difficult tasks. Rewards may include things as simple as vanity pets, but may also include objects as impressive as giant dragon mounts. However, in order to obtain such success, WoW players must learn a great deal about the game. Some things that players can learn about in order to progress include geography, politics, racial and cultural differences, zoology, botany, geology, history, magic, religion, occupations, weaponry, armaments, finances, and social norms relevant to the game. WoW is also known for how well it adapts to the skill of each player, which is accomplished by limiting where players can go with their avatars, while still providing them with the sense of agency in the matter. This is accomplished in two primary ways: by (1) making 'mobs' (computer controlled, non-special characters) in zones that are of a level too high for the player kill them on sight, which discourages them from staying in that zone until they are ready; and (2) creating situations of mobs in zones that are of a level too low for the player, which does not provide any experience points when killed. This factor discourages players who are too powerful for a certain zone to move on to better environments where they will actually be rewarded for their power. Nevertheless, the players still have the option of going into zones that are not appropriate for their level; in fact, this is a fairly common occurrence. Players may engage in 'corpse dragging' (entering a zone that is too difficult, running as far as possible before being killed, then coming back into the game at that point and again running as far as possible, etc.), as well as 'ganking' (entering a zone that is too low level and slaughtering all the low level players found there). The player in WoW is also provided with agency in many other ways. These include allowing players to decide what

race they are, what professions they pursue, what quests they will and will not complete, and how they will spend their hard-earned gold.

As just shown, good digital games, such as WoW, do indeed follow all of the principles indicated by theoreticians interested in the development of serious games. Thus it is easy to see why researchers have begun looking for ways to incorporate digital games into the SLL classroom. To date, most of this research has been theoretical, but several notable empirical findings have been made. Among these are the findings that: (1) digital games do seem to be able to teach SLLs linguistic targets, including vocabulary (Cornillie et al., 2012a; Neville et al., 2009; Rama et al., 2012; Sundqvist & Sylvén, 2012; Sylvén & Sundqvist, 2012); (2) learners seem to be more willing to communicate with each other when such communication is mediated by a digital game (Reinders & Wattana, 2012); and (3) learners do not seem to enjoy custom-made digital games as much as was anticipated (Cornillie et al., 2012a; Neville et al., 2009). Thus, empirical research has confirmed the theoretical ideas that digital games can be effective at helping SLLs acquire target linguistic features, and motivate them to participate more fully than they would in class. However, more research is needed to determine how custom digital games can best be designed in order to be more effective at increasing student enjoyment when used within the SLL classroom. The question of motivation is not central to the present dissertation, but due to its importance in the literature, it is addressed in the research questions and data related to the issue of motivation are examined herein.

This research is critical because administrators and teachers are increasingly demanding that publishers provide them with online materials, such as digital games. But if game developers cannot create the games in such a way that they will be enjoyable, then these materials will lose much, if not all, of their value as pedagogical tools.

2.6. Research questions

As mentioned in the introduction above, a study like the present one is needed for several reasons. First, there is an important gap in the literature regarding which of these two types of feedback is most effective at promoting SLL, and SLVA in particular, via a digital game (Cornillie et al., 2012a; Markey et al., 2008). Second, the results from this study will help either confirm or challenge the merits of digital games for pedagogical purposes as espoused by theoreticians (Gee, 2012; Thorne et al., 2009) for SLVA. Third, the results will help guide the ever-increasing development of online SLVA materials by publishers, who are being motivated by requests from administrators and teachers. Fourth, the results will help clarify further the question as to whether digital games designed for pedagogical purposes can indeed motivate and benefit adult SLLs (Cornillie et al., 2012a; Neville et al., 2009). Fifth, although there are many digital games available that intend to teach vocabulary, there is little empirical research to support them collectively, and none to support specific games or game design principles. Finally, the results will help SLL educators make more informed choices when choosing between different available materials and pedagogies for SLVA. The results from this study will be especially useful because this study is focused on design principles rather than specific design choices. This means that the results from this study should be applicable across a wide range of SLVA digital games, regardless of the specific context or content of any given digital game.

Additionally, as also pointed out earlier, SLL research tends to focus on one subfield of the language learning process at a time, because (1) it is possible that each subfield will benefit differently from each pedagogical approach, and (2) it is difficult to evaluate learner performance with regard to each of these aspects of the language at once. In this project I maintain this tradition by focusing on a single area of language learning. I

have chosen vocabulary as the focus of study because: (1) designing a game based on currently accepted best practices in SLVA pedagogy, as discussed above, is much simpler than designing one based on the currently accepted pedagogy of the other subfields; and (2) vocabulary is an important component of SLL (Nation, 2001; Sundqvist & Sylvén, 2012).

The present study investigates the potential difference in the pedagogical effectiveness of using traditional SLL classroom feedback mechanisms and classic digital game feedback mechanisms within a digital game in answer to the call put forth by Cornillie et al. (2012a). Traditional SLL classroom feedback is represented within this game by the most commonly occurring type of feedback found in language classrooms: recasts (Lyster & Ranta, 1997). The second type of feedback, classic digital game feedback, is implemented based on the techniques and principles discussed by Schell (2008), Salen and Zimmerman (Schell, 2008), and Reinhardt and Sykes (2012). The feedback in MuCo was designed to follow the principles in which there is: (1) equal emphasis on negative and positive feedback; (2) playful feedback; (3) no metadata as part of the core feedback, but rather as optional material; (4) feedback to show that the learner's choices have made a real difference; (5) feedback that provides just-in-time information as needed by the learner; (6) continuous feedback; (7) 'juicy' feedback in which learners are rewarded in multiple ways at once; (8) feedback with endogenous value, or importance for the game itself; (9) feedback usually in the form of as many rewards as possible; and (10) as little feedback in the form of punishments as possible, with an emphasis on positive feedback. Therefore, according to the taxonomy proposed by Reinhardt and Sykes (2012), this study falls under the umbrella of Game-based L2 pedagogy research. The two digital game designs I use are identical in all regards other

than the feedback employed, and are designed to help learners acquire food vocabulary in Spanish.

The research questions posed for this study are:

(1) Does MuCo help participants learn the target vocabulary?

(2) If so, does the MuCo game with traditional classroom feedback (Lyster & Ranta 1997) or the feedback designed according to accepted digital game design principles (Schell 2008) help participants learn the target vocabulary more effectively?

(3) How well does MuCo motivate participants, as evidenced by the participants' self-reported enjoyment of the game and the extent of their use of the game?

(4) Can a participant's improvement from pretest to posttest, perceived helpfulness of the game, and/or reported enjoyment, vary according to self-reported gaming habits, motivation, and/or personality traits either collectively or according to feedback type?

It is difficult to hypothesize at this point exactly what the present study will find as answers to these questions due to the scarcity of similar empirical research. Nevertheless, with regards to the first question of whether digital games will be helpful for vocabulary learning, I hypothesize that these games will indeed be successful at helping participants learn the vocabulary, as suggested both by theoreticians and the few other related empirical studies (Cornillie et al., 2012a; Gee, 2003, 2005, 2007; Glazer, 2006; Heift, 2001; Heift & Chapelle, 2012; Johnson et al., 2012; Nation, 2001; Nelson, 1998; Prensky, 2001; Purushotma et al., 2008; Squire, 2011; Thorne, 2008; Thorne et al., 2009; Van Bussel, 1994). The effectiveness of the games will be determined by whether participants from both groups score higher on the posttest than on the pretest.

With regards to the second question of which feedback type will be more successful, I will assume the null hypothesis: neither feedback type will be more effective. I take this position because there is no other empirical work available on which to base a conclusion, and because this scenario is a hybrid of the two different fields of SLL and digital games. Both of these fields have developed and extensively tested their respective feedback techniques within their corresponding environments. Within SLL, different feedback techniques have been tested in empirical studies, while in digital gaming, different feedback techniques have been tested during game development by play-testing different options, and copying and refining techniques that have been used in existing, successful, digital games. However, the present study includes a hybrid environment and it is therefore difficult to determine which methodology will result as more effective.

The third question of whether the games will be motivating is perhaps the most difficult to answer and is only included herein due to the great interest in this question within the field of game-based language learning. On the one hand, theoreticians have made strong arguments to the effect that digital games should be highly motivating for language learners. On the other hand, the few other empirical studies that have looked at the relationship between SLL digital games and motivation have found the opposite; namely, that the participants have not enjoyed the games (Cornillie et al., 2012a; Neville et al., 2009). Moreover, several pilot studies that I conducted before beginning work on this project also suggest that students do not enjoy pedagogical digital games as much as theoreticians have proposed they would. Nevertheless, in each of these studies, the game was either played outside of the classroom context or playing the game had no relevance to the participants' grades for their language class (Cornillie et al., 2012a; Neville et al., 2009). In the present study participants either receive a 100 as an optional quiz grade, or a

1% bonus added to their final grade for the course upon completing their participation in this experiment, depending on which university they are attending. Moreover, the material targeted in this game is specifically tailored to include vocabulary that students will need to learn in order to do well in their first- or second-semester Spanish course. Thus this game will have more endogenous value within the actual class. Therefore, I hypothesize that students will report enjoying the game and demonstrate their motivation by playing the game more than is required.

Finally, with regards to the last research question about the potential effects of differences among the participants, it is again difficult to hypothesize as to what the findings will be due to the lack of both previous theoretical work and previous empirical work. Additionally, as with the question about motivation, this research question is only included due to its theoretical interest and not because its answer is central to the present dissertation. Nevertheless, because this topic has not been looked at extensively in previous works and because the arguments in favor of the use of digital games as educational tools do not make any claims as to variable effectiveness based on player differences, I once again assume the null hypothesis and predict that there will be no differences in the effectiveness of MuCo nor in the motivational qualities based on player differences. If this is true, this would be an important finding because it would indicate that digital games are equally effective regardless of the individual differences of the learners.

To conclude this chapter, although I am making a concerted effort to make the findings from this study as generalizable as possible, I emphasize that it is important not to overgeneralize the results. Future studies will be needed in order to determine whether the findings from this study hold true if certain key variables are changed. These variables include the frequency of the vocabulary items, the semantic families from

which vocabulary targets are pulled, and the number of vocabulary items taught per game. We must also consider the importance of comparing different game design principles with their classroom counterparts, comparing the results if the game is designed to teach as opposed to review material, and comparing results for games that are designed to teach other aspects of a language, such as grammar or pragmatics. Unfortunately, answers to these questions are beyond the scope of this paper.

2.7. Summary

In this chapter I have reviewed other background literature that was necessary in order to carry out the present study. This included literature related to the Communicative Approach, Task-Based Language Teaching, Second Language Vocabulary Acquisition, the history of digital gaming and serious games, the relevant design principles behind the development of commercial digital games, and the current state of the use of digital games as language instruction tools. Moreover, in this chapter I also presented reasons why a study such as the present one is needed, my research questions, and hypotheses about what the answers to those research questions could be. In the next chapter I discuss the tools and methodologies created and employed to carry out the study.

Chapter 3: Methodology

In this chapter I describe in detail the tools and methodology employed in the current study to answer the indicated research questions. First we look at the participants and how they were recruited. The pedagogical content that was used is then presented, followed by an in-depth description and discussion of the game and questionnaire developed for this study. Finally, I give a detailed account of the procedures I used to carry out the data collection.

The methodology employed was constrained by the effort to minimize the effects of confounding variables. To this end, both quantitative and qualitative data were collected in a mixed-methods approach. The study was also designed to make the procedures the participants followed as much like an actual class activity as possible, which is important because the purpose of this investigation was to determine the answers of the research questions within a classroom, and not within a laboratory, setting.

This experiment was carried out during three separate semesters. The first semester that data collection measures were used, the spring of 2013 (SP13), various defects came to light in the materials and procedures. Several students who participated in SP13 also made some recommendations on how to improve the materials in minor ways, which were incorporated in the materials and procedures used the following semester. Unfortunately, during this second data collection in the summer of 2013 (SU13), the number of participants was small. As a result, I carried out the experiment a third time two semesters later, during the spring of 2014 (SP14).

During the intervening semester in which no data collection took place, one of the universities changed their textbook and the other changed their language program coordinator. As a result, for the third attempt, data were collected from only one

university, and the content of the game had to be adjusted to match the content within the newly adopted textbook. A distinction is made in this chapter, where appropriate, between the different methodologies and materials used for each of these three semesters.

3.1. Participants and recruitment

3.1.1. Spring 2013

During SP13 I carried out this study at two large public universities in the southwestern United States: the University of Texas at Austin (UT) and the University of New Mexico (UNM). Each spring and fall semester approximately 600 – 800 students enroll in UT's first-semester Spanish course. Several hundred more enroll in the equivalent courses at UNM. These groups formed the pool from which participants were recruited.

All participants at UT were offered the opportunity to play MuCo in exchange for an 'all-or-nothing' extra quiz grade. Students were also allowed to do a worksheet provided by their instructor in place of participating in this study in order to receive the course credit. The IRB at UT required that this alternative method of obtaining the course credit be offered to the students. No record was obtained as to how many students used this alternate method. I recruited at UT in SP13 by visiting all first-semester Spanish courses at UT and giving a brief presentation about my study, after which I passed around a sign-up sheet for any students who were interested. I then emailed the students who signed the sheet with the link to the game and the link to the questionnaire. That message also requested that participants not access the questionnaire until after completing the posttest within the game.

The students at UNM were offered a 1% bonus to be added to their final course grade for participation in this study. These students were also provided the option of

completing the worksheet in order to receive this in-class credit instead of participating in my study. Recruitment at UNM was handled by the language program coordinator, who emailed all the students in the appropriate courses with a recruitment email that I had prepared. This email contained a flyer and links to the game, the worksheet, and the questionnaire.

Because students were offered significant in-class credit for participation in my study, I was hopeful that I would get a much better response than I had received in previous pilot studies that had not offered course credit. In order to receive the in-course credit, the participants needed to participate before they began work in class on the vocabulary covered within the game. This requirement was necessary so that participants would not be 'contaminated' with exposure to the target vocabulary outside of the game. Nevertheless, it is possible that this requirement may have lowered the number of participants as they were not yet working on the relevant material in class, which means that they needed to study other material for their class. On the other hand, all participation was done completely online so that participation could be as natural as possible, and students could access the materials at their leisure. I wanted them to do so when and where they would normally complete online homework. I also hoped that this access would inspire a higher rate of participation. In any case, in SP13 only 176 students actually accessed the game. Of these participants, anyone who did not complete enough of the game, did not finish the questionnaire, or used outside materials was excluded from the rest of the study. The result included 46 viable participants from SP13. Due to a flaw in the game (as discussed below), no participants were excluded for prior knowledge of the vocabulary because the game failed to record the pretest and posttest data correctly. Moreover, the game also failed to record what type of feedback each participant received, and therefore the data collected from these participants were not applicable to any of the

research questions that looked at pretest/posttest scores or type of feedback. Finally, no data were collected about the participants' previous study of languages, language background, age, gender, grades, or GPA during this iteration.

3.1.2. Summer 2013

Because there were far fewer sections of the target course in the summer at UT and only one section of the appropriate course at UNM, I decided to expand the recruitment pool. Therefore I created three recruitment videos: one video only for UT, a general video promising credit, and a general video that did not promise credit. Each video contained hyperlinks that would take viewers directly to the game. No direct link was provided for the questionnaire used in the summer; instead, the link was provided within the game itself. I forwarded the link to the first UT video to all targeted instructors at UT. They then showed the video to their classes and forwarded the link to the video to each member of their class. Next, I forwarded the link to the second recruitment video to the instructor at UNM, who also showed that recruitment video to the students in the class and then forwarded the link to the recruitment video to them. Finally, I emailed a recruitment message with both of the later videos to the Spanish departments at each university and college that had used the same textbook as UNM within the last year. The message asked them to forward the link to the appropriate video to the members of their classes, with the anticipation that one or two students from most of the 150+ institutions would participate. Had I been successful, this measure would have helped offset the fact that there was a much smaller potential pool of participants at UT and UNM during the summer than there had been during the previous spring.

Once again all participation was online, students were provided with the worksheet as an alternative activity, and participation was required to take place before

the target vocabulary was covered in class. Unfortunately there was no participation from any of the outside universities that were contacted. There was a total of 70 students from UT and UNM that accessed the game. Once again students who accessed outside materials or did not complete the required portions of the study were eliminated. Additionally, no data were collected about the participants' previous study of languages, language background, age, gender, grades, or GPA. However, during SU13 the game did correctly record pretest and posttest data, so participants who achieved an accuracy rating of 80% or more during the pretest were also excluded from the study. The result was a total of 31 participating students for SU13.

3.1.3. Spring 2014

Data collection was carried out a final time during SP14. This time participants were recruited only from UT. Recruitment once again involved my forwarding a link to the recruitment video to the instructors of each of the sections. The instructors then showed the video in their classes and then forwarded the link to the recruitment video to their students. The recruitment video contained the link to the game, and the game provided the link to the questionnaire once a participant had completed the posttest. Once again, no data were collected about the participants' previous study of languages, language background, age, or gender. However, information was collected regarding the students' grades and GPA, but this information was not used to eliminate participants. During SP14 there were 137 students who accessed the game. After eliminating all students who had used outside materials, scored an 80% or higher on the pretest, or did not complete all required portions of the study, usable data from 45 participants resulted.

3.2. Content

3.2.1. Spring 2013 and summer 2013

During the spring and summer of 2013, instructors in the target course at UT used the textbook *Impresiones* (R. Salaberry, Barrette, Elliott, & Fernández-García, 2004), while those at UNM used *Mosaicos* (Castells, Guzmán, Lapuerta, & Liskin-Gasparro, 2009) as the primary source of content and materials for their courses. In *Mosaicos*, food vocabulary is covered in Chapter 10. One interesting aspect of *Impresiones* is that vocabulary related to food is covered twice; once in Chapter 6 and again in Chapter 10. Only food-related vocabulary that is taught in Chapter 10 of either textbook, but not before, was included in the game. Thus, no participant would have been exposed to any of the vocabulary presented in the game in any of the textbook vocabulary lists they had used up to the time of their participation in the study. This measure was taken in order to prevent any confounding effects that previous exposure to the words might cause. For example, while *agua* 'water' is presented in Chapter 10 of both books, it was not included because it is also taught earlier in both books. On the other hand, *aguacate* 'avocado' is covered only in Chapter 10 of *Impresiones*, and was therefore included. Any vocabulary words that were deemed to be too orthographically similar to their English counterparts were also excluded in order to make it as difficult as possible for participants to guess correctly. This procedure resulted in the inclusion of 23 Spanish food vocabulary items. All but one of these lemmas ranks either as occurring in low frequency rates (between the top 3,000 and 5,000 most frequent words in Spanish) or in extremely low frequency rates (does not appear within the top 5,000 most frequent words in Spanish). The single exception is *miel* 'honey', which has a relative frequency of 2990. This means that it is only 10 slots away from being of low frequency, as indicated by Davies (2006). Due to the fact that it is right on the border of being low-frequency, I kept *miel* as one of

the target words. The list of vocabulary items along with their relative frequencies can be seen in Table 3.1 below. The low frequency of these words is likely due to the fact that most high-frequency food vocabulary is taught before Chapter 10 in at least one of the textbooks. This aspect is ideal because, as indicated earlier, Nation (2001) has suggested that the best way to learn low-frequency vocabulary is through frequent and shallow practice, which is exactly the type of practice that takes place in this study.

