Crystallize: An Immersive, Collaborative Game for Second Language Learning

Gabriel Culbertson[1], **Erik Andersen**[2], **Walker White**[2], **Daniel Zhang**[2], **and Malte Jung**[1]

[1]Department of Information Science, [2]Department of Computer Science, Cornell University

{grc74, ela63, wmw2, dwz24, mfj28}@cornell.edu

ABSTRACT

Learning a second language is challenging. Becoming fluent requires learning contextual information about how language should be used as well as word meanings and grammar. The majority of existing language learning applications provide only thin context around content. In this paper, we present *Crystallize*, a collaborative 3D game that provides rich context along with scaffolded learning and engaging gameplay mechanics. Players collaborate through joint tasks, or quests. We present a user study with 42 participants that examined the impact of low and high levels of task interdependence on language learning experience and outcomes. We found that requiring players to help each other led to improved collaborative partner interactions, learning outcomes, and gameplay. A detailed analysis of the chat-logs further revealed that changes in task interdependence affected learning behaviors.

Author Keywords

language learning, video games, collaboration

ACM Classification Keywords

K.3.0. Computers and Education: General

INTRODUCTION

Learning a second language is a complex and challenging process. Most adult learners never achieve proficiency [7] and research on the mechanisms of second language learning is inconclusive [33]. Traditionally, language has been conceptualized as abstract mental structures that exist independently of context. However, in recent years, some limitations of this view have led language acquisition researchers to develop a more holistic view of language knowledge [19]. This view suggests that learning a second language is not only the process of memorizing abstract meanings and syntax, but also knowledge situated in a physical and cultural context.

Most existing language learning tools take the traditional view that language can be learned in terms of abstract linguistic components. For example, DuoLingo primarily uses translation for learning. In translation, the student learns how to say words and put them in an order that fits the language grammar, but ideas are separated from the context. In another tool, Rosetta Stone, users learn using pictures. While this offers more context, there is discontinuity between problems and the application lacks other important aspects of context present in the real world.

In the social, situated view of language learning, language knowledge cannot be separated from the context it is learned in. Researchers have pointed to the importance of considering a more holistic context in learning. Gee [21] argues that humans comprehend linguistic meaning through concrete experiences that are situated in the real world rather than abstract propositional reasoning. This is further supported by the theory of encoding specificity [52], which argues that recall is highest when there are perceptual similarities between the context of initial learning and eventual use.

Both perspectives have advantages. A situated approach can give learners a better sense of how to use language in practice, while a cognitive approach can make learning more approachable and engaging. Ideally, we would have tools that can provide rich visual and situational context for learners while engaging learners with scaffolded learning experiences and fun game mechanics.

A game called *Influent* recently explored this idea, by having the player walk around a 3D environment and learn the vocabulary associated with objects in that environment [23]. This offered an exciting new way to interact with language, but we believe there is potential to enrich context with other aspects such as conversation and narrative. More importantly, the game lacked real interactions with other people.

Since language is inherently social, ideally multiple learners should be able to interact in a shared virtual environment. Massively multiplayer online games (MMOG) do just this. Work by Yee [54] has shown that social behaviors and cultural norms can cross the barrier into virtual environments like Second Life. Online collaboration not only enhances context, but also multiple studies have established a link between collaboration and motivation for learning in MOOCs [27, 30] and STEM classes [49].

In this paper, we present *Crystallize*, a 3D video game for second language learning. *Crystallize* is a learning experience that prioritizes observation, inference, experimentation, and feedback over rote memorization. Players control an avatar and navigate a virtual environment that simulates being immersed in a real target language environment. Players collaborate to complete language "quests" that require them to find words in the environment and use them to accomplish objectives. Using *Crystallize*, we can both situate and structure learning experiences.

CSCW '16, February 27-March 02, 2016, San Francisco, CA, USA
2016 ACM. ISBN 978-1-4503-3592-8/16/02 $15.00
DOI: http://dx.doi.org/10.1145/2818048.2820020

The primary difference between Crystallize and existing collaborative learning approaches is that Crystallize is built around core game mechanics that: 1) incentivize users to explore both the physical environment and the space of sentences and phrases that can be constructed, 2) create engaging feedback loops, and 3) facilitate synchronous collaboration. To our understanding, these mechanics have not been combined in previous language learning games.