Table 3.1: Vocabulary items in MuCo in the spring and summer of 2013

Moderate Frequency	Low Frequency	Extremely Low Frequency
Miel 'honey'	*Trigo* 'wheat'	*Aguacate* 'Avocado', *Cordero* 'Lamb', *Lentejas* 'Lentils', *Piñas* 'Pineapples', *Plátanos* 'Bananas', *Nueces* 'Nuts', *Maracuyás* 'Passion Fruit', *Cerdo* 'Pork', *Manteca* 'Butter', *Cacahuate* 'Peanut', *Aderezo* 'Dressing', *Ajíes* 'Peppers', *Pepinos* 'Cucumbers', *Mostaza* 'Mustard', *Espinaca* 'Spinach', *Cerezas* 'Cherries', *Pomelos* 'Grapefruit', *Avena* 'oats', s*alchichas* 'sausages', *aves* 'poultry', *especia* 'spice'

3.2.2. Spring 2014

The textbook used in the target course in SP14 at UT was *Conectándonos* (M. Salaberry, Barrette, María, & Nevárez, 2013). Like *Impresiones*, *Conectándonos* covers food in both Chapters 6 and 10. However, this time the content for MuCo was pulled from Chapter 6 instead of Chapter 10. Like the previous version of MuCo, any words that were deemed too orthographically similar to their English counterparts were omitted. Additionally, any words that were used in the textbook prior to Chapter 6 were also omitted. In order to determine whether a word had been used prior to Chapter 6, the first five chapters were digitized and then searched by a Python script for any occurrences of the food vocabulary from Chapter 6. Finally, any words that were among the 5000 most frequent words in Spanish according to Davies' (2006) *Frequency Dictionary of Spanish*

were also omitted. This procedure resulted in the inclusion of 40 Spanish food items with extremely low frequency rates. These words are: *aceituna* 'olive', *aderezo* 'salad dressing', *albahaca* 'basil', *albaricoque* 'apricot', *almejas* 'clams', *almendras* 'almonds', *azafrán* 'saffron', *bacalao* 'cod', *besugo* 'sea bream', *bollería* 'bread', *cacahuate* 'peanut', *canela en polvo* 'cinnamon powder', *caracoles* 'snails', *cebolla* 'onion', *champiñones* 'mushrooms', *chorizo* 'sausage', *chuletón* 'chops', *costillas* 'ribs', *espinaca* 'spinach', *fresas* 'strawberries', *gambas* 'shrimp', *helado*, 'ice cream', *langosta* 'lobster', *lechuga* 'lettuce', *lentejas* 'lentils', *mejillones* 'mussels', *melocotón* 'peach', *merluza* 'hake', *nuez* 'walnut', *ostras* 'oysters', *pasas* 'raisins', *patilla* 'watermelon', *pepino* 'cucumber', *perejil* 'parsley', *pimentón* 'pepper', *pimiento negra* 'black pepper', *porotos* 'beans', *refresco* 'soda', *trucha* 'trout', and *zanahorias* 'carrots'.

3.3. Instrument: Mundo de Comida

3.3.1. Game design

Mundo de Comida (MuCo)[2] is a browser-based, single-player, strategy game. In MuCo the players assume the role of entrepreneurs who are just starting out with their own food kiosk franchise. MuCo was custom designed and created by myself specifically for use in this study. The fact that MuCo is a browser-based game allows the learners to access the game from any device that can handle interactive, media-rich websites. This accessibility is advantageous because downloadable games need to be adjusted for the operating system of the user and it can be much more difficult to pull the data from them, while browser-based games do not have these limitations. MuCo is a single-player game for two reasons. First, having a multiplayer game would add several more variables and

[2]The development version of MuCo can be accessed on the following site: http://www.wooglie.com/games/Strategy/MuCo

levels of complexity to the study's results, possibly obscuring any findings about feedback. Second, it is far easier to design and coordinate a single-player digital game.

The basic concept behind MuCo is loosely based on the design of the award-winning board game *The Settlers of Catan*©. The player's goal when playing MuCo is to accumulate 1,000 *pesos* (a monetary unit) before any of their opponents. The opponents are three virtual, non-player characters with a very simple artificial intelligence. When a game of MuCo is started, a city map is randomly generated for that instance of the game. This map consists of 19 randomly-arranged hexagonal spaces and an un-owned restaurant, represented by a white cube, at the corner of each hexagon, as exemplified in Figure 3.1 below.

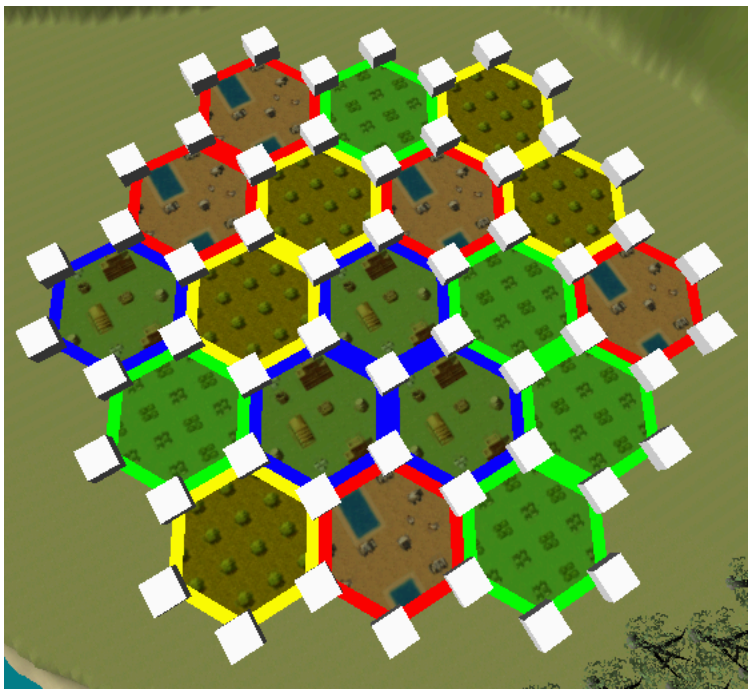

Figure 3.1: Example of MuCo Randomly Generated Map

As shown in Figure 3.1, there are four different types of hexagonal spaces. These different types of spaces represent different categories of food: meat and seafood (red), vegetables (green), fruit (yellow), and miscellaneous (blue). These food categories were chosen based on the food vocabulary taught in *Impresiones*, *Mosaicos*, and *Conectándonos*.

3.3.2. Game setup

When play begins, players are first taken to the IRB consent form, which they must acknowledge in order to play the game. They are then taken to the introduction screen as depicted in Figure 3.2 below. The purpose of this screen is to introduce the participants to the story and goal of the game, and to collect information that can be used to identify them to their instructor.

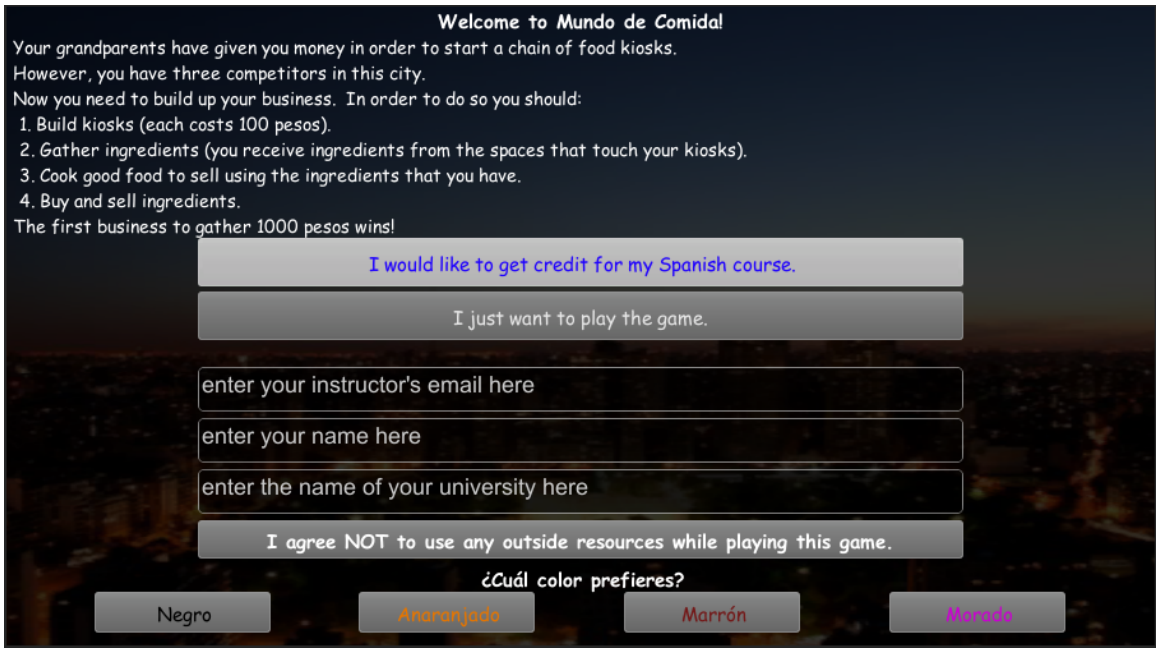

Figure 3.2: Opening screen in MuCo in summer 2013 and spring 2014

On this screen the participants are informed that their grandparents have given them 100 pesos with which to start a chain of food kiosks. They are informed that in order to establish their franchise successfully they will need to build kiosks, gather ingredients, cook recipes, and sell or purchase ingredients as needed. In SP13 all content on this page was written in Spanish, while in SU13 and spring of 2014 this content was in English, as shown in 3.2. This change was made because the participants in SP13 had difficulty understanding the instructions in Spanish. This difficulty lead to trouble during gameplay, which possibly added some confounding factors to the results obtained. Since the purpose of the study had nothing to do with participants' comprehension of the game's instructions, these instructions were translated into English in order to eliminate any such confounding variables in the subsequent iterations of data collection.

In addition to being introduced to the game, participants need to choose which color they would like to play, and to provide personally identifying information if they are playing MuCo for course credit. The chosen color is used to identify the participant's kiosks, score, and amount of resources. The personal information includes the participant's name and their instructor's email address. Players are also informed that by choosing a color they are agreeing not to use any external resources such as dictionaries, friends, etc. during the game. If they are not playing the game for course credit, they are allowed to choose which version of the game they would like to play, although the difference between the two different options is not explained to them.

3.3.3. Pretest

After they choose a color, players who have indicated that they are playing the game for in-class credit are taken to the pretest. The pretest and posttest both take place within MuCo's kitchen interface, as illustrated in Figure 3.3 below. This is the same

interface that is used during the treatment portion of the game. Before a participant can begin to use the interface for the first time they must go through a mandatory tutorial that explains the interface to the participant in detail. Additionally, before the tutorial begins for the pretest, and before the posttest begins, the players are informed, both with text and an audio recording, that their performance on the pretest/posttest will have no effect on the credit they receive or their success in the game in any way, and that they will not receive any feedback during the pretest/posttest. After viewing the tutorial the first time participants are given the option to see the tutorial again at any time that they wish.

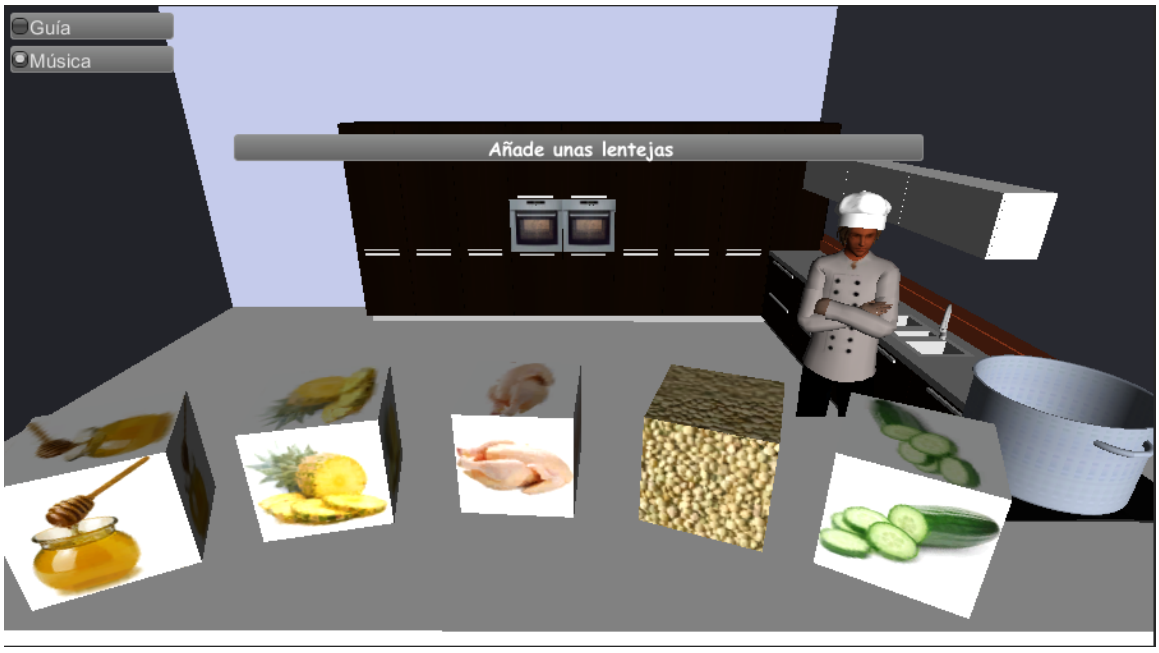

Figure 3.3: Example of cooking interface

As shown in Figure 3.3, the cooking interface displays the name of a food that the participant is supposed to identify. An audio recording of the name of the food is also played for the participant. On the bottom of the screen the participant is shown five pictures of different foods. In order to identify the designated food, the participant needs

to click on its corresponding image. The other four images are randomly selected. This procedure of food identification is repeated ten times for both the pretest and the posttest. During the pretest, participants are introduced to the kitchen interface and are tested on their prior knowledge of the target food vocabulary. I removed the results from the study by participants who demonstrated a good control over the target vocabulary during the pretest by scoring 80% or higher correct since they knew too much of the vocabulary prior to the study for their data to be of any use. As discussed below, the game records each choice each participant makes along with each corresponding correct answer, so that a score can be calculated for each participant.

3.3.4. Main game interface

After completing the pretest the player is taken to the main interface for the game, as depicted in Figure 3.4 below. In this interface the player is shown the map of the city, everyone's current score, and the number of each type of resource currently owned by each player. This is also the interface in which the participants can purchase new kiosks, buy and sell resources, and choose to attempt to cook a recipe or to end their turn.

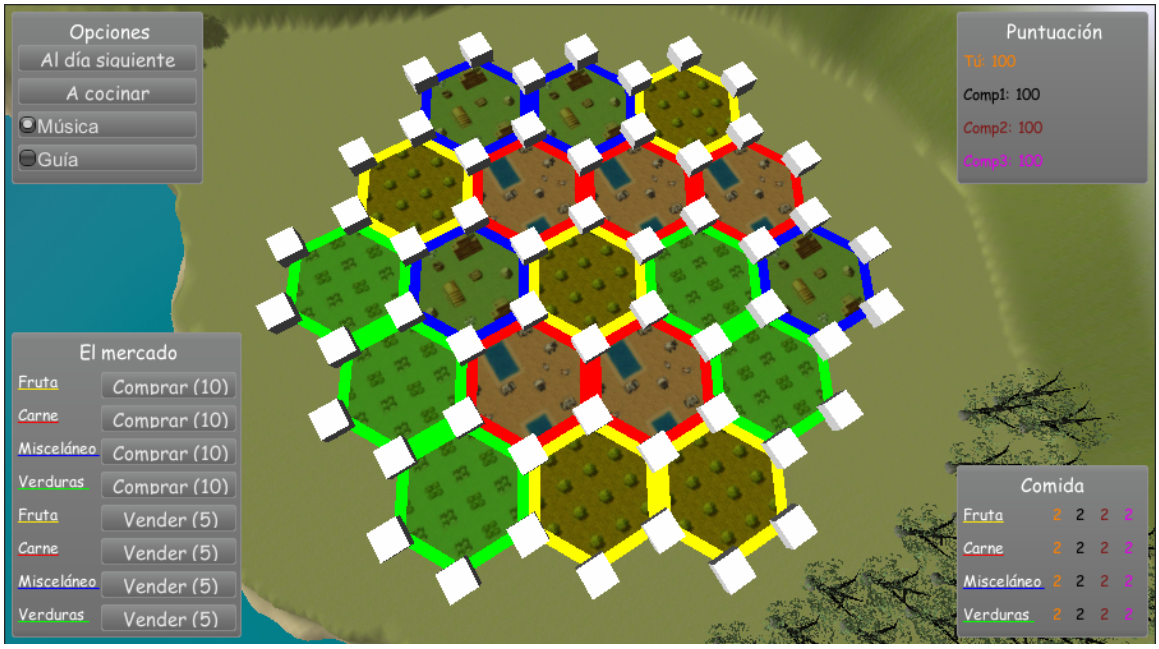

Figure 3.4: Main MuCo interface at the beginning of the game

Each player starts out with two of each resource in addition to the 100 pesos they received from their grandparents. In order to end their turn, players can click on *Al día siguiente* 'The next day'. This move allows them to receive resources from any kiosks they own. These resources are dependent on the hexagonal spaces contiguous to any kiosks they own. Thus, if they have one kiosk contiguous to a blue space, they will receive one 'miscellaneous' resource. This resource can be used to help prepare any recipe that requires a 'miscellaneous' food item. After the player is given their allotted resources, the other (computer) players take their turns. Participants can click on any kiosk that is not already owned in order to purchase that kiosk for 100 pesos. Any kiosk they purchase will turn to their chosen color in order to indicate that the kiosk then belongs to them. They can also visit the *mercado* 'market' by clicking on one of the *Comprar* 'Purchase' or *Vender* 'Sell' buttons. They pay the amount indicated in parentheses for each resource they purchase and receive the amount in parentheses for

each resource they sell. These values are dynamically adjusted throughout the game, so that resources that are purchased frequently will be more expensive and resources that are sold frequently will be cheaper. Finally, they can click on *A cocinar* 'On to cooking' in order to attempt to cook a recipe with the ingredients they started with, purchased or received from their kiosks.

3.3.5. Recipe book interface

When a player clicks on *A cocinar* 'On to cooking', they are taken to the recipe book interface, as illustrated in Figure 3.5 below. Players use the recipe book interface to select the next recipe that they would like to attempt to prepare. The interface shows the players how many *pesos* they would earn for cooking each recipe as well as how many resources of each type are needed for each recipe. In the spring and summer of 2013 there were 20 recipes that players could choose, displayed on two different pages. In SP14 there were 48 recipes displayed on five different pages. In both versions of MuCo, each individual food item is part of exactly three different recipes.

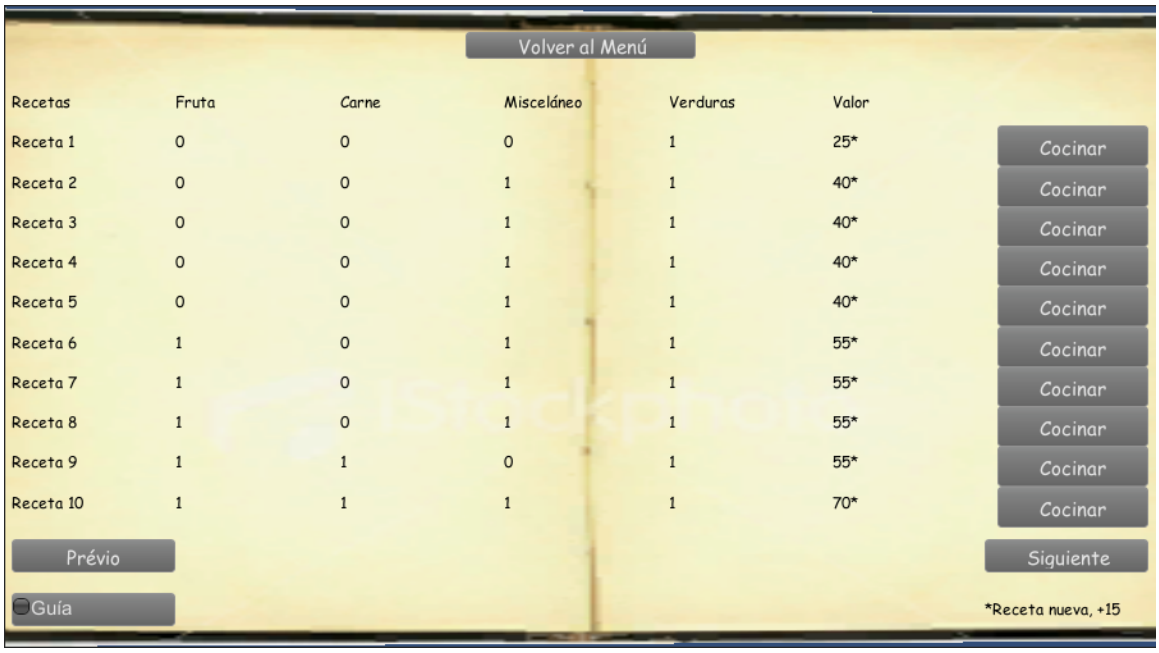

Figure 3.5: Recipe book interface for spring and summer 2013

The player can flip through the various pages of the recipe book by clicking either *Prévio* 'Previous' or *Siguiente* 'Next'. On each page is a list of several different recipes and the quantity of each type of resource required to cook each recipe. So, as shown in Figure 3.5, the recipe for *Receta 1* 'Recipe 1' requires only one *verdura* 'vegetable'. If a player does not have enough ingredients to cook a recipe, the text for that recipe is displayed in red. In the spring and summer of 2013, each recipe required between one and six ingredients, with a mean of 3.8 ingredients and a standard deviation of 1.2 ingredients. In SP14 each recipe required between one and nine ingredients, with a mean of 3.8 ingredients and a standard deviation of 1.4 ingredients. Additionally, each recipe has an indicated number of *pesos* that the player would receive for successfully cooking that recipe. These values are dynamically adjusted throughout the course of the game. The adjustments are made such that each time a specific food item, such as *aguacate* 'avocado', is used in a successful attempt at a recipe by either the player or a computer,

the value of all recipes that contain that item is decreased. The purpose of this game mechanic was to make it more tempting for players to attempt recipes that had food items to which they had not yet been exposed. A value with an asterisk next to it has a bonus of ten *pesos* added into the value because the player has not cooked the recipe before. The purpose of this game mechanic was to encourage players to cook as many different recipes as possible. Once the player has chosen a recipe they would like to cook, they simply click on the *Cocinar* 'Cook' button at the end of that recipe's description in order to be taken to the cooking interface for that recipe. This cooking interface is identical to that used in the pretest (see Figure 3.3 above).

3.3.6. Cooking interface

Like the pretest, in the regular cooking interface the player sees and hears the name of one food item at a time. At the bottom of the screen, they see the identified food item randomly positioned among four other randomly selected food items. The difference between this cooking interface and the pretest's cooking interface is that in this cooking interface, players receive feedback according to their performance. Four different types of feedback are provided according to the group the player is in and whether they correctly identified the food or not. The various feedback types are identified and illustrated in Table 3.2 below.

Table 3.2: Types of feedback given according to group and correct identification of food item

	Classroom Feedback Group (CFG)	Digital Game Feedback Group (DGFG)
Food Identified Correctly	Feedback: Spoken and written congratulations Example: ¡*Muy bien*! 'very good!'	Feedback (short term): Customized spoken and written confirmation accompanied Example: ¡*Qué tomate más bonito*!, ¿*no*? 'What a pretty tomato, right?' Feedback (medium term): A positive visual and oral cue is provided. Example: (The player sees a picture of the recipe they have just cooked and hears the chef congratulate them on their success.)
Food Identified Incorrectly	Feedback: Spoken and written recast Example: *No, aquí está el tomate*. 'No, here is the tomato.' (The picture of the tomato is enlarged and a button appears. When the player presses the button they are taken back to where they left off in the recipe.)	Feedback: Spoken and written outcry accompanied by negative visual and oral cues as well as a resource penalty Example: ¡*No! ¡La cereza no*! 'No, not the cherry!' (The pot is seen to explode accompanied by the sound of an explosion, the player loses the resources they have put into the pot so far for this recipe, they are shown an image of the dish they were trying to cook, along with the incorrect food item super-imposed on it. The player is then given the option is be told what the correct food was or to be returned to the main interface. In either case they are not allowed to finish the recipe.)

Each feedback option was created based on the principle for the feedback for that group as discussed in section 1.2 above. These principles are repeated here in Table 3.3 for convenience.

Table 3.3. Principles of CF and DGF		
	CF	DGF
Frequency of Feedback	Provided in response to an error.	Provided for each and every player action.
Mode of Feedback	Verbal.	'Juicy'.
Practical Integration of Feedback	Not integral to what happens in class.	Has endogenous value.

The CF principles state that CF should be provided in response to an error, should be spoken, and should not be integral to what happens in class. On the other hand, the DGF principles state that DGF should be provided for each and every player action, should be 'juicy', and should have endogenous value. I will now discuss how these principles were used in the creation of all four feedback types.

To mirror how CF is provided in the classroom, the CF group in this study received spoken feedback. However, it is also a known fact that many individuals who play digital games do so with the sound turned off (McGowan, 2011; Totilo, 2011). Thus, if the CF group received only oral feedback, it was very likely that several participants would in fact not receive the feedback. In order to counter this possibility I included captions of all spoken text in the game so that those who played without sound could still see the feedback in written form.

3.3.7. Negative CF

The negative feedback provided to the CFG consists of a written and oral message that indicates that an error has been committed and an enlargement of the picture of the correct answer. Hence, the player is essentially provided with a recast of the vocabulary word. The participant is required to acknowledge this feedback by clicking on a button. When players click the button, they are asked to identify the same food item from among

the same selection of images. This process repeats until the participant chooses the correct image. An example of this feedback is illustrated in Figure 3.6 below.

Figure 3.6: Negative feedback for the CFG

The negative CF follows the first principle of CF because it is provided after an error. This feedback design complies with the second principle because it is oral in nature. However, it was also necessary to provide the linguistic portion of the feedback in written form due to the fact that many individuals play digital games without sound (see 1.1.2 above). Finally, this feedback design complied with the third principle because it did not truly affect what happens in the game. That is, whether the player receives positive or negative feedback, there is no difference in the number of resources consumed for the recipe or the number of *pesos* earned for cooking the recipe. The decision was made to use a recast for this feedback mechanism because recasts are (1) the most common type of negative feedback found in the classroom, (2) especially common with

vocabulary, and (3) designed to help learners associate the correct picture with the indicated food vocabulary.

3.3.8. Positive CF

For the CFG, positive feedback is simply a form of written and oral congratulations, as shown in Figure 3.7 below. In order to move on after receiving this feedback, the player acknowledges the feedback by clicking on a button. At this point the players are either taken to the next ingredient in the recipe or, if they have finished the recipe, they are taken back to the main game interface.

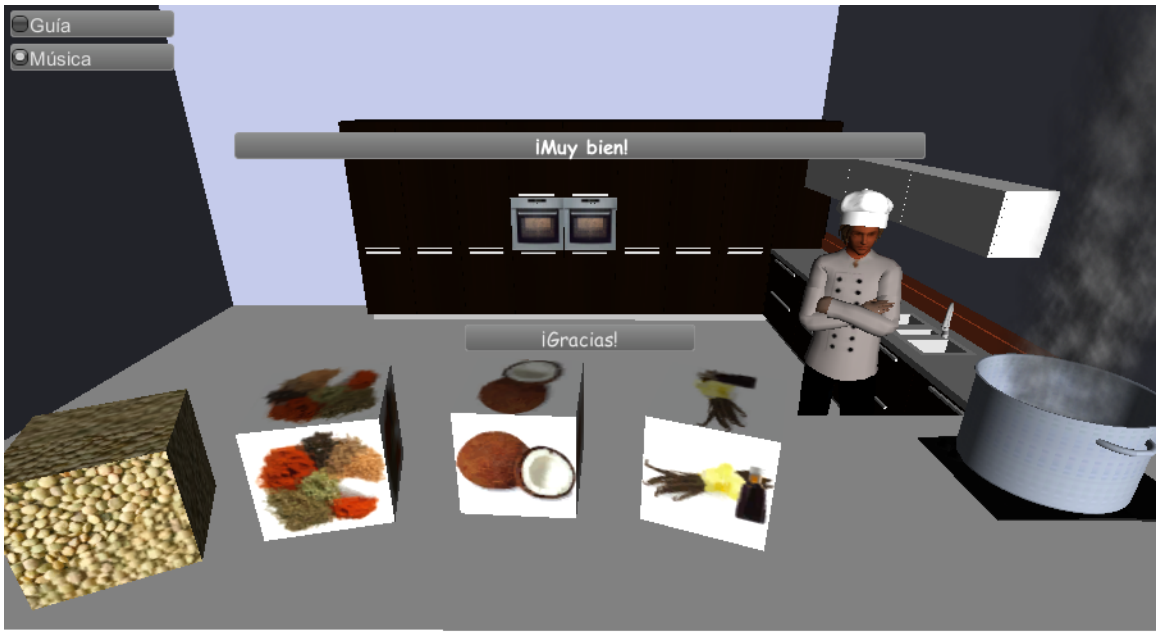

Figure 3.7: Positive feedback for the CFG

Admittedly the mere existence of positive feedback for the CFG goes against the first principle of CF, which is that feedback should be provided only after an error. Unfortunately, this problem was unavoidable because of the fact that negative feedback was provided consistently after every single error. Thus, even if no explicit positive

feedback were provided, the absence of any negative feedback would in effect serve as implicit positive feedback. The only way to avoid this result would have been to provide inconsistent negative feedback. However, because this game was being used as a teaching tool for actual classroom learning, I decided it would be almost unethical to provide the negative feedback inconsistently.

In spite of non-compliance with the first principle of CF, the positive feedback supplied to this group did comply with the other two principles. It followed the second principle by being spoken although, like the negative feedback, it was also provided in written form in order to address any participants who played with the sound turned off. This feedback complied with the third principle in the same way that the negative CF did. That is, regardless of whether the participant made a correct choice or an incorrect choice, there was no effect at all on gameplay.

This feedback was also not customized in any way, which means that all the participants received the exact same message of *¡Muy bien!* 'Very good!' each time that they made a correct choice. This is similar to the way that positive classroom feedback is often provided. Thus the positive feedback provided is deemphasized, which also reflects the treatment of positive feedback in the literature on SLVA (Laufer & Nation, 2012; Nation, 2001). The idea here seems to be that the players do not need a lot of positive feedback because they have already correctly identified the answer. Instead of receiving feedback they need to continue so that they can practice other, potentially more troublesome, vocabulary.

3.3.9. Negative DGF

When a player in DGFG does not correctly identify a food, there is an oral and written outcry accompanied by a negative sound and a negative image (the sound and

image of an explosion), as well as a depletion of the player's resources (they lose all resources they have already used for this recipe), as shown in Figure 3.8 below and exemplified in Table 3.1 above. In SU13 and SP14, the players then had the choice to acknowledge the feedback and move on with the game, or to have the correct answer identified for them and then move on, while in SP13 they did not have the option to have the correct answer identified for them and could only move on.

Figure 3.8: Negative feedback for DGFG

In either case they are then provided with more negative feedback before being taken back to the main game interface. This time, the players are shown the food that they incorrectly chose on the recipe they were trying to cook, as well as a brief textual message expressing the cook's disgust with the result of the player's efforts to cook, as shown in Figure 3.9 below. At this point the player can click on a button to be taken back to the main game interface.

Figure 3.9: Negative end-of-recipe feedback for DGFG

This negative feedback is designed in accordance with the first principle of DGF in that it is provided after player errors, which are the result of one possible player action. Additionally, this feedback complies with the second principle of being 'juicy' by utilizing speech, written text, images, sound effects, a depletion of the player's resources, and the termination of the attempt to prepare the recipe. Moreover, the 'juiciness' of the feedback is enhanced by the fact that almost all of the feedback elements are dependent in one way or another on the player's specific recipe and food choices. Finally, because two of the feedback elements of this feedback event were a depletion of the player's resources and a termination of the current attempt to prepare the recipe, this feedback directly affected what was happening in the game, and thus had endogenous value.

3.3.10. Positive DGF

There are two different positive feedback events for the DGFG. The first event takes place each time the player correctly identifies a food. When this happens, there is an oral and written confirmation of their choice, as shown in Figure 3.10 below. The player can then move on by clicking a button.

Figure 3.10: Positive feedback in DGFG

The second positive feedback event occurs when a player successfully completes an entire recipe. When this happens, an image of the recipe that the player has just finished is displayed along with a positive oral and written message from the cook on the quality of the end product, as illustrated in Figure 3.11 below. Moreover, the player receives the number of *pesos* that the recipe was worth. The player can move on to the main game interface by clicking on a button.

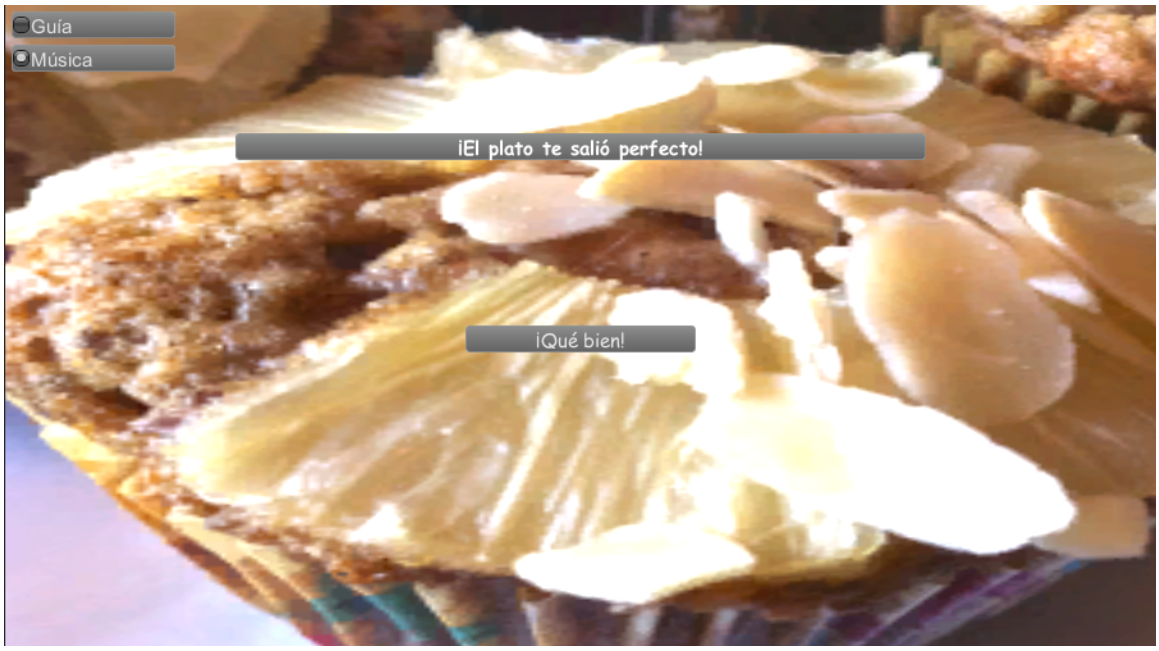

Figure 3.11: Positive end product image

Once again the feedback is designed in accordance with the DGF principles identified earlier. The positive feedback complies with the first principle by being provided each time the players correctly select a food and each time they complete a recipe, which are both player actions. The second principle of 'juiciness' is fulfilled by the use of several different feedback elements, including speech, written text, images, and an increase in resources. Finally, the third principle is satisfied in both of these feedback events by directly affecting what happens in the game. In the first event, which takes place when players correctly identify a food, they are allowed to continue on with the recipe. This is in contrast with the negative feedback that they would have received after an error that would have terminated the recipe prematurely. The second event also has endogenous value because the player's number of *pesos* is increased.