Collaboration has been shown to be an effective tool for learning, however structuring tasks to be collaborative without adding complexity that may take away from learning can be challenging. To address this, we use joint quests to encourage players to work together. The quests include both language learning objectives and objectives that encourage players to help their partner complete those objectives as well. Players can communicate through a standard chat interface and the players' avatars share the same virtual space.

In addition to the design of the game itself, we present results from a user study with 42 players showing that increased task interdependence led to improved partner interactions, improved subjective and objective language learning outcomes, and slower but more careful engagement with the game's content. *Task interdependence* refers to the degree to which players must depend on one another to reach a common goal and therefore structures collaborative interactions. The structure of interaction has been shown to heavily influence learning outcomes [26], and designing collaborative tasks without careful consideration of this structure can have negative consequences for learning [4].

The core contributions of this paper are: (1) A novel collaborative game platform that enables situated language learning in tea, (2) three game mechanics that encourage collaboration, and (3) a user study suggesting that collaborative learning in games can benefit from increased interdependence.

RELATED WORK

Researchers have explored many methods of designing interactive instructional materials to teach language. In this section, we provide an overview of existing work.

Individual language learning technology

Traditionally, digital tools made use of only written or audio context and focused only on specific aspects of learning. Tip Tap Tones trains users to recognize tones in Chinese [17]. MemReflex adaptively changes flashcard learning schedules to maximize retention [16]. Dearman et al. use a desktop wallpaper to reinforce vocabulary meanings [10]. DuoLingo [53] is another successful tool that teaches language by asking the learner to translate sentences. DuoLingo also has a highly structured learning progression and has achievements and point systems to motivate users.

More recently, other tools have begun to integrate other contextual information. Rosetta Stone teaches language through a series of pictures. A typical task features a set of four or more pictures that each show a certain situation, such as a boy eating or a girl running, and ask the user to identify the picture that most closely matches a phrase in the target language. Rosetta Stone offers many advantages over traditional curricula: the learner receives immediate feedback, information is presented in a visual context, and meaning is often learned through inference.

Other tools have pushed this idea even further. MicroMandarin, for example, makes use of the user's real world location as a source for vocabulary words [18]. A video game, *Influent*, uses a 3D environment to teach vocabulary words [23]. This game makes use of visual context to teach vocabulary meanings, but does not provide a set of deep, structured, and layered challenges to motivate the player. Furthermore, it only deals with individual words whereas we propose to cover phrases and sentences.

Collaborative learning

Some tools have been developed to support collaborative language learning. For example, in LiveMocha[1], learners can post exercises or other learning content which other users can purchase for virtual currency. Users can also use their virtual currency to purchase live tutoring sessions with other users. However, this system does not use deeply embedded game mechanics to engage users, and the system uses tutor-student collaboration rather than peer-to-peer collaboration.

While synchronous collaboration as a learning approach has been less explored through learning games, multiple studies of traditional learning approaches have shown that collaborative learning and peer learning is conducive to learning. For example, Springer et al. [49] conducted a meta-analysis of STEM-focused classes and found that learning in small groups improves student motivation and learning outcomes. In another meta-analysis of collaborative learning approaches, Johnson et al. [26] found additional support for these findings and also determined that the structure of collaboration heavily influences learning outcomes.

The benefits of collaboration extend to software based learning environments. For example, studies on massive open online courses, or MOOCs, have shown that adding elements of synchronous collaboration improves learning outcomes. For example, Kizilcec and Schneider [27] found that taking MOOCs with peers increased engagement and participation. Kulkarni et al. [30] found that collaborative discussions led to improved engagement and performance in an online class. In both of these studies, short synchronous collaborative activities were used to support learning.

The benefits of collaborative learning have also motivated several researchers to integrate collaborative learning features into software based learning tools. For example, Piper and colleagues [42] used collaboration as a core feature in a tabletop game for social skills development for children with Asperger's Syndrome. Also, Singley and colleagues [48] integrated synchronous collaboration aspects into an algebra learning software intended to support collaborative learning through features that support the establishment of common ground and the maintenance of group focus. Finally, Larson

[1] www.livemocha.com

Figure 1. *Crystallize* game prototype. Counter-clockwise beginning from the middle-left: a description of current objectives, a chatbox for communicating with the other player, a word inventory that contains all of the words the player has learned so far, and (center) a prompt for constructing a sentence. The green and blue lines show players where their partners are.

et al. [34] created a mobile game for learning English pronunciation that used collaboration and competition as core motivating features.