3.3.11. Posttest and questionnaire

After the player has attempted to cook 10 recipes, which was estimated to take between 20 and 30 minutes, participants who were seeking in-course credit were taken to the posttest. This test is identical to the pretest in that it uses the same cooking interface, feedback is not provided to the players on their performance, and their performance has no effect on the rest of the game or the in-class credit they are being offered. After they completed the posttest, players in SP13 were asked to complete the questionnaire and were then taken back to the game. They had received the link to the questionnaire in the recruitment email. On the other hand, after completing the posttest in SU13 and SP14, players were provided with a hyperlink within the game that took them directly to the questionnaire. This questionnaire was presented as a Google Form, as shown in Appendix A. Upon completing the questionnaire, players were informed that their official participation had ended and that their instructor would be informed that they had successfully completed the optional assignment. They were also informed that they now had the option of continuing to play the game.

3.3.12. Data collection

The questionnaire responses were automatically saved to a Google spreadsheet by the Google form. During gameplay, the game recorded information to an external database when the players chose which color they wanted to play, and whenever they clicked on the image of a food item. When players chose their color at the beginning of the game, the game recorded the time, their name, their instructor's email, whether the game was being played for course credit and, for SU13 and SP14 participants, which type of feedback would be administered, to the external database. Each time the player clicked on a food, the game recorded the time, the player's name, the food clicked, the recipe being cooked, the target food that the player was supposed to click, and whether they

were cooking for the pretest, posttest, treatment, or post-treatment part of the game, to the external database. In the SP13 version of the game the correct answers for the pretest and posttest items were not recorded correctly. This flaw was fixed for the SU13 and SP14 groups. All of the information recorded by the game was stored in a database accessible only by myself. These data were used in order to answer the research questions, as discussed in the following chapter.

3.4. Procedure

Although MuCo is designed as a single-player game, it was possible that several hundred participants would be logged on to the game on the same server at the same time. Consequently, the game was hosted on a specialized game server that is designed to accommodate such activity. Individuals accessed the game either through the hyperlink in a recruitment video or through a hyperlink in a recruitment email. At this point participation was fairly straightforward. Participants simply needed to follow all in-game directions as detailed in 3.3 above.

The decision to allow players to access the game remotely has its advantages and disadvantages. One advantage was that it was more likely that there would be more participation since students did not have to schedule a time to come in to a lab. Also, the study could be conducted with ease at more than one university because face-to-face interaction with the players was not required. The most important advantage was that the data collected more correctly reflect 'real-life' use on the part of the learners, or the way in which they would use the materials within a normal class setting. For example, I am able to see what time of day and how close to the deadline students chose to play the game.

The primary disadvantage of conducting the study remotely was not being able to confirm who the participants were or whether they were using outside resources, such as the textbook, a dictionary, or Spanish-speaking friends. It was hoped that these risks were minimized by the fact that accuracy would not affect their in-course credit received and that they had agreed not to allow others to play for them or to use external resources themselves when they accessed the game.

Students were advised of the opportunity to participate in the study one week before they began work on the corresponding vocabulary in their class. They were told they could participate in the study until they began work on the target chapter. I emailed the instructors the information about which of their students should be given course credit on the day they started work on the target chapter. Then I collected, compiled, and analyzed all of the data that had been recorded by the game and the questionnaire. The results of these analyses are presented in the following chapter.

3.5. Summary

In this chapter I have discussed in detail the procedures and materials used for this study. This discussion covered the topics of who the participants were, how they were recruited, how they participated in the present study, how the game and questionnaire were designed and built, and how the data were recorded for this study. Each of these topics was discussed with reference to each of the three semesters in which this study took place. The primary purposes of discussing the materials and procedures are to aid with interpretation of the results, and to allow for duplication of this study by other researchers.

Chapter 4: Results

4.1. Introduction

In this chapter I discuss the results from the research carried out for this project. I respond to each of the research questions proposed in Chapter 2, and repeated below, in turn.

(1) Does MuCo help participants learn the target vocabulary?

(2) If so, does the MuCo game with traditional classroom feedback or the feedback designed according to accepted digital game design principles help participants learn the target vocabulary more effectively?

(3) How well does MuCo motivate participants, as evidenced by the participants' self-reported enjoyment of the game and the extent of their use of the game?

(4) Can a participant's improvement from pretest to posttest, perceived helpfulness of the game, and/or reported enjoyment, vary according to self-reported gaming habits, motivation, and/or personality traits either collectively or according to feedback type?

4.2. Did MuCo work?

4.2.1. Empirical evidence

Although the principle goal of this study is to compare the relative effectiveness of the two feedback types, it is not possible to make a valid comparison of the two feedback types before we establish whether or not MuCo itself is an effective pedagogical tool for vocabulary learning. Therefore the first research question of necessity cannot take feedback into account. The goal of the first research question was to determine whether MuCo was successful in helping participants improve in their

accuracy from the pretest to the posttest when they identified the food vocabulary used within the game, regardless of which feedback type was used. To answer this question, the average of the pretest scores for SU13 and SP14 were compared with the average of the posttest scores for these same students, using paired samples t-tests. As shown below in Table 4.1, the average pretest score was 52.41% correct in SU13 and 51.76% correct in SP14, while the average posttest scores were 82.33% correct and 74.44% correct, respectively. The results from the t-tests indicate that these differences are indeed statistically significant at $p = 0.00$ for SU13 and at $p = 0.00$ for SP14. This result suggests that, empirically speaking, MuCo is indeed successful at helping players learn the target food vocabulary.

Table 4.1: Average pretest and posttest scores

	Pretest	Posttest	Change
Summer 2013	52.41%	82.33%	29.92%
Spring 2014	51.76%	74.44%	22.67%

In order to confirm this finding, I reexamined the posttest scores in a more fine-grained analysis. I looked exclusively at tokens that a speaker was tested on both in the pretest and in the posttest, and that they missed in the pretest. I compared the scores they received for these items in the posttest according to whether they had also been exposed to the item during the treatment phase. The results are shown in Table 4.2 below. As can be seen, when the individuals in SU13 missed an item in the pretest and were not exposed to it during the treatment phase, they got the item correct only 25% of the time in the posttest. However, if they missed an item in the pretest and were exposed to it in the treatment, then they got that item right 85.48% of the time in the posttest. Similarly,

items that were missed in the pretest and not present in the treatment for participants in SP14 were correctly identified only 59.52% of the time in the posttest. However, items that were missed in the pretest and were also present in the treatment for the participants were identified correctly 81.81% of the time in the posttest. The numbers in parentheses indicate the number of different food items that were present in both the pretest and the posttest for all participants, depending on whether they were also present in the treatment. Thus, for example, in SU13 one participant missed only 12 instances of food items in the pretest that were not seen not in treatment. Two-sample t-tests were conducted in order to determine whether the differences between then scores for items that were and were not present in the treatment were significant. The results indicate that both in SU13 ($p = 0.00$) and SP14 ($p = 0.03$), the differences were statistically significant. However, due to the small number of tokens, these figures should be interpreted with care.

Table 4.2: Average posttest scores for items missed in the pretest according to treatment

	Posttest / not in Treatment	Posttest / in Treatment	Difference
Summer 2013	25% (12)	85.48% (31)	60.48%
Spring 2014	59.52% (21)	81.81% (33)	22.29%

These results show that participants performed better in the posttest for items that they missed in the pretest if they were exposed to those items during the treatment phase. This confirms that the treatment phase helped the participants acquire items they were not familiar with before beginning their participation in this study.

The fact that MuCo may be successful at helping these learners acquire the target vocabulary indicates that we can compare different factors in game design and in

participant characteristics in order to determine whether these had any influence on the effectiveness of MuCo. Moreover, this result allows us to look at the perceptions of the participants about the effectiveness of the game as a vocabulary teaching tool, in order to determine whether they recognized if MuCo had helped them learn the vocabulary.

4.2.2. Perceptual evidence

We now look at the participants' questionnaire responses in order to determine whether learners also perceived the game as being helpful. When asked how helpful it was on a Likert scale from 1 to 5, with 5 representing 'Very Helpful', the participants in SP13 on average gave MuCo a score of 4.23, while participants in SU13 gave an average score of 4.09 and participants in SP14 gave an average score of 4.16. Thus participants in each semester perceived of the game on average as being between 'Somewhat Helpful' and 'Very Helpful'. Taking a closer look, we examine the subjectively observable trends in the comments according to the helpfulness score given. A total of 43 participants gave MuCo a rating of 5 in helpfulness across all semesters. In commenting on why they had given a high rating, these participants tended to remark that MuCo provided repetitive practice and was motivating because it was fun and/or immersive. Representative comments from two participants who gave a rating of 5 include the following:

Example 11. Comments from those who awarded MuCo a rating of 5 for helpfulness

> "The repetition, as well as the visual style of the game really helped me learn these terms with little stress."
>
> "Making it into a competitive game gave me a little more motivation for learning the vocabulary, and even made it kind of fun."

These comments are interesting because they indicate two different things. First, participants intuitively agreed with SLVA researchers that frequent, shallow processing is a good way to learn these low-frequency vocabulary items. Second, participants considered the enjoyableness of the game to be an important factor in its effectiveness in learning the target vocabulary. This inability to distinguish between helpfulness and enjoyment of the game is something to which we will return throughout the rest of the study.

The most common helpfulness rating for MuCo was a 4 for 'Somewhat Helpful', which was given by 83 participants (59.71%). The comments on the helpfulness of MuCo by those who gave a helpfulness rating of 4 tended to focus either on functional defects in the game, such as the poor image quality, or that they would have preferred the game to have additional features such as English glosses, or that the game had helped them learn the target vocabulary. Three representative comments from this group include the following:

Example 12. Comments from those who awarded a rating of 4 for helpfulness

> "I would absolutely use it for vocab memorization. It makes repetition much less mind numbing. However, the pictures were so low quality it became difficult to recognize what some of them were"
>
> "…having the English word underneath the picture would have clarified what I was matching with what"
>
> "I like the concept and I learned some new words."

The comments from those who gave a helpfulness score of 4 are perhaps the most important to understand since it was the most commonly-assigned score. In general they were positive, in spite of the fact that they frequently pointed out ways in which the game

could be improved. They suggest that it is important that a digital game be extremely well designed if it is to be perceived of as more than simply 'Somewhat Helpful'.

There were nine participants who gave a neutral helpfulness rating of 3, two who gave a rating of 2, and two who gave a rating of 1 (Very Unhelpful). These comments indicated that those that had given them had either not liked the overall design of MuCo or had perceived major flaws in its construction. I use the word 'perceived' because the identified flaws were actually non-existent. For example, one participant remarked that "It [MuCo] did not give the correct answer after [I made an error] so I didn['t know how to fix it". However, this individual was a member of DGFG in SU13, and therefore was indeed provided with a way to obtain the correct answer after making an error. In order to have the correct answer identified for them they had to click on a button that appeared as soon as they made the mistake.

Finally, when looking at all of the comments on the helpfulness of MuCo, regardless of the rating given, it is interesting to note that many participants seemed to confuse the idea of learning the meaning of a word with the idea of learning the translation of a word. Learning the meaning of a word is a process by which a learner associates that word with its meaning in the target language. This process is independent of the first language. Thus, a language learner can learn a word for which they do not know the closest equivalent word in their first language. On the other hand, learning the translation of a word consists of forming a direct connection within the mind of the learner between the word in the second language and the best possible analogous counterpart of that word within the first language. Thus, in this latter process a concept must be understood in the first language before its translation can be learned in the second language. Representative of this trend is the comment from a participant who stated "I just saw the word and knew to match it to the picture instead of actually learning

it". The apparent inability of the participants to distinguish between helpfulness and enjoyment, discussed earlier in this section, as well as the inability to distinguish between meaning and translation, as just discussed, make the subjective data provided by learners at least somewhat questionable. Future research should be conducted in an attempt to see if these factors can be separated within the mind of the learner.

As a whole, the comments on MuCo's perceived helpfulness indicate not only that the quality of game design is of utmost importance if a serious game is to be perceived as very helpful, but also that this 'quality' refers to pedagogy, aesthetics, and the ability to keep the participants in a state of 'flow'. The term 'flow' is used in digital game development to describe a state in which the game is not so easy that the player gets bored and at the same time is not so hard that the player gets frustrated (Schell, 2008). This concept is similar to Vygotsky's (1980) concept of the Zone of Proximal Development, in which he states that language instruction is maximized when learners are challenged to perform beyond their ability, and are provided with the tools and assistance that they need in order to do so. But most importantly, these subjective ratings and comments indicate that on average participants did perceive MuCo as helping them acquire the vocabulary.

4.3. Classroom-style feedback vs. digital game-style feedback

4.3.1. Empirical evidence

The second research question asks whether there was a significant difference between how well the two feedback types, CF and DGF, helped participants acquire the target vocabulary. A comparison of the pretest and posttest results for these two feedback types in SU13 and SP14 is shown in Table 4.3 below. Two-tailed t-tests for equality of means were conducted in order to determine whether there was a statistically significant

difference between these two groups in either semester. The results indicated that there was statistically no difference between the average score changes from the pretest to the posttest for these two groups in SU13 ($p = 0.2$) or in SP14 ($p = 0.6$).

Table 4.3: Pretest and posttest score averages per feedback type

	Pretest	Posttest	Change
Summer 2013 CF	55.8%	81.8%	26%
Summer 2013 DGF	47.73%	83.07%	35.34%
Spring 2014 CF	50.69%	72.22%	21.53%
Spring 2014 DGF	53.53%	78.11%	24.57%

This result confirms the hypothesis that there would be no difference between the effectiveness of the two different approaches to feedback. As proposed earlier, this result may be due to the fact that both the classroom and the digital game approaches to feedback are based on research and theory within their respective fields. To confirm this finding I again looked at the posttest scores for only those items that were missed in the pretest for any given participant and then either present or absent in their treatment phase. These results can be seen in Table 4.4 below.

Table 4.4: Average posttest scores for each group for items missed in the pretest according to their presence in the treatment

	Posttest if not in Treatment	Posttest if in Treatment	Difference
Summer 2013 CF	25% (4)	82.5% (20)	57.5%
Summer 2013 DGF	25% (8)	90.9% (11)	65.9%
Spring 2014 CF	44.44% (9)	79.16% (24)	34.72%
Spring 2014 DGF	70.83% (12)	88.88% (9)	18.05%

As the results in Table 4.4 illustrate, for each group in each semester, the performance on the posttest was better when the item was present in the treatment as opposed to when the item was not present during the treatment for that particular player. Unfortunately, once again there were too few tokens to allow for reliable statistical measures to be performed on these numbers.

The above results confirm that MuCo helped each group of participants regardless of the type of feedback that they received or the semester in which they participated. However, they fail to confirm any difference in how effective the different feedback types were. Nevertheless, there are two other factors that should be kept in mind when looking at these results. First, given that the participants in this study self-selected themselves, future research should be conducted in which participation is not self-selected. It is possible that the type of students who would not self-select themselves for participation in this study would benefit more notably from one type of feedback more than those who did self-select themselves. Second, we note that in both semesters, the DGF groups improved more than their corresponding CF groups. Thus, it is possible that a pattern actually does exist and that it is simply not emerging as statistically significant due to the small number of participants. Future research should use larger numbers of participants in order to answer this question more concretely.

Since there was no difference in the actual pretest and posttest results according to feedback type, another test was conducted to see whether there was any significant difference. For this second test, the average amount of time it took each participant to answer each pretest and posttest question was averaged and compared for each different group. First, outliers were identified and removed from each group's pretest and posttest times by calculating the z-score. The times for each group were then averaged. The results from this comparison can be seen in Table 4.5 below. The purpose of this test was

to determine whether one group improved more than the other in how quickly they selected their answers in the pretest and posttest. Two-tailed t-tests for equality of means were run for each semester in order to determine whether there was a difference. The result was again negative for both SU13 ($p = 0.88$) and SP14 ($p = 0.22$).

Table 4.5: Average reaction time in seconds during the pretest and posttest per semester and feedback type

	Pretest t	Posttest t	Delta t
SU13 CF	10.38	5.86	-4.52
SU13 DGF	10.31	5.92	-4.39
SP14 CF	10.38	6.70	-3.68
SP14 DGF	10.71	8.14	-2.57

There is no statistically significant difference in the change in the average amount of time that participants used to select pretest and posttest items. This result suggests that neither feedback type helped the members of either group become more confident in their own knowledge of the tested vocabulary. This finding again confirms the hypothesis that there would be no statistically significant difference between the effectiveness of the two different feedback types. As discussed in section 2.6, the basis for this hypothesis was that both forms of feedback had been developed for their respective components, educational and entertainment, of serious games and therefore their effectiveness would be similar. The evidence thus far suggests that this is indeed the case. Both types of feedback are, statistically, equally effective at helping the participants acquire the target vocabulary.

4.3.2. Perceptual evidence

The results from these two empirical tests indicate that there was no significant difference in how well either type of feedback within MuCo helped the participants acquire the target vocabulary. The next step taken was to determine whether there was a significant difference in how helpful the participants perceived the different feedback types. In order to examine this possibility, the average scores on the 5-point Likert scales were compared, as illustrated in Table 4.6 below. A t-test for equality of means was run for both semesters in order to determine whether there was a statistically significant difference between the two groups. The difference was not found to be significant for either SU13 ($p = 0.91$) or SP14 ($p = 0.47$).

Table 4.6: Perceived helpfulness of MuCo per feedback type and semester

	Perceived Helpfulness
SU13 CF	4.11
SU13 DGF	4.07
SP14 CF	4.10
SP14 DGF	4.26

The absence of any statistically significant difference in the helpfulness ratings provided by the different groups indicates that the participants did not consider the differences in the feedback types to be related to the helpfulness of the game.

All three of the preceding comparisons between the CF and DGF groups for both semesters indicate that there is no statistical difference in the benefit of either feedback type, either empirically or perceptually. One final comparison that can be made is a subjective comparison of the comments provided by the participants in the questionnaire. Therefore I looked at all comments that referred explicitly to the feedback provided in

either the responses to the question about how helpful MuCo was, or in the responses about what the players' favorite and least favorite parts of the game were, for both groups. In sifting through the comments, I ignored those that referred to the pretest and posttest (5 comments), as well as those that stated that they wanted to be told what the correct answer was (3 comments), since this feedback was already provided. The three comments that stated that they wanted to be told what the correct answers were likely refer either to the pretest and posttest sections or indicate that the participants simply did not notice the large button in the middle of the screen that they could push in order to be provided with the correct answer.

Seven participants in the CF groups referred explicitly to feedback in their comments on the perceived helpfulness of MuCo. Each of these comments was made by players who had given a helpfulness rating of either 4 or 5. Here are three examples of the comments provided by the CF group:

Example 13. Helpfulness comments that specifically mention feedback by members of the CF groups

a. "When choosing the incorrect answer, you were shown what the correct answer was."

b. "I wasn't completely sure of the names of the food and having the game help me out and give chances was useful."

c. "The game uses pictures and corrects the player without penalizing them during the game play."

As illustrated in these responses, the members of the CF groups who referred to feedback in their comments on the helpfulness rating that they had provided tended to focus on two aspects of the feedback: (1) the correct answer was provided; and (2)

multiple attempts were allowed. This suggests that the players found the systematic use of recasts as helpful in learning the target vocabulary within the game.

In the feedback regarding helpfulness from the DGF groups, the feedback was mentioned five times. Three of these comments are similar to those from the CF group in that they refer to the trial-and-error aspect of the game. One of these is presented in (14) below, although in this case, the players had taken an extra step in order to find out what the correct answer was. This extra step of pushing a button was required because the feedback was provided on an as-requested basis, rather than consistently. This approach in providing feedback is consistent with the principles of DGF, which prescribe that feedback that may or may not be helpful should only be provided when requested. In this case, since users may in their minds have narrowed down the correct answer to one of two choices, they would not want to be shown the correct answer since they would already knew what it was merely by knowing that they had made a mistake. The presence of comments about how helpful it was that the game provided the correct answers indicates that the participants perceived the correction of their incorrect choices as helpful, regardless of whether or not the help came automatically or required them to ask for it.

Example 14. Helpfulness comments that specifically mention feedback by members of the DGF groups

a. “I learned new vocabulary through trial and error.”

b. “[W]ithout learning you have no chance to win the game[,] which essentially is what every person that plays wants to do.”

c. “I also liked that when you chose the wrong ingredient it would tell you what it was, that helped a lot.”

In addition, comment (4b) referred to the need to learn the vocabulary in order to win the game. This comment indexes the fact that the players did not earn the money for a recipe if they got it wrong, and yet still lost the ingredients that they had used up to that point in the recipe. Thus, making mistakes in this version of the game could lead to losing the game, whereas this consequence was not possible in the other version of the game. This comment suggests that increasing the stakes of getting an answer correct may be perceived of as increasing how the helpfulness of the game is perceived. The final helpfulness comment (4c) from this group refers to an additional aspect of the negative feedback, which was that when players incorrectly identified a food, they were told which food they had selected. Comment (4c) suggests that providing feedback with more than just the correct answer was also considered by at least one participant as an important part of the helpfulness of MuCo.

As illustrated in these comments, the fact that the game could provide the correct answer was seen as integral to the helpfulness of the game. Moreover, some participants also considered additional aspects of the richer feedback provided in the DGF version of the game to add to the helpfulness of the game. Since members of both groups mentioned the same aspects of the feedback as helpful, but the DGF groups also found additional aspects of the game's feedback to be helpful, it appears that the digital game approach to feedback is perceptually more helpful than classroom feedback within this context. More research is needed in order to confirm this hypothesis.

Because we previously discovered that participants did not distinguish between helpfulness and enjoyment of the game, the next question to address is whether participants in either group found the feedback as either contributing to or detracting from their enjoyment of the game. In response to the question about why they had provided a particular enjoyment rating, only one member of the CF group specifically

mentioned any aspect of the feedback. This participant, who had given the game an enjoyableness rating of 4 (Somewhat Enjoyable), stated that "The [repetition] of the questions and encouragement were sort of off putting". This comment is interesting because it suggests that providing uniform positive feedback that is not customized in accordance with the player's actions may potentially inhibit the player from being fully immersed in the game.

However, this possibility is called into question by one of the comments related to feedback made by a member of the DGFG who gave the game an enjoyment rating of 5 (Very Enjoyable). This participant stated that "The only part I didn't like was the repetition of the question '(Blank) es muy rico, no?' It was unnecessary to ask that every time". As indicated by the participant, this positive feedback was customized according to the player's actions. Therefore, because the comment suggests that the positive oral feedback is unpleasant because it is repetitive, and because all unpleasantness can distract from the game and thus impede immersion within the game, this comment suggests that the repeated feedback may have impeded immersion within the game, albeit to a small degree. It would be interesting to pursue this question further. For example, studies could look at whether positive oral feedback is perceived as more enjoyable when it is provided only after the first time each item is correctly identified, when it is reduced to a single randomly exclamatory word such as "Awesome!" or when it is omitted completely. My guess is that positive oral feedback would be perceived as most enjoyable if it is provided only sporadically; i.e. only after the first time each item is correctly identified, or when it is reduced to a single, quick, exclamatory word, which would limit the amount of time that it requires and therefore would not be as likely to be perceived as wasting the player's time.

Two other participants also mentioned aspects of the DGF groups' feedback in their explanations of why they had provided the specific enjoyableness ratings that they had provided. Both of these comments were made by participants who provided an enjoyableness rating of 4, and can be seen in (15) below.

Example 15. Comments on specific feedback elements that contributed to enjoyment of MuCo by members of the DGF group

a. "It was also interesting to see what kinds of dishes were made."

b. "The strategy of trying to get the most points was fun. Also the sound of the cooking pot blowing up made me not want to get anything wrong and slightly nervous before [I] clicked an ingredient."

Comment (15a) indicates that the images of the completed recipes made the game more enjoyable. This comment is particularly interesting given that a player in the CF group who gave a helpfulness rating of 2 stated that they had provided this low rating because "I did not understand the point of the game; why did we have to put certain ingredients into the pot; didn't really connect or make sense to me". Taken together, these two comments support the idea that language learning is perceived as both more helpful and enjoyable when the learning process involves meaningful, communicative work.

The second comment by a member of the DGFG (15b) about why the person provided an enjoyableness rating of 4 is somewhat ambiguous. This participant's comment begins by focusing on a positive aspect of the game. However, the comment also points out that the feedback element in question — the explosion provided when a mistake was made — caused them to feel nervous, which is generally considered to be a negative feeling. However, upon examining this participant's other comments, the person reported the favorite part of the game as "making recipes", which is where the explosions took place, and they also suggested that the researcher "make it [the game] more visually

interesting". Since informal conversations about MuCo have shown that the visual depiction of the explosion is almost universally regarded as being the most visually interesting element of the game, it would seem likely that this participant did actually enjoy the explosions. So why would they state that it made them nervous? To answer this question I return to one of the definitions of games discussed earlier and repeated here: "A game is a problem-solving activity, approached with a playful attitude" (Schell, 2008). People play games in order to solve problems; they want to be challenged. Thus, by making this player feel nervous, MuCo was actually succeeding as a game because it provided a challenge for the player. And it was a challenge that was meaningful and important enough to this player to affect their emotional state. I therefore surmise that this player did enjoy this feedback element.

Taken together then, the responses as to why a specific enjoyableness rating was provided that referred specifically to the type feedback subjectively suggest that feedback that has no endogenous value and takes up a noticeable amount of time, such as the positive oral responses, was not enjoyable. This type of feedback was present in both the CF as well as the DGF versions of MuCo. On the other hand, those types that intensified immersion within the game did contribute to the enjoyment of the game, and this type of feedback was only provided in the DGF version of MuCo. Thus, although both types of feedback had elements that were not enjoyed by the participants, only the DGF version also had feedback elements that were enjoyed by the participants.

One final point can be made regarding the comments about why each enjoyment rating was provided. That is, six participants in the DGF groups mentioned the aspect of competition as vital to the enjoyment of the game. Moreover, seven participants in the CF groups also mentioned the aspects of competition and challenge as vital to the game's enjoyment. This aspect is significant because the level of competition/challenge was

markedly greater for the DGF groups due to the feedback provided. In the CF groups, if one made a mistake in the cooking process, one was corrected; there was no negative consequence. On the other hand, in the DGF groups, if one made a mistake, one's chances of succeeding in the game were negatively impacted because the player's turn ended, and the player lost resources and did not earn the money that was promised at completion. Thus, it is likely that the DGF may be more enjoyable because it increases the challenge of winning the game. As one player stated: "It's all about beating the computer".

To conclude this section, we now look at the comments provided by the participants about their favorite and least favorite parts of the game that explicitly refer to feedback elements. To begin, I should point out that specific feedback elements were never mentioned as either the favorite or the least favorite part of the game by anyone in the CF groups.

The members of the DGF groups, on the other hand, did mention specific feedback elements both as their favorite and least favorite parts of the game. Eight different participants in these groups mentioned specific feedback elements, such as the explosions and earning money, as their favorite part of the game. However, we note that the explosions were also mentioned as the least favorite part of the game by two other participants in these groups. In order to determine why the explosions were listed as both the favorite and least favorite part of the game, I reviewed all of the comments provided by all the participants before the process of eliminating any participants for various reasons had begun. I discovered that two of the participants who had been excluded from the final results mentioned that they did not like the explosions because they were too loud. It was not possible to adjust the volume of specific MuCo game elements. Thus regardless of the volume at which the game was played, the explosions were always just

as loud in comparison to the other game elements. In contrast, many commercial games allow the player to adjust the volume of specific game elements individually. For example, in *World of Warcraft* it is possible to adjust the volume independently for the sound effects, the music, the speech, and the ambience. Future research should examine whether adding this volume functionality helps increase the perceived enjoyment and/or helpfulness of a serious language game. In any case, the comments on the favorite and least favorite parts of the game indicate that feedback was much more salient for members of the DGF groups than for members of the CF groups.

The next question that arises is: If the digital game feedback was subjectively more helpful, more enjoyable, and more salient, why was it not found to be empirically more helpful and enjoyable? Unfortunately, this question cannot be answered with the data collected in this study. However, I hypothesize that even though I was comparing two different versions of MuCo according to their feedback type, the participants themselves, as seen in their helpfulness and enjoyment ratings, were most likely comparing MuCo with traditional classroom activities, such as worksheets. And, even though I based the design of the feedback provided to the CF groups on principles underlying the design of feedback within the traditional language classroom, the rest of MuCo was designed as much as possible like a commercial digital game. Thus, I suggest that regardless of how the feedback was provided, the rest of the game was perceived as being much more helpful and enjoyable than more traditional language classroom activities. Any difference caused by the feedback alone was insignificant next to the difference between MuCo and traditional materials, such as worksheets. Future research could investigate this question in several different ways. Participants could be exposed to both versions of the game and then asked to compare them. A 9-point Likert scale could be used instead of a 5-point scale in order to capture possible smaller differences in

perceived enjoyment and helpfulness. A baseline comparison could be made between the CF version of MuCo and a more traditional classroom activity in order to determine how much of the variation in the ratings can be accounted for due to the other aspects of MuCo. And, finally, the study could be repeated with a larger number of subjects.

4.5. How well does MuCo motivate?

4.5.1. Average enjoyment of MuCo

The third research question seeks to determine how well MuCo succeeded at motivating participants. Once again I point out that, while the issue of the ability of digital games to motivate language learners is of great theoretical and practical importance, this research question is not central to the present dissertation. It is included herein largely anecdotally because of its importance within the field. However, future research that focuses specifically on the issue of motivation is needed in order to test all findings made and discussed herein. This question is critical because, as mentioned in chapter 2, previous theoretical work has suggested that digital games should be highly motivating for SLLs, while previous empirical work has suggested that serious language games are in fact not as motivational as was hoped (Cornillie et al., 2012a; Neville et al., 2009). The question of motivation was approached in several different ways in this study. The first approach was simply to ask participants to rate their enjoyment of MuCo on a 5-point Likert scale, with 5 as "Very Enjoyable". The average rating provided by all participants from SP13, SU13, and SP14 combined was 3.43, which is slightly lower than halfway between "Neither Enjoyable nor Unenjoyable" and "Somewhat Enjoyable". This finding would seem to confirm previous findings that digital games are not subjectively as enjoyable as has been hoped. However, if we break down the results by semester, there is an interesting result. In SP13 the average enjoyment rating was 2.76, while in SU13 it

was 3.74 and in SP14 it was 3.80. Moreover, the mode enjoyment rating for both SU13 and SP14 was 4, which means that in both these cases the most commonly given enjoyment rating was a 4, which is "Somewhat Enjoyable". The average enjoyment rating increased dramatically between the spring and summer of 2013, which is when several major alterations were made to MuCo as discussed in Chapter 3. Basically, the same occurred in SU13 and SP14, during which semesters the actual structure of the game was identical, while some of the content was changed. This finding suggests that as the game was improved, enjoyment of the game increased. This finding is not surprising, and it suggests that the game design does have a notable influence on the enjoyment of the game. All serious games are not created equal, which is important because it means that more research such as the present study should be conducted in order to determine how best to design serious games.

4.5.2. Enjoyment according to feedback type

The next step was to compare the enjoyment ratings for the CF and DGF groups in order to determine whether changing the type of feedback provided had any effect. In Table 4.7 below we can see the average rates for each group.

Table 4.7: Average enjoyment rating per feedback type and semester

	CF		DGF	
	Average	Standard Deviation	Average	Standard Deviation
Summer 2013	3.55	0.98	4	0.91
Spring 2014	3.68	0.87	4	0.79

As shown in Table 4.7, in both semesters the CF group provided an average enjoyment rating between neither enjoyable nor unpleasant and somewhat enjoyable, while the average rating for both semesters for the DGF groups was somewhat enjoyable. Moreover, the standard deviation for each group is similar, which indicates that the variation in responses is similar for each group. The difference between the different feedback groups was not significant for either SU13 ($p = 0.22$) or SP14 ($p = 0.16$) as shown by t-tests. Thus, further research with a larger number of participants is needed in order to determine whether the pattern shown in these results is significant.

4.5.3. Comments about enjoyment of MuCo

To probe more deeply into the question of the motivation inspired by MuCo in general, I examined the comments made about the enjoyment of MuCo by all groups combined. In this section I discuss the general patterns that I observed in the comments provided by the participants and share representative comments that illustrate these patterns. The danger in doing so is the temptation to 'cherry pick' comments that support my own arguments. Being aware of this temptation, I have made a concerted effort to empirically seek out the patterns in the comments, regardless of whether or not they support my arguments. In general, those who gave a rating of 5 focused their comments either on the fact that MuCo was very game-like, or on the fact that it was a good way to learn, as the following representative quotes in (16) demonstrate.

Example 16. Comments on enjoyment of MuCo by individuals who gave an enjoyment rating of "Very Enjoyable"

a. "I play games often and could not believe the amount of depth in the game. The houses and and recipes were very thought out."

b. “It was a fun way to learn the new vocabulary as opposed to memorizing it out of a book.”

Those who gave a rating of 4 tended to say that MuCo was good as a learning game, though they suggested that it could use some improvement, as these representative quotes in (17) indicate.

Example 17. Comments on enjoyment of Muco by individuals who gave an enjoyment rating of “Somewhat Enjoyable”

a. “I wouldn't call it Game of the Year, but it was a good learning game!”

b. “I had some issues with the pacing and interface, but I still enjoy pretty much all games, so I had some fun.”

The comments of those who gave a rating of 3 suggested that they had a difficult time immersing themselves into the flow of the game, as shown by the representative comments in (18).

Example 18. Comments on enjoyment of MuCo by individuals who gave an enjoyment rating of “Neither Enjoyable nor Unpleasant”

a. “It wasn't anything super fun or exciting, but it wasn't completely boring, either.”

b. “I wasn't super excited about the game, but it did its job.”

Finally, those who gave a rating of 2 or 1 mentioned problems with game elements, such as the music, but would still sometimes mention that it was good for a learning game, as shown in (19).

Example 19. Comments on enjoyment of MuCo by individuals who gave an enjoyment rating of “Somewhat Unpleasant” or “Very Unpleasant”

a. “The graphics were terrible. The music was super annoying but thankfully could be turned off.”

b. “It was tedious, but effective.”

c. “Obviously AI isn't superb, but there just isn't much challenge. It's less of a competition to see who can get 1,000 points first, rather than a grind to get there. The games were repetitive and abusable; you could click on multiple objects at once to just get to the right answer. I suggest lowering the score to something like 500 to make it seem like less of a drag.”

A common thread that runs through the comments provided at all rating levels is that this game was enjoyable. That is, the participants seemed to expect that this game, as a serious game, would not be enjoyable, and they were therefore often pleasantly surprised at how enjoyable MuCo actually was. This finding suggests that when it comes to being perceived as enjoyable, a serious game starts out at a disadvantage when compared with commercial entertainment games.

The accuracy of this finding was further confirmed by the finding that there is a significant correlation coefficient of 0.42 for the perceived helpfulness and the perceived enjoyment ($p = 0.00$). This finding indicates that those participants who perceived the game as more helpful were more likely to view the game as more enjoyable, and likewise, those who perceived the game as less helpful were more likely to see it as less enjoyable. Thus, it appears that these SLLs were not able to distinguish between how much they enjoy a serious game and how helpful it is at accomplishing the predetermined goal. Future research should be conducted to confirm this finding, and also to determine whether this effect goes both ways. That is, does making the game more enjoyable also make it seem more helpful?

4.5.4. Playing MuCo without seeking course credit

The next method employed to measure motivation elicited by MuCo involved looking at whether any participants either played the game without seeking course credit, or continued to play it after they had played the number of turns of the game that was required in order for them to receive the promised course credit. These data can be seen in Table 4.8 below.

Table 4.8: Extra participation per semester

	Played without seeking credit	Played longer than necessary
Spring 2013	25	6
Summer 2013	11	1
Spring 2014	8	6

These data show that there were 13 participants who continued to play the game after they had completed the required portion for course credit. Moreover, MuCo was played for at least 10 turns, 44 times, in non-course-credit mode. Although these numbers are small, they do suggest that it is possible that MuCo can motivate participants to take advantage of this tool even when they are not receiving any direct compensation for their participation. This result is notable since anecdotal experience suggests that SLLs will often not even complete activities that are assigned and graded as part of a language course.

Because participants who played the game without seeking course credit made that decision before they were assigned to either the CF or DGF group, there is no reason to look at the group to which they belonged. However, it is interesting to note that of the seven participants who played longer than necessary in the summer of 2013 and the spring of 2014, only two were in the CF groups while five were in the DGF groups. This

does suggest that participants in the DGF may have been more likely to continue playing MuCo after having completed the required portion of the game, which suggests a greater level of immersion.