To our understanding, none of the studies and learning platforms examined how synchronous collaboration as a continuous, core feature of a learning platform affects motivation and learning outcomes, especially for language learning. While prior work suggests that collaboration can support learning, it is less clear how to design for effective collaboration in language learning and how collaboration should be structured in educational technology.

Collaboration and games

There are multiple studies in HCI focusing on optimizing player experience in video games [2, 3, 36]. Since video games often use collaboration and multiplayer game mechanics to engage players, some studies have focused specifically on social interaction in games. Johnson et al. [24] found that playing with human teammates led to greater sense of relatedness than playing with computer-controlled teammates, but actually led to less competence and flow. Furthermore, Ducheneaut et al. [15] found that although MMORPGs like World of Warcraft facilitate and benefit from social interaction, a lot of the player experience is actually more solitary than one might expect. Ducheneaut and Moore found that design choices, such as forcing players to take a break from game activities to interact with other players, significantly influenced the types of interactions that took place in Star Wars Galaxies [14]. Toups et al. found that a carefully designed game could improve communication in crisis response teams [51]. This motivated us to carefully design the collaborative aspects of *Crystallize* and explore the importance of interdependence through a user study.

GAME DESIGN

We have constructed an early prototype game, *Crystallize*, which leverages some aspects of language immersion to engage students and enrich learning. *Crystallize* can be downloaded and played from the game website[2]. *Crystallize* has four primary learning goals: (1) motivate players with a sequence of clear and achievable goals, (2) facilitate learning of word meanings through active logical inference rather than passive memorization, (3) provide opportunities for structured exploration of grammatical structures, and (4) enable collaborative language learning. A screenshot of *Crystallize* can be seen in Figure 1. The game is designed to work with any target language and source language, but this particular prototype teaches Japanese to English speakers. Japanese grammar is very different from English grammar, which allows us to evaluate the ability of the game to facilitate learning of unfamiliar grammatical structures.

A primary goal of *Crystallize* is to use game mechanics to motivate players to learn. We use Koster's definition of game mechanics: "Game mechanics are rule based systems / simulations that facilitate and encourage a user to explore and learn the properties of their possibility space through the use of feedback mechanisms" [28]. In *Crystallize*, the core mechanics require players to build sentences from words in order to gain information from the computer controlled characters. Players are rewarded with positive feedback when they are able to successfully communicate with these characters. This leads to two core feedback loops. First, players learn new words by using their current words to interact with the computer characters. Second, by completing conversations with characters in one level, new levels are unlocked with new available conversations.

Motivate players through quests

Games often motivate players through achievements and quests, which are challenges that can be completed and tracked. Denny found that adding student achievements to an educational game increased participation [12]. Therefore, *Crystallize* uses a questing system that guides the player through a set of learning objectives while still providing freedom to explore. Quests are completed by successfully conversing with computer controlled characters. Figure 2 also shows the two stages of each quest. In the first stage, the player must find and learn the words necessary to complete the conversation. In the second, the player must construct the phrases needed for completing the conversation.

Infer word meanings from context

Schneider et al. suggest that learning words through challenges rather than being given the meaning increases long term retention [47]. Therefore, *Crystallize's* challenges are designed to help the player learn word meanings from context. A primary mechanism for this is overhearing conversations between computer controlled characters. Each location is filled with dialogue that the player can overhear. If someone says "What are you <*unknown word*>?" and another character responds with "I am <*unknown word*> sushi," the player might be able to guess that the missing word is "eating". If a character says "This is an <*unknown word*>." while pointing to a cat, the word is probably "cat". This process can be seen in Figure 2.