4.6. Did gaming habits, motivation, personality traits, GPA, and/or expected course grade have any effect on helpfulness or enjoyment?

The purpose of the final research question was to find out whether variation in self-reported gaming habits, motivation, personality traits, overall GPA at the university, or the grade the students expected to make in the course would have any effect on the score changes from pretest to posttest, or the reported enjoyment or helpfulness of the game. Thus this section examines fifteen different pairings of independent and dependent variables as illustrated in Table 4.9 for both SU13 and SP14. The results from these analyses help clarify whether the usefulness of the game is different for students with different gaming habits, motivation levels, or personalities.

Table 4.9: Pairings of independent and dependent variables to be examined

		Independent Variables				
		Gaming Habits (GH)	**Motivation (M)**	**Personality (P)**	**Current GPA (CG)**	**Expected Grade (EG)**
Dependent Variables	**Score Change (SC)**	Effect of GH on SC	Effect of M on SC	Effect of P on SC	Effect of CG on SC	Effect of SC on SC
	Perceived Enjoyment (PE)	Effect of GH on PE	Effect of M on PE	Effect of P on PE	Effect of CG on PE	Effect of SC on PE
	Perceived Helpfulness (PH)	Effect of GH on PH	Effect of M on PH	Effect of P on PH	Effect of CG on PH	Effect of SC on PH

Whenever possible, statistical tests were conducted to test the significance of any observed patterns within the data. However, when one or more of the cells for any comparison contained fewer than 15 tokens, then no reliable statistical tests could be performed. Future research should be conducted with larger numbers of participants in order to test the significance of these findings.

4.6.1. The effects of gaming habits

The first of the independent factors is that of gaming habits. Participants indicated on the questionnaire how often they played several different types of digital games on average. For this questionnaire, digital games were broken into six different types: social media games (e.g. Farmville, Mafia Wars), casual computer games (e.g. Mahjong, Solitaire), hard core computer games (e.g. World of Warcraft, Skyrim), casual console games (e.g. Wii Sports, Guitar Hero), 'hard core' console games (e.g. Final Fantasy XIII,

Super Mario Brothers Galaxy), and phone app games (e.g. Angry Birds, Bejeweled). The results from this part of the questionnaire are illustrated in Table 4.10 below. As these data show, by far the most common type of game played by participants in this study is phone app games, with 37 participants playing them weekly or daily on average. In contrast, no more than 11 participants played other game types weekly or daily on average; casual computer games and casual console games were played with the least frequency, with the combined number of weekly and daily players at only four each.

Table 4.10: Gaming habits for spring 2013/ summer 2013/ spring 2014

	Never	Rarely	Monthly	Weekly	Daily
Social Media	30/ 24/ 29	12/ 3/ 8	0/ 1/ 3	2/ 2/ 1	2/ 1/ 4
Casual Computer	19/ 12/ 18	17/ 12/ 16	8/ 5/ 4	1/ 1/ 6	1/ 1/ 1
Hard Core Computer	34/ 23/ 31	5/ 2/ 2	2/ 0/ 4	4/ 3/ 5	1/ 3/ 3
Casual Console	11/ 9/ 12	18/ 15/ 18	15/ 5/ 5	1/ 2/ 7	1/ 0/ 3
Hard Core Console	21/ 13/ 23	11/ 7/ 8	10/ 5/ 5	4/ 3/ 7	0/ 3/ 2
Phone App	0/5/ 8	6/ 6/ 12	8/ 4/ 5	5/ 8/ 11	16/ 8/ 9
Any Game Type	**1/ 1/ 1**	**9/ 5/ 10**	**5/ 3/ 6**	**17/ 9/ 14**	**14/ 13/ 14**

As shown in Table 4.10, it was not possible to perform a statistical test on the data divided in this way because many of the cells had fewer than 15 participants. To overcome this obstacle, I decided to sacrifice detail in favor of statistical power. Therefore all of the games were collapsed into a single category, identified as "Any Game Type" in Table 4.11, and all frequencies were collapsed into two different ratings: "Infrequently", comprising Never, Rarely, and Monthly, and "Frequently", including Weekly and Daily. Combining the data in this manner produced the data shown in Table 4.11 below.

Table 4.11: Aggregated gaming habits for spring 2013/ summer 2013/ spring 2014

	Infrequently	Frequently
Any Game Type	15/ 9/ 17	31/ 22/ 28

As shown in Table 4.11, even after collapsing all of these data, there is still one cell, including participants in SU13 who play games infrequently, that has fewer than the minimum of 15 participants required for the results to be statistically valid. Thus only the data from SP14 is examined in the remainder of this section. A t-test was performed in order to determine whether gaming habits had a statistically significant effect on the score change from pretest to posttest, shown in Table 4.12.

Table 4.12: Influence of gaming habits on score change

	Pretest Score	Posttest Score	Difference
Non-Gamers	58.90%	77.27%	18.36%
Gamers	44.87%	68.48%	23.61%

The results from the t-test indicate that there was no significant difference in how much gamers and non-gamers improved on average from the pretest to the posttest ($p = 0.45$). This finding means that neither gamers nor non-gamers improved significantly more than the other group from the pretest to the posttest in their mastery of the target vocabulary. It suggests that the utility of serious language games is not dependent on the pre-existing gaming habits of the students, which is exactly what we had hypothesized to be the case.

The next two comparisons look at whether gaming habits had an effect on the perceived enjoyment or the perceived helpfulness of MuCo in SP14. To answer this question, the results from Gamers and Non-Gamers are compared in Table 4.13.

Table 4.13: Influence of gaming habits on perception of enjoyment and helpfulness

	Helpfulness	Enjoyment
Non-Gamers	4.00	3.58
Gamers	4.14	3.82

T-tests on the difference between the perceived helpfulness and enjoyment of MuCo for gamers and non-gamers indicate that there is no significant difference for either perceived helpfulness ($p = 0.60$) or perceived enjoyment ($p = 0.39$). These data suggest that a student's perception of the helpfulness and pleasure of a serious game is not influenced by their gaming habits. Once again, the lack of a distinction based on gaming habits can be interpreted as an indication that whether our students are gamers when we consider using serious games in class is probably of little importance.

4.6.2. The effects of motivation

Next, we look at the effect of self-reported motivation on score changes, perceived helpfulness and perceived enjoyment. The data on motivation were collected from participants' answers to 11 questions (see Appendix A) derived from the online version of the *French Mini AMTB* (see Appendix B). Each question required the participant to provide an answer on a 7-point Likert scale. The same procedure was followed to calculate each participant's motivation score, as suggested by Tennant and Gardner (2004). This procedure involves first obtaining the total of the Likert scale numbers for questions that represent positive motivation such as "My attitude toward

learning Spanish is:", then subtracting the numbers chosen by the participant for questions that indicated negative motivation, such as "My anxiety level in Spanish is:", and finally dividing the total by the number of questions. See Table 4.14 below for a depiction of the motivation scores for SU13 and spring of 2014.

Table 4.14: Motivation scores for summer 2013 and spring 2014

	Average Motivation Score	Motivation Score Standard Deviation
Summer 2013	3.55	0.75
Spring 2014	3.72	0.64

In order to determine whether there is a significant relationship between motivation and score change, a linear regression was performed for each semester. The results from the regression confirm that there is no significant relationship between these two variables in either SU13 ($p = 0.13$) or SP14 ($p = 0.36$). The lack of any significant correlation between the participants' motivation and score change is interesting given the fact that motivation has been cited as a very important factor in SLA (Cook, 2008). I believe that this result is due, at least in part, to the fact that the participants self-selected themselves for participation in this study, which resulted in a small range of motivation scores, as indicated by standard deviations shown in Table 4.14 (0.75 and 0.64).

Similarly, linear regressions were performed in order to determine whether there was a relationship between motivation and perceived enjoyment or between motivation and perceived helpfulness for either SU13 or SP14, as seen in Table 4.13. The results from these regressions indicate that in SU13, the motivation scores had no significant effect on either perceived enjoyableness ($p = 0.26$) or on perceived helpfulness ($p = 0.34$). There was also no significant correlation for SP14 between motivation and

perceived helpfulness ($p = 0.29$) or perceived enjoyment ($p = 0.16$). Taken together, these results suggest that the ability of MuCo to help participants learn the target vocabulary and to motivate them is not affected by their degree of self-motivation. This finding is important because it suggests that the same game can be used for the same learners regardless of their initial motivation. Nevertheless, this conclusion must be tested further because all of the participants in this study were self-selected, which may have resulted in an abnormal distribution of motivation scores.

4.6.3. The effects of personality

The next set of data is used to determine if there is any significant correlation between each participant's personality and their score changes, perception of the helpfulness of the game, or their enjoyment of the game. Personality was measured using Saucier's (1994) 40 Mini Markers. Each Mini Marker is an adjective that is strongly correlated with a specific personality trait. There are 8 adjectives for each of the 5 different personality traits. Personality scores were collected by asking participants to indicate how well each of Saucier's 40 Mini Markers describes them on a 9-point Likert scale. One example of a question for each personality trait is displayed in (10) below.

Example 20. One example of a question for each of the five personality traits measured by Saucier's (1994) Mini Markers

a. Extraversion: Describe yourself on a scale from 1 (Extremely Inaccurate) to 9 (Extremely Accurate) for the word "Bashful."

b. Agreeableness: Describe yourself on a scale from 1 (Extremely Inaccurate) to 9 (Extremely Accurate) for the word "Cold."

c. Conscientiousness: Describe yourself on a scale from 1 (Extremely Inaccurate) to 9 (Extremely Accurate) for the word "Disorganized."

d. Emotional Stability: Describe yourself on a scale from 1 (Extremely Inaccurate) to 9 (Extremely Accurate) for the word "Anxious."

e. Intellect, Openess, or Imagination: Describe yourself on a scale from 1 (Extremely Inaccurate) to 9 (Extremely Accurate) for the word "Creative."

The full set of questions can be seen in Appendix A. There are eight Mini Markers for each personality trait tested. Some Mini Markers are positive while others are negative. Personality scores were calculated by adding together all positive markers for each personality trait, subtracting the negative markers for each trait, and finally dividing the resulting number by 8. The five personality traits tested by Saucier's Mini Markers are: Extraversion, Agreeableness, Conscientiousness, Emotional Stability, and Intellect/Openness/Imagination. This final personality trait may seem somewhat confusing since the three terms used are not synonyms. However, they are grouped together by Saucier due to the finding that the individual Mini Markers that correlate with each of the three different terms also strongly correlate with each other. In order to determine whether personality had any effect on score change, perceived enjoyment, or perceived helpfulness, three separate multiple regressions were performed for both semesters, as seen in Table 4.15 below.

Table 4.15: Significance of effect of personality on score changes, perceived enjoyment and perceived helpfulness

	Extraversion	**Agreeableness**	**Conscientiousness**	**Emotional Stability**	**Intellect or Openness**
Summer 2013 Score Changes	p = 0.35	p = 0.60	p = 0.88	p = 0.57	p = 0.04
Summer 2013 Perceived Enjoyment	p = 0.56	p = 0.34	p = 0.84	p = 0.59	p = 0.78
Summer 2013 Perceived Helpfulness	p = 0.11	p = 0.09	p = 0.37	p = 0.58	p = 0.18
Spring 2014 Score Changes	p = 0.50	p = 0.14	p = 0.33	p = 0.49	p = 0.50
Spring 2014 Perceived Enjoyment	p = 0.25	p = 0.02	p = 0.25	p = 0.83	p = 0.77
Spring 2014 Perceived Helpfulness	p = 0.82	p = 0.65	p = 0.89	p = 0.16	p = 0.82

The only two relationships found in these regressions to be statistically significant in Table 4.15 are those of "Intellect or Openness" on score changes in SU13 and of "Agreeableness" on the perception of enjoyment in SP14. However, there are other data that suggest that these two figures are not important, the most notable of which are that neither of these findings is reflected in the data collected for the other semester. That is, the effect of "Intellect or Openness" on score changes in SP14 is not significant (p = 0.50), nor is the effect of "Agreeableness" on perceived enjoyment significant in the summer of 2014 (p = 0.34). Additionally, the r-squared values of the regressions for the

effect of personality traits on the score changes in SU13 and the effect of personality traits on the enjoyment in SP14 are 0.18 and 0.22, respectively. These low r-squared figures indicate that the magnitude of effect for these models is negligible. Regarding how much vocabulary was learned by playing the game, how much the participants perceived that they had enjoyed the game, and how helpful they perceived it, these results from the effects of personality traits indicate that MuCo was equally effective regardless of individual differences. The results mirror those data that looked for any potential effect for gaming habits or motivation.

4.6.4. The effects of GPA and expected course grade

The final two variables we examined to see whether they have an effect on the score change from the pretest to the posttest, on the perceived helpfulness, or on the perceived enjoyment were the participant's current GPA and the grade that they expected to make in their current Spanish course. The data for these variables were collected only in SP14. The average self-reported GPA for participants was 3.27, with a standard deviation of 0.47. The average self-reported expected grade for the current Spanish course was an 87.22, with a standard deviation of 5.93. Linear regressions were used in order to determine whether these factors had a significant effect on the independent variables, as shown in Table 4.16 below.

Table 4.16: Significance of effect of GPA and expected course grade on score changes, perceived enjoyment and perceived helpfulness

	GPA	**Expected Course Grade**
Score Changes	p = 0.77	p = 0.19
Perceived Enjoyment	p = 0.71	p = 0.75
Perceived Helpfulness	p = 0.71	p = 0.40

The findings illustrated in Table 4.16 reveal that neither the participants' GPA, nor the grade they expected to earn in their Spanish course that semester, had any significant effect on their improvement from the pretest to the posttest, their perceived enjoyment of the game, or their perception of the helpfulness of the game. Like the other results in this section, this finding indicates that the differences among these participants had no effect on the effectiveness, whether measured or perceived, or the perceived enjoyment, of MuCo. Given that MuCo was measured and perceived to be helpful and enjoyable on average, this result once again indicates that MuCo was both helpful and enjoyable for these participants, regardless of their individual differences.

4.7. Summary and conclusion

In this chapter, I have presented and discussed the results from the research carried out in order to answer my research questions. This section is a summary of these findings. The first research question asked whether MuCo was effective at helping participants improve in their command of the target vocabulary as measured empirically by the pretest and posttest within the game, and subjectively by the ratings provided in the questionnaire. The results indicate that MuCo was indeed successful both empirically and subjectively at helping the participants improve in their command of the target vocabulary. This promising result suggests that more research should be conducted on the use of serious games in the formal language learning environment.

The next research question asked whether there was a difference in how effective the two different types of feedback were, either perceptually or empirically. This question is important because the answer will aid in the design of future serious language-learning games. The results indicated that there was no statistically significant difference in how well either feedback group improved in their pretest to posttest scores, in how quickly

they answered the test items, nor in how helpful they rated MuCo to be on a Likert scale. However, a close comparison of the comments provided by both groups revealed that the CF group did not mention anything specific to their group as being helpful or enjoyable, while several different aspects of the DGF were mentioned as either being helpful, enjoyable, and not enjoyable. This finding suggests that the DGF is more salient, and perceptually more helpful, than the CF. The increased salience is most commonly referenced in the comments with relating to the visual and audio effects of the explosion, which suggests that spending the extra effort to make the feedback as juicy as possible, as specified in the DGF guidelines listed in section 1.2, does in fact aid with the immersion of the participants within the game environment. The comments on the helpfulness of the DGF referred to the option to have the correct option provided and the benefit of having the game provide the name of the selected food when an error was made. Thus, participants found added feedback to be helpful, but they noticed it more when it was optional. This again brings us back to the DGF principle of providing feedback on an as-needed basis.

The purpose of the third research question was to determine whether participants enjoyed and/or were motivated by MuCo. In order to answer this question we looked at the ratings provided by the participants on how much they liked MuCo, the comments provided by the participants, as well as whether participants played the game without receiving credit for it. The data indicated that participants enjoyed MuCo according to both their ratings and comments, and that they were motived by MuCo as evidenced by the use of the game without the opportunity to earn course credit. This finding is important in light of previous findings on enjoyment of and motivation by serious language games, neither of which have had unambiguously positive responses. This

finding suggests that MuCo may be a step in the right direction as far as the design of serious language games is concerned.

The objective of the final research question was to determine whether individual differences among participants could account for any of the variation in measures of success, perceived helpfulness, or perceived enjoyment of MuCo. Therefore, we looked at the effect exercised by five different independent variables on three different dependent variables. The independent variables examined were the gaming habits, motivation, personality traits, GPA, and expected course grade of the participants. The dependent variables included improvement from pretest to posttest, perceived helpfulness, and perceived enjoyment. Because the difference between CF and DGF had not had a significant effect on these same dependent variables, no distinction was made between the CF and DGF participants in these analyses. The results indicated that none of these independent variables had a significant effect on any of the dependent variables. This finding suggests that the effectiveness of serious games does not depend on any of the independent factors examined, which in turn suggests that serious games may be equally effective and enjoyable for all SLLs, regardless of individual differences.

Chapter 5: Discussion and Conclusion

5.1. Introduction

This last chapter examines how the data help answer the research questions listed in 2.4, and proposes future lines of research based on the findings obtained herein. The research questions are repeated below for the sake of convenience for the reader:

(1) Does MuCo help participants learn the target vocabulary?

(2) If so, does the MuCo game with traditional classroom feedback (CF) or the MuCo game with feedback designed on accepted video game design principles (DGF) do a better job of helping participants learn the target vocabulary?

(3) How well does MuCo motivate participants, as evidenced by the participants' self-reported enjoyment of the game and the extent of their use of the game?

(4) Can a participant's improvement from pretest to posttest, perceived helpfulness of the game, and/or reported enjoyment vary according to their self-reported gaming habits, motivation, and/or personality traits either collectively or according to feedback type?

5.2. Does MuCo work in helping learners with L2 vocabulary?

The results from this study would not be valid if MuCo were not successful at helping participants acquire the target vocabulary. Simply put, the relative effectiveness of different conditions cannot be compared if neither condition is effective. Fortunately, the fact that participants' posttest scores on average were 28.14% more accurate than their pretest scores indicate that MuCo was successful at helping participants learn the target vocabulary. This indication thus allows us to compare the different versions of

MuCo as well as the different participant variables in order to see whether any of them had any significant effect on MuCo's success.

Nevertheless there is also another important tangential finding that can be derived from the confirmed effectiveness of MuCo, especially in light of the fact that participants played MuCo outside of class. That is, the fact that MuCo helped participants acquire the target vocabulary outside of the classroom suggests that tools such as MuCo could be used in a 'flipped classroom' environment. A flipped classroom is a one in which students are initially exposed to new course material outside of the classroom and then practice with the content during class. In such an environment MuCo could serve initially as a tool to teach the new content to learners before they practice it in the classroom itself. Thus instructors could use activities such as MuCo to help students learn their vocabulary at home, and then they could dedicate class time to practicing that vocabulary, instead of having to teach it. This would allow instructors to use class time more effectively because the learners would already be prepared with a basic familiarity of the vocabulary to be used that day.

5.3. Classroom-style feedback vs. digital game-style feedback

One of the features that distinguish this study is that I compared the relative effectiveness of the principles behind classroom-style feedback with those behind digital game-style feedback. This is important because it helps us to get a better idea of how serious digital games should be designed in order to be as useful as possible. To accomplish this, I created two versions of MuCo. The version based on classroom-style feedback provided feedback in the form of recasts when the participant made a mistake, and provided a brief acknowledgement of success when they correctly identified a food. In the digital game-style version, the game provided negative oral and visual cues as well

as resource depletion when the player made a mistake, and the opportunity to obtain just-in-time information. When participants in this group got the answer correct, they were provided with positive oral and visual cues. Thus equal emphasis was placed on both the positive and negative versions of the feedback.

As discussed previously, the results from this study failed to identify any difference in effectiveness of either feedback style in helping participants acquire the target vocabulary. Before beginning this study, I hypothesized that this might be the result due to the fact that I was comparing the accepted design principles of digital game feedback with the accepted norms of classroom feedback within an environment that was a hybrid of both of these environments. The comments provided by participants suggest that this may indeed have been the case since participants who received classroom-style feedback referred to it as only a pedagogical tool, while those in the digital game-style feedback group referred to it as only a game element. Future research is needed in order to determine whether this is in fact true or not.

Another interesting result found in the participants' comments was that many of the participants found the competition with the computer players to be the most engaging aspect of the game. Since the practical integration of the feedback in the DGF version of MuCo made that version much more competitive than the CF version it is probable that participants who enjoy competition would enjoy the DGF version more than the CF version. Further research is needed in order to determine whether the use of feedback to increase the level of competition increases overall participant enjoyment of serious games.

Also it should be kept in mind that I designed both versions of MuCo based on digital game design principles in all other regards. Therefore this study actually compared a game that was completely based on digital game design principles with one that was

based almost completely on those same principles. It is quite possible that looking at such a small piece of the pie has a negligible effect, even with the number of participants used in this study. Future studies that compare CF and DGF should be carried out with larger numbers of participants in order to determine whether an effect for feedback type emerges. Additionally, future research should be conducted that compares DGF with pure CF, rather than the hybrid of CF and DGF that was used in the present study. Future studies should also be carried out that compare the relative effectiveness of other game components when they are based on either classroom practices or digital game design principles. There are many such possible game components that need to be examined. These include, but are not limited to: (1) the story, in that digital games tend to be based on fantastic storylines while classroom activities tend to be designed to be as realistic as possible; (2) skill testing, because digital games skills are tested in the same environment in which they are learned while classroom skills tend to be tested artificially with tests that tend to be distinct from the activities employed in the learning process; (3) continuity, because in digital games each activity is very strongly related with each other activity through the storyline and the mechanics behind the activities while classrooms activities tend to have disparate stories and mechanics; and (4) competition, in that in digital games, competition between players is common even when participants are working together, whereas in classrooms, competition is rare.

One final point that should be made is that the design principles discussed should also be examined in environments other than digital games. What I am suggesting is that the perceived helpfulness of digital games may be due simply to these differences in design principles rather than to the fact that they are digital. That is, if it is demonstrated that design principles have a significant effect on the effectiveness of pedagogical activities, then it could be that digitizing these activities is not necessary. Instead, these

principles could simply be employed through 'gamification' in order to improve regular classroom design. Gamification is the process of applying the principles that make games enjoyable to the classroom in an attempt to make the classroom more enjoyable and engaging. However, even if this is the case, research should still be carried out in order to determine whether these principles are not more effective and/or easily implemented in a digital environment. I propose that since digital tools are so effective at providing uniformity and consistency, and since digital games can be used *ad infinitum* after they are created, it is quite probable that even if digital game design principles can be used to improve regular classroom activities, they will be even more effective when used in digital games.

5.4. Motivation

Perhaps the most frequently touted potential benefit of serious games is the fact that they should be highly motivational for learners. Previous research has not found this to be the case, and has even found that study participants have rated enjoyment of digital games as lower than that of worksheets (Neville et al., 2009). In this study, this was not the case. Although enjoyment ratings were still not stellar with MuCo, they were better than neutral. It is possible that the group with the feedback based on game design principles enjoyed the game more than the other group, although this hypothesis needs to be tested with a larger number of participants in order to determine whether the difference is significant.

There are several possible explanations behind the fact that MuCo was perceived as more enjoyable than digital games, as found in previous studies such as Neville et al. (2009). One possibility is the game genre. While all previous serious language learning games have been designed as role-playing games, MuCo was created as a strategy game.

Due to this genre difference it is very likely that exposure to content was much faster paced in MuCo than in the games employed in previous studies. And participants' comments in previous pilot studies have shown that when playing serious games, participants do not like to waste time on purely game-related elements, such as moving around the game world, which is prevalent in most role-playing games. It is also possible that participants were more likely to be familiar with the strategy genre game since these games can be found on phones, consoles, and computers, while role-playing games are not generally found on phones, the platform most frequently used by my participants to play digital games for leisure.

Another possibility is the relative amount of effort put into developing MuCo as a digital game. As pointed out earlier, I worked diligently to follow accepted game design procedures and principles in the development of MuCo. Although previous studies have not revealed exactly how much effort was put into making their games game-like, the very fact of this omission suggests that the effort was not as great as it could have been. This difference in effort is important because, as indicated earlier, just creating a digital game does not make it enjoyable; there have been thousands of failed attempts at creating entertaining digital games.

Finally, another important factor that may have affected the relative enjoyment of these serious games is the manner in which they were tested. All previous studies that I have examined have tested the enjoyment of digital games in a laboratory setting. This means that the participants were playing the game completely independently of their actual language class. Since participants were likely not studying the target language on their own time, any teaching tool would be perceived as just that: a teaching tool. And this would be a teaching tool for use during leisure time. Outside of class it is likely that participants are operating with the frame of activities for pure enjoyment and are

therefore biased against enjoying any activity that appears to be educational. On the other hand, in this study the game was tested within the classroom environment and participants were given in-class credit in exchange for their participation. Thus when the participants were playing the game, they were doing so during study time. Instead of looking at the game through the frame of entertainment, they were likely viewing it through the frame of education and study. This means that they were probably classifying their enjoyment of MuCo along with their enjoyment of other classwork, rather than with their enjoyment of other forms of entertainment. Since they were approaching MuCo from such a different perspective, it seems likely that they would be more generous in their judgment of how enjoyable it was.

One of the important findings of this study that supports this last possibility is the fact that the participants in this study had a significant positive correlation between their ratings for perceived helpfulness and perceived enjoyment. Thus how enjoyable the game was perceived as being was somehow connected with how helpful the game was viewed. It is likely that such a correlation does not exist with commercial digital games. Hence individuals who tend to play digital games for the reasons listed in chapter 1, only one of which has anything related to learning, are playing serious games for a completely different reason; that needs to be considered when examining enjoyment of serious games. As this discussion suggests, it is difficult to get someone to change their frame of perspective and one to look at a pedagogical tool as a form of entertainment. Nevertheless, several of the comments provided by the participants in this study suggest that this is indeed a possibility, at least to a limited extent. There were several participants whose comments clearly indicated that they had been able to 'get into flow' with MuCo. And in fact, several participants made suggestions about how MuCo could have been even more enjoyable. These suggestions included making the game capable for

multiplayer format in order to facilitate interpersonal communication and competition, and improving game elements, such as the music, graphics, and artificial intelligence.

5.5. Conclusion

The findings from this study are very important for the field of game-based language learning for several reasons. These reasons were originally presented in 2.4 and are summarized below for convenience. After reviewing what the reasons are, I highlight how the present study has helped to address each point and discuss directions for future research.

The reasons for such a study as the present one are:

(1) There is a gap in the literature regarding about how to design feedback for serious games.

(2) The merits of serious games need to be further tested.

(3) Empirical data is needed to guide the development of online SLVA materials.

(4) The question as to whether serious games can motivate and benefit adult SLLs remains unanswered.

(5) Empirical research is needed to test the merits of different design principles.

(6) SLL educators need to be informed so that they can make informed choices when selecting between different available SLVA materials.

The gap in the literature about how best to design feedback for serious language games is important because feedback is such an important part of both second language instruction and digital game design. In this study, I compared the relative effectiveness of classroom-style feedback with digital game-style feedback within a single serious game.

However, it should be kept in mind that, because this game was designed to be as game-like as possible, there were certain elements of the feedback that were based on game design principles for both groups. These include the fact that successful completion of a recipe resulted in an increase in the number of *pesos* owned by the player. Nevertheless, the findings from this study suggest that designing feedback for digital games based on accepted digital game design principles may be more helpful as far as getting the player into the flow of the game and increasing enjoyment. These findings need to be confirmed by further investigation.

This study successfully tested the merits of serious digital games. The results confirm that digital games do indeed have promise as pedagogical tools. In fact, since the participants played MuCo as a learning tool, and not as a review tool, outside of the classroom it is quite probable that serious digital games would be a powerful asset in flipped classrooms, allowing instructors to dedicate more class time to practicing what students have already learned at home.

The empirical data collected during the course of this study confirm the fact that frequent, shallow processing is effective for low-frequency vocabulary items. Moreover, the results also indicate that participants seemed to be aware of this fact since several of them commented on their perception of MuCo as useful because it allowed for frequent, shallow processing of these items. Also, learners appreciated the fact that MuCo was consistent with the feedback it provided. These empirical findings suggest that online SLVA materials should focus on the frequent, shallow processing of low-frequency vocabulary items and that they should be particularly successful at aiding SLVA because they can provide very consistent feedback.

The important question of how well digital games can motivate adult SLLs has also been clarified by the present study. The findings in this study suggest that adult SLLs

can indeed be motivated by serious digital games: they can enjoy these games, find them useful, and even use them more than they are required to use them. When viewed in light of the failure of previous serious language games to motivate adult learners in similar ways, these findings suggest that although digital games can be motivational, they must be well designed, and based on both solid game design principles and pedagogical principles in order to achieve this motivation.

Although the differences in effectiveness of the different game design principles employed failed to provide empirically significant results, they do suggest that larger studies may find such differences. Moreover, they suggest that digital game design principles may be more effective than classic classroom activity development and execution principles.

Finally, the present study provides information that is useful for educators who must select pedagogical tools. The results suggest that educators should find activities that neither compromise their pedagogical merit nor sacrifice how enjoyable they are. Because both pedagogical the merit of serious games as well as the ease with which they engage learners are crucial to their success as effective language teaching tools.

In conclusion, this study has effectively answered all of the research questions and addressed all of the reasons indicated that confirm the need for such a study. The key finding of the present study is that feedback within the context of serious language learning games is equally effective at teaching Spanish vocabulary both when designed with the principles behind classroom feedback and with those behind digital game feedback, but that the feedback leads to greater immersion within the game when designed using the principles behind digital game feedback. Nevertheless, the present study has only barely scratched the surface of the research that needs to be continued in the field of serious language games. Future studies should examine such variables as the

level of proficiency of the participants, compare methodologies for high-frequency and low-frequency items, compare different types of tasks and game genres, and of course look at the relative importance of how different game elements are designed. There are many such elements, though perhaps some of the most important to look at are the story, the continuity, competition, and skill testing.

In the next iteration of this study, I plan to increase the complexity by introducing the element of multiplayer competition within MuCo. There are three primary reasons for pursuing this line of research. First, modifying MuCo to become a multiplayer environment and giving the players the ability to interact with text and/or audio chat features will provide a new source of feedback for the learners; namely their peers. This means that the feedback obtained by the participants will be more naturalistic, which in turn means that the findings from the research will be more applicable for practical language pedagogy. The second reason to make this modification is that changing MuCo into a multiplayer game will increase the challenge of the game, and will thus likely result in increased immersion on the part of the participants. Because immersion is linked to acquisition of language, I hypothesize that MuCo as a multiplayer game will increase its effectiveness as a pedagogical tool as well as increase the level of motivation and entertainment that it provides. Finally, although some researchers hypothesize about the possible advantages of multiplayer digital games as pedagogical tools (see for example Thorne et al. 2012), so far, to my knowledge, no multiplayer serious language games have been tested and proven to provide them. Given the great interest in serious gaming (see for example Johnson et al. 2012), the gap in the literature related to the testing of multiplayer serious language games needs to be addressed as soon as possible.

Appendix A

Game Questionnaire
Submit this questionnaire after playing the game in order to get quiz credit!
What is your name?
When does your class officially start work on Chapter 10?
What is your instructor's email address?
1. What external resources (e.g. friends, textbook, dictionary, etc.) did you use during the game and how extensively did you use them (e.g. the whole time, just once or twice, etc.)?
2. How helpful do you think the game was at helping you learn the target vocabulary? Very helpful <---> Very unhelpful
3. Why did you give the rating you chose in question 2?
4. How enjoyable do you think the game was? Very enjoyable <---> Very unpleasant
5. Why did you give the rating you chose in question 4?
6. Please indicate how often you play each type of game.
Social media games (Games on social sites such as Facebook; e.g. Mafia Wars, Farmville). On average at least once a day <---> Never
Casual computer games (Games that take less than one hour to play; e.g. Solitaire, Mahjong.) On average at least once a day <---> Never
Serious computer games (Games that take more than one hour to play; e.g. World of Warcraft, Skyrim.) On average at least once a day <---> Never
Casual console games (Games on a console such as the Wii or PSP that take less than one hour to play; e.g. Wii Sports, Guitar Hero.) On average at least once a day <---> Never
Serious console games (Games on a console such as the Wii or PSP that take more than one hour to play; e.g. Final Fantasy XIII, Super Mario Brothers Galaxy.) On average at least once a day <---> Never
Phone apps (e.g. Angry Birds, Plants vs. Zombies.) On average at least once a day <---> Never
7. Please use the list of common human traits to describe yourself as accurately as possible. Describe yourself as you see yourself at the present time, not as you wish to be in the future. Describe yourself as you are generally or typically, as compared with other persons you know of the same sex and of roughly the same age.
Bashful: Extremely Inaccurate <---> Extremely Accurate
Bold: Extremely Inaccurate <---> Extremely Accurate
Careless: Extremely Inaccurate <---> Extremely Accurate
Cold: Extremely Inaccurate <---> Extremely Accurate

Complex: Extremely Inaccurate <---> Extremely Accurate
Cooperative: Extremely Inaccurate <---> Extremely Accurate
Creative: Extremely Inaccurate <---> Extremely Accurate
Deep: Extremely Inaccurate <---> Extremely Accurate
Disorganized: Extremely Inaccurate <---> Extremely Accurate
Efficient: Extremely Inaccurate <---> Extremely Accurate
Energetic: Extremely Inaccurate <---> Extremely Accurate
Envious: Extremely Inaccurate <---> Extremely Accurate
Extroverted: Extremely Inaccurate <---> Extremely Accurate
Fretful: Extremely Inaccurate <---> Extremely Accurate
Harsh: Extremely Inaccurate <---> Extremely Accurate
Imaginative: Extremely Inaccurate <---> Extremely Accurate
Inefficient: Extremely Inaccurate <---> Extremely Accurate
Intellectual: Extremely Inaccurate <---> Extremely Accurate
Jealous: Extremely Inaccurate <---> Extremely Accurate
Kind: Extremely Inaccurate <---> Extremely Accurate
Moody: Extremely Inaccurate <---> Extremely Accurate
Organized: Extremely Inaccurate <---> Extremely Accurate
Philosophical: Extremely Inaccurate <---> Extremely Accurate
Practical: Extremely Inaccurate <---> Extremely Accurate
Quiet: Extremely Inaccurate <---> Extremely Accurate
Relaxed: Extremely Inaccurate <---> Extremely Accurate
Rude: Extremely Inaccurate <---> Extremely Accurate
Shy: Extremely Inaccurate <---> Extremely Accurate
Sloppy: Extremely Inaccurate <---> Extremely Accurate
Sympathetic: Extremely Inaccurate <---> Extremely Accurate
Systematic: Extremely Inaccurate <---> Extremely Accurate
Talkative: Extremely Inaccurate <---> Extremely Accurate
Temperamental: Extremely Inaccurate <---> Extremely Accurate
Touchy: Extremely Inaccurate <---> Extremely Accurate
Uncreative: Extremely Inaccurate <---> Extremely Accurate
Unenvious: Extremely Inaccurate <---> Extremely Accurate
Unintellectual: Extremely Inaccurate <---> Extremely Accurate
Unsympathetic: Extremely Inaccurate <---> Extremely Accurate
Warm: Extremely Inaccurate <---> Extremely Accurate
Withdrawn: Extremely Inaccurate <---> Extremely Accurate
8. Complete each sentence truthfully. Remember that your instructor will not see your answers.
My feelings about learning Spanish in order to interact with Spanish speakers are: Weak <---> Strong
My attitude toward Spanish speakers is: Unfavorable <---> Favorable

My interest in languages other than Spanish and English is: Very low <---> Very high
My desire to learn Spanish is: Weak <---> Strong
My attitude toward learning Spanish is: Unfavorable <---> Favorable
My attitude toward my Spanish teacher is: Unfavorable <---> Favorable
My feelings toward learning Spanish for practical purposes, such as to improve my occupational opportunities, are: Weak <---> Strong
My anxiety in speaking Spanish outside of class is: Very low <---> Very high
My attitude toward my Spanish class is: Unfavorable <---> Favorable
My anxiety level in Spanish class is: Very low <---> Very high
I would characterize how hard I work at learning Spanish as: Very little <---> Very much
8. For how much of the game (percentage) did you listen to the sound and music?
9. What was your favorite part of the game?
10. What was your least favorite part of the game?
11. Are there any other comments (praise, suggestions, critiques, complaints, etc.) that you would like to make about the game and/or the study?