[2]http://gdiac.cs.cornell.edu/gallery/download.php?name=crystallize

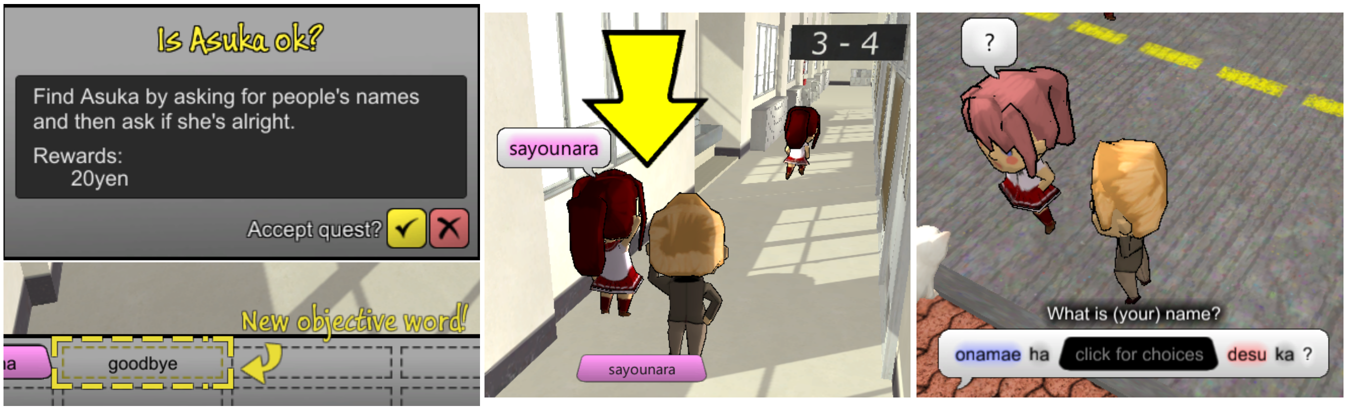

Figure 2. ***Crystallize*** **is composed of a series of language learning quests. The top left image shows the quest overview panel that is shown to players before deciding whether or not to take on the task. Each quest has two parts: find the relevant words and use them to construct a sentence. The words the player needs to find (bottom left) are indicated by English words with a dashed outline. Players collect words by overhearing dialogues. Word meanings are inferred from context. In the middle image, the player can learn "sayounara," which means "goodbye," because a character says it while leaving. Once the words have been collected, the player needs to complete a sentence using the prompt provided (right image).**

Figure 3. The player can communicate with computer controlled characters and also with each other by constructing sentences. The player does this by dragging words from the inventory into an interface widget that prompts the player to construct a desired sentence. Some of the words in this sentence may be filled in already. These challenges are scaffolded so that the player is gradually asked to complete more of the sentence over time. This technique, gradually increasing the portion of the problem that the student is asked to complete, has been shown to increase student performance on near-transfer tasks and reduce the abruptness of the transition from observing solved demonstrations to solving problems independently [43].

Discover grammatical structures

The theory of the Zone of Proximal Development (ZDP) states that each learner has a set of concepts that can be learned if provided guidance and that learning materials should continually target this set [38]. As a result, the game requires the player to construct increasingly complex sentences. This balance of challenge with skill is not only good for learning, but also enables the player to reach a flow state [9]. In *Crystallize*, some words are given to the player and verbs are already conjugated in order to build challenges at rate which is challenging without being overwhelming. As shown in Figure 3, when learning the question "What is your job?", the player is first asked to fill in the word "what", and later asked to fill in both "what" and "is". Eventually the player must complete the whole question. By gradually increasing the portion of the problem that the student is asked to complete in this way, the game eases learning of difficult sentences. This technique, which is called fading worked-out demonstrations, has been shown to increase student performance on near-transfer tasks [43]. If the player fills in the sentence incorrectly, the player receives immediate feedback and must fix the sentence before continuing the conversation.

Drag and drop sentence construction allows for quick experimentation. By minimizing the effort of rearranging words, emphasis is placed on grammatical structures, allowing players to see patterns that may have otherwise been unclear. Currently, the player can only complete the sentences as specified by the game, which is a limitation. In the future, we plan to develop a more flexible system that can interpret a wider range of potentially correct user input.

Learn with another player

In education, collaborative learning has been found to produce higher levels of academic achievement and to encourage the development of supportive relationships with co-learners [31]. However, the implementation details are important for the success of a collaborative learning program. Activities must be flexible enough for students to establish social roles and negotiate understanding, but be structured enough to guide students toward learning goals [44]. However, supporting the nuance of social interactions and grounding activity can be challenging in computer supported collaborative work [1].

In order to ground communication, the player avatars share the same virtual space. This allows players to easily guide other players to destinations and refer to virtual environment elements. Arrows connect players so that each player can locate the other player and have some idea of what they are doing. In cases where a player feels lost or confused, this makes it possible to easily locate the other player and learn from their actions.