Appendix B

Survey Items in the Mini-AMTB Online Questionnaire (7-point scales)	
Item	Content
IO (Integrative Orientation)	If I were to rate my feelings about learning French in order to interact with French Canadians, I would have to say they are: Weak <---> Strong
AFC (Attitude toward French Canadians)	My attitude toward French Canadians is: Unfavorable <---> Favorable
IFL (Interest in Foreign Languages)	My interest in languages other than French and English is: Very Low <---> Very High
D (Desire to Learn French)	My desire to learn French is: Weak <---> Strong
ALF (Attitude toward Learning French)	My attitude toward learning French is: Unfavorable <---> Favorable
PROF (Attitude toward French Instructor)	My attitude toward my French professor is: Unfavorable <---> Favorable
INST (Instrumental orientation)	If I were to rate my feelings about learning French for practical purposes such as to improve my occupational opportunities, I would have to say they are: Weak <---> Strong
FUA (French Use Anxiety)	My anxiety in speaking French outside of class is: Very Low <---> Very High
COURSE (Attitude toward French Course)	My attitude toward my French class is: Very Low <---> Very High
FCA (French Course Anxiety)	My anxiety level in my French classes is: Very Low <---> Very High
MI (Motivational Intensity)	I would characterize how hard I work at learning French as: Very Little <---> Very Much

Appendix C

Abbreviation	Expansion
AMTB	Attitude Motivation Test Battery
CALL	Computer Assisted Language Learning
CF	Classroom Feedback
CFG	Classroom Feedback Group
CLL	Communicative Language Learning
DGF	Digital Game Feedback
DGFG	Digital Game Feedback Group
GPA	Grade Point Average
IH	Interaction Hypothesis
IRB	Internal Review Board
L2	Second Language
NES	Nintendo Entertainment System
SLA	Second Language Acquisition
SLL	Second Language Learner
SLVA	Second Language Vocabulary Acquisition
SP13	Spring of 2013
SP14	Spring of 2014
SU13	Summer of 2013
TBLT	Task-Based Language Teaching
TL	Target Language
UNM	The University of New Mexico at Albuquerque
UT	The University of Texas at Austin

References

Allwood, J., Nivre, J., & Ahlsén, E. (1992). On the Semantics and Pragmatics of Linguistic Feedback. *Journal of Semantics, 9*(1), 1-26.

Arnseth, H. C. (2006). Learning to Play or Playing to Learn -- A Critical Account of the Models of Communication Informing Educational Research on Computer Gameplay. *Game Studies, 6*(1), Retrieved from gamestudies.org on 9/16/2014.

Bargiela, M. M. (2003). Teacher Feedback and Learners' Uptake. *Linguagem em (Dis)curso, 4*(1), 81-96.

Bartle, R. (1996). Hearts, Clubs, Diamonds, Spades: Players Who Suit MUDs. *Journal of MUD research, 1*(1), Retrieved from brandeis.edu on 9/17/2014.

Brandl, K. K. (1995). Strong and Weak Students' Preferences for Error Feedback Options and Responses. *The Modern Language Journal, 79*(2), 194-211.

Carrier, M. (1991). Simulations in English Language Teaching: A Cooperative Approach. *Simulation & Gaming: An Interdisciplinary Journal of Theory, Practice and Research, 22*(2), 224-233.

Carroll, S., & Swain, M. (1993). Explicit and Implicit Negative Feedback. *Studies in Second Language Acquisition, 15*(3), 357-386.

Castells, M. O., Guzmán, E. E., Lapuerta, P., & Liskin-Gasparro, J. E. (2009). *Mosaicos: Spanish as a World Language* (5th ed.). New Jersey: Prentice Hall.

Cathcart, R., & Olsen, J. (1976). Teachers' and Students' Preferences for Correction of Classroom Conversation Errors. *on TESOL, 76*, 41-53.

Chenoweth, N., 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.

Coffield, F., Moseley, D., Hall, E., & Ecclestone, K. (2004). Learning Styles and Pedagogy in Post-16 Learning: A Systematic and Critical Review. *Learning and Skills Research Centre*, Retrieved from lsda.org on 9/17/2014.

Coleman, D. (2002). On Foot in Sim City: Using Sim Copter as the Basis for an ESL Writing Assignment *Simulation & Gaming: An Interdisciplinary Journal of Theory, Practice and Research 33*(2), 217-230.

Cook, V. (2008). *Second Language Learning and Language Teaching* (4th ed.). Hachette, UK: Hodder Education.

Cornillie, F., Clarebout, G., & Desmet, P. (2012a). Between Learning and Playing? Exploring Learners' Perceptions of Corrective Feedback in an Immersive Game for English Pragmatics. *ReCALL, 24*(3), 257-278.

Cornillie, F., Thorne, S. L., & Desmet, P. (2012b). ReCALL Special Issue: Digital Games for Language Learning: Challenges and Opportunities. *ReCALL, 24*(3), 243-256.

Costikyan, G. (2002). *Toward a Critical Vocabulary for Games*. Paper presented at the Computer Games and Digital Cultures Conference Proceedings.

Davies, M. (2006). *A Frequency Dictionary of Spanish: Core Vocabulary for Learners*. New York, NY: Routledge.

deHaan, J. (2011). Teaching and Learning English Through Digital Game Projects. *Digital Culture & Education 3*(1), 46-55.

DeKeyser, R. (2008). Skill Acquisition Theory. In B. VanPatten & J. Williams (Eds.), *Theories in Second Language Acquisition: An Introduction* (pp. 97-113). Mahwah, NJ: Routledge.

Dilans, G. (2010). Corrective Feedback and L2 Vocabulary Development: Prompts and Recasts in the Adult ESL Classroom. *The Canadian Modern Language Review, 66*(6), 787-815.

Ellis, R. (2003). *Task-Based Language Learning and Teaching*. Oxford: Oxford University Press.

Ellis, R. (2009a). Corrective Feedback and Teacher Development. *L2 Journal, 1*(1), Retrieved from escholarship.org on 9/16/2014.

Ellis, R. (2009b). The Differential Effects of Three Types of Task Planning on the Fluency, Complexity, and Accuracy in L2 Oral Production. *Applied Linguistics, 30*(4), 474-509.

Espinar, K. K., & Baxter, P. (2012). Task-Based Language Teaching (TBLT). Retrieved from umb.edu on 9/16/2014.

Folse, K. S. (2006). The Effect of type of Written Exercise on L2 Vocabulary Retention. *TESOL Quarterly, 40*, 273-293.

Gass, S. M., & Selinker, L. (2008). *Second Language Acquisition: An Introductory Course* (3rd ed.). New York: Routledge.

Gee, J. P. (2003). *What Video Games Have to Teach us About Learning and Literacy* (1st ed.). New York, NY: Palgrave Macmillan.

Gee, J. P. (2005). Good Video Games and Good Learning. *Phi Kappa Phi Forum, 85*(2), 33-37.

Gee, J. P. (2007). *What Video Games Have to Teach Us About Learning and Literacy* (2nd ed.). New York, NY: Palgrave Macmillan.

Gee, J. P. (2012). Foreword. In H. Reinders (Ed.), *Digital Games in Language Learning and Teaching* (pp. xii-xiv). New York, NY: Palgrave Macmillan.

Glazer, S. (2006). Video Games: Do They Have Educational Value? *CQ Researcher, 16*(40), 937-960.

Greenwald, A. G., & Gillmore, G. M. (1997). Grading Leniency is a Removable Contaminant of Student Ratings. *American Psychologist, 52*(11), 1209-1217.

Harper, D. (2001). Online Etymology Dictionary. Retrieved from etymonline.com on 9/17/2014.

Havranek, G. (2002). When is Corrective Feedback Most Likely to Succeed? *International Journal of Educational Research, 37*(3), 255-270.

Hedgcock, J., & Lefkowitz, N. (1994). Feedback on Feedback: Assessing Learner 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*(1), 99-109.

Heift, T. (2004). Corrective Feedback and Learner Uptake in CALL. *ReCALL, 16*(2), 416-431.

Heift, T., & Chapelle, C. A. (2012). Language Learning through Technology. In S. M. Gass & A. Mackey (Eds.), *The Routledge Handbook of Second Language Acquisition* (pp. 555-570). New York, NY: Routledge.

Hilgendorf, S. K. (2012). History of Language Teaching Methods. *The Encyclopedia of Applied Linguistics*, Retrieved from onlinelibrary.wiley.com on 9/17/2014.

Howatt, A. P. R., & Widdowson, H. G. (1984). *A History of English Language Teaching* (Vol. 6). Oxford: Oxford University Press.

Hubbard, P. (2002). Interactive Participatory Dramas for Language Learning. *Simulation & Gaming, 33*(2), 210-216.

Jackson, G. T., Dempsey, K. B., & McNamara, D. S. (2012). Game-Based Practice in a Reading Strategy Tutoring System: Showdown in iSTART-ME. In H. Reinders (Ed.), *Digital Games in Language Learning and Teaching* (pp. 115-138). New York, NY: Palgrave Macmillan.

Johnson, L., Adams, S., & Cummins, M. (2012). The NMC Horizon Report: 2012 Higher Education Edition *The NMC Horizon Report*. Austin, TX: The New Media Consortium.

Jordan, G. (1992). Exploiting Computer-Based Simulations for Language-Learning Purposes *Simulation & Gaming: An Interdisciplinary Journal of Theory, Practice and Research 22*(1), 88-98.

Krashen, S. (1982). *Principles and Practice in Second Language Acquisition*. Oxford: Pergamon.

Larsen-Freeman, D. (2011). *Techniques And Principles In Language Teaching*. Oxford: Oxford University Press.

Laufer, B., & Nation, I. S. P. (2012). Vocabulary. In S. M. Gass & A. Mackey (Eds.), *The Routledge Handbook of Second Language Acquisition* (pp. 163-176). New York, NY: Routledge.

Lee, J. F., & VanPatten, B. (2003). *Making Communicative Language Teaching Happen* (2nd ed.). New York, NY: McGraw-Hill.

Loewen, S. (2012). The Role of Feedback. In S. M. Gass & A. Mackey (Eds.), *The Routledge Handbook of Second Language Acquisition* (pp. 24-40). New York, NY: Routledge.

Loewen, S., & Philp, J. (2006). Recasts in the Adult English L2 Classroom: Characteristics, Explicitness, and Effectiveness. *The Modern Language Journal, 90*(4), 536-556.

Long, M. H. (1996). The Role of the Linguistic Environment in Second Language Acquisition. In W. C. Ritchie & T. K. Bhatia (Eds.), *Handbook of Second Language Acquisition* (pp. 413-468). New York: Academic Press.

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.

Lyster, R., & Ranta, L. (1997). Corrective Feedback and Learner Uptake: Negotiation of Form in Communicative Classrooms. *Studies in Second Language Acquisition, 20*, 37-66.

Lyster, R., & Ranta, L. (2013). Counterpoint Piece: The Case for Variety in Corrective Feedback Research. *Studies in Second Language Acquisition, 35*(1), 167-184.

Markey, K., Swanson, F., Jenkins, A., Jennings, B. J., Jean, B. S., Rosenberg, V., & Frost, R. L. (2008). Designing and Testing a Web-Based Board Game for Teaching Information Literacy Skills and Concepts. *Library Hi Tech, 26*(4), 663-681.

McGonigal, J. (2011). *Reality is Broken: Why Games Make us Better and How They Can Change the World*. New York: The Penguin Press.

McGowan, J. (2011). Survey: Turning Off Game Music. *Bleeps and Pops*, Retrieved from bleepsandpops.com on 9/17/2014.

Merriam-Webster. (2014). Merriam-Webster Dictionary. Retrieved from merriam-webster.com on 9/17/2014.

Nagata, N. (1993). Intelligent Computer Feedback for Second Language Instruction. *The Modern Language Journal, 77*(3), 330-339.

Nation, I. S. P. (2001). *Learning Vocabulary in Another Language*. Cambridge: Cambridge.

Nelson, B. (1998). Web-Based Vocabulary Activities: Pedagogy and Practice. *Computer Assisted Language Learning, 11*(4), 427-435.

Neville, D. O., Shelton, B. E., & McInnis, B. (2009). Cybertext Redux: Using Digital Gamebased Learning to Teach L2 Vocabulary, Reading, and Culture. *Computer Assisted Language Learning, 22*, 409-424.

Novak, J. (2012). *Game Development Essentials* (3rd ed.). Canada: Delmar Cengage Learning.

Peterson, M. (2012). Learner Interaction in a Massively Multiplayer Online Role Playing Game (MMORPG): A Sociocultural Discourse Analysis. *ReCALL, 24*(3), 361-380.

Polio, C. (2012). The Acquisition of Second Language Writing. In S. M. Gass & A. Mackey (Eds.), *The Routledge Handbook of Second Language Acquisition* (pp. 319-334). New York, NY: Routledge.

Prabhu, N. S. (1987). *Second Language Pedagogy*. Oxford: Oxford University Press.

Prensky, M. (2001). *Digital Game-Based Learning*. New York, NY: McGraw-Hill.

Pujolà, J.-T. (2001). Did CALL Feedback Feed Back? Researching Learners' Use of Feedback. *ReCALL, 13*(1), 79-98.

Purushotma, R., Thorne, S., & Wheatley, J. (2008). 10 Key Principles for Designing Video Games for Foreign Language Learning. Retrieved from knol.google.com/k/ravi-purushotma on 5/10/2009.

Radecki, P. M., & Swales, J. M. (1988). ESL Student Reaction to Written Comments on Their Written Work. *System, 16*(3), 355-365.

Rama, P. S., Black, R. W., van Es, E., & Warschauer, M. (2012). Affordances for Second Language Learning in World of Warcraft. *ReCALL, 24*(3), 322-338.

Reid, J. M. (1995). *Learning Styles in the ESL/EFL Classroom*. Florence, KY: Heinle & Heinle Publishers.

Reinders, H. (2009). Using Computer Games to Teach Writing. *English Teaching Professional 63*, 6-58.

Reinders, H. (2012). Digital Games in Language Learning and Teaching: Introduction. Retrieved from UTXA.eblib.com on 9/17/2014.

Reinders, H., & Wattana, S. (2012). Talk to Me! Games and Students' Willingness to Communicate. In H. Reinders (Ed.), *Digital Games in Language Learning and Teaching* (pp. 156-188). New York, NY: Palgrave Macmillan.

Reinhardt, J., & Sykes, J. (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 Technology* (pp. 32-49). New York, NY: Palgrave Macmillan.

Richards, J. C., & Rodgers, T. S. (2002). *Approaches and Methods in Language Teaching*: Ernst Klett Sprachen.

Salaberry, M., Barrette, C., María, F.-G., & Nevárez, I. C. (2013). *Conectándonos*. Debuque, IA: Kendall Hunt.

Salaberry, R., Barrette, C. M., Elliott, P., & Fernández-García, M. (2004). *Impresiones*. New Jersey: Pearson.

Saucier, G. (1994). Mini-Markers: A Brief Version of Goldberg's Unipolar Big-Five Markers. *Journal of Personality Assessment, 63*(3), 506-516.

Schell, J. (2008). *The Art of Game Design: A Book of Lenses*. USA: Elsevier.

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.

Shuler, C. (2012). What in the World Happened to Carmen Sandiego? *The Edutainment Era*, Retrieved from joanganzcooneycenter.org on 9/16/2014.

Squire, K. (2002). Cultural Framing of Computer/Video Games. *Game Studies: The International Journal of Computer Games Research 2*(1), Retrieved from gamestudies.org on 9/16/2014.

Squire, K. (2011). *Video Games and Learning: Teaching and Participatory Culture in the Digital Age*. New York, NY: Teachers College Press.

Sundqvist, P., & Kerstin Sylvén, L. (2012). World of VocCraft: Computer Games and Swedish Learners' L2 English Vocabulary. In H. Reinders (Ed.), *Digital Games in Language Learning and Teaching* (pp. 189-208). New York, NY: Palgrave Macmillan.

Sykes, J. M., & Reinhardt, J. (2012). *Language at Play: Digital Games in Second and Foreign Language Teaching and Learning*. Boston: Pearson Higher Ed.

Sylvén, L. K., & Sundqvist, P. (2012). Gaming as Extramural English L2 Learning and L2 Proficiency Among Young Learners. *ReCALL, 24*(3), 302-321.

Takimoto, M. (2006). The Effects of Explicit Feedback on the Development of Pragmatic Proficiency. *Language Teaching Research, 10*(4), 393-417.

Tennant, J., & Gardner, R. (2004). The Computerized Mini-AMTB. *CALICO Journal, 21*(2), 245-263.

Thomas, M. (2012). Contextualizing Digital Game-Based Language Learning: Transformational Paradigm Shift or Business as Usual? In H. Reinders (Ed.), *Digital Games in Language Learning and Teaching* (pp. 11-31). New York, NY: Palgrave Macmillan.

Thorne, S. L. (2008). Mediating Technologies and Second Language Learning. In D. Leu, J. Coiro, C. Lankshear & M. Knobel (Eds.), *Handbook of Research on New Literacies* (pp. 417-449). Mahwah, NJ: Lawrence Erlbaum.

Thorne, S. L., Black, R. W., & Sykes, J. M. (2009). Second Language Use, Socialization, and Learning in Internet Interest Communities and Online Gaming. *The Modern Language Journal, 93*, 802-821.

Thorne, S. L., Fischer, I., & Lu, X. (2012). The Semiotic Ecology and Linguistic Complexity of an Online Game World. *ReCALL, 24*(3), 279-301.

Totilo, S. (2011). The Year I Gained The Courage To Ignore Video Game Music. *2014*(2/19), Retrieved from kotaku.com on 2/19/2014.

Truscott, J. (1996). The Case Against Grammar Correction in L2 Writing Classes. *Language Learning, 46*(2), 327-369.

Van Bussel, F. J. J. (1994). Design Rules for Computer-Aided Learning of Vocabulary Items in a Second Language. *Computers in Human Behavior, 10*, 61-76.

Van den Branden, K. (2012). Task-Based Language Education. In A. Burns & J. C. Richards (Eds.), *The Cambridge Guide to Pedagogy and Practice in Second Language Teaching* (pp. 132-139). Cambridge, MA: Cambridge University Press.

Van den Branden, K., Bygate, M., & Norris, J. (2009). Task-Based Language Teaching: Introducing the Reader. In K. Van den Branden, M. Bygate & J. M. Norris (Eds.), *Task-Based Language Teaching: A Reader* (pp. 1-13). Philadelphia: John Benjamins .

Vygotsky, L. S. (1980). *Mind in Society: The Development of Higher Psychological Processes*. Cambridge: Harvard University Press.

Weigand, E. (2010). *Dialogue: The Mixed Game*. Amsterdam: John Benjamins.