Communication takes place over an independent chat channel rather than tying communication to virtual location. Players may recognize their in-game avatar as a tool or prop rather than a virtual representation of themselves [35], therefore spatially limiting chat unnecessarily constrains communication. We also included 3 emote buttons ('hi!', 'thanks :)', and 'nice!') to encourage supportive interactions. These buttons cause the player avatar to perform an animation (wave, bow and clap respectively) as well as write text to the chat box.

The collaborative quests provide structure to guide the interaction. The quests open opportunities for players to discuss language and gameplay as well as encourage one another. For example, one quest requires players to ask the name of a computer controlled character. If one of the players can't remember how to ask this question, he or she might ask "where does 'ka' go in the sentence again?". This discussion reinforces grammatical concepts for for both players. In another quest, players learn the word for 'table' by asking a computer controlled character "what is this?". If one of the players cannot figure out the challenge, the other player can offer help by approaching the correct character and saying "you need to talk to this person here". This type of assistance not only smooths out gameplay experiences, but can also bring the learners closer together. Finally, when players receive help or complete a difficult challenge together, they can use the emote buttons to give positive feedback.

Breakdown of gameplay

The game is broken down into levels. A level can contain multiple quests and contains other computer controlled characters that may not be directly related to the player's objectives. After the tutorial, each level uses a different environment. Once all of the objectives for a level are completed, the player is automatically moved to the next level.

Players begin by completing a series of tutorial levels. All of these levels take place in the same environment. The objectives in these levels are intended to teach the gameplay mechanics, although there is some Japanese language in these levels as well.

After the introduction to the game, all progress through the game takes place in the form of quests. Quests are given to the player by 'quest clients' (indicated by exclamation marks above their heads) and must be accepted by the player before they can begin. Once the players complete a quest, they need to seek out the next 'quest client' to begin the next quest. Once Players must complete all of the quests in level of the game before than can proceed to the next level. A summary of the content in the quests is shown in Table 1.

Level	Quest	Vocabulary, phrases and other skills learned by the player
T1	N/A	hello, *drag words from speech bubbles*, *make phrases*
T2	N/A	*move*
T3	N/A	goodbye, *overhear dialogues*
T4	N/A	student, to be, question marker, (I) am a student, (teacher), *sentences*
T5	N/A	I am a student, I, also, (particle ha)
T6	N/A	(yes), *see all learned words*
T7	N/A	*use dictionary*
1	1	(she), (what), (possessive particle), (what's her name?), *accept quests*, *interact with partner*
2	2	name, what's your name?, (he), (what's his name?), (I am)
3	3	(at), (where), (to be at somewhere), (where is she?)
	4	what, this, what is this?, (new), (word), (direct object particle), (learn), (cat), (plant), table
4	5	alright, are you alright?, (where is he?)
	6	(is she alright?), (yes, I'm alright)
5	7	(car)
	8	cat
	9	job, chef, what is your job?, (shop staff), (what is his job?)
	10	shop staff, (buy), (ramen)
	11	buy, please, ramen, I'd like some ramen, (tasty), (similar)
	12	water, drink, I'd like some water, (coffee)
	13	university student, what are you eating?, (I will eat sushi), (tempura)
	14	I would like some sushi, (what would you like to eat?)

Table 1. Game content divided into levels and quests. Levels beginning with T are tutorial levels. New words and phrases that must be used to complete the give quest or level are shown in normal font. Words and phrases that are available, but do not need to be learned are shown in parentheses. Some quests do not have new words and phrases because they are used to review old content. Skills that the player must learn to play the game are shown in italics.

USER STUDY: THE IMPACT OF INTERDEPENDENCE

As collaboration is the key characteristic of language learning in *Crystallize*, the focus of our user study is on gaining insights into the effect collaboration has on language learning. We were particularly interested in finding out how to find the right degree and type of collaboration that is conducive to language learning, and how to design for collaborative language learning experiences.

To compare different types of collaboration we drew from McGrath's Group Task Circumplex, a widely used framework, that distinguishes types of group tasks [37]. One of the two main dimensions for distinguishing types of group tasks in the task circumplex is the level of task interdependence [50, 37]. Task interdependence refers to the degree to which team members must depend on each other to perform their tasks in order to reach a goal [45]. Therefore the higher the level of task interdependence the more coordination is necessary by a team to perform well [46]. Studying task interdependence is particularly interesting for us as it is not clear whether increased interdependence is beneficial or harmful for collaboration in a language learning game.

We might expect increased task interdependence to improve collaborative language learning along three dimensions: (1) Partner interactions, (2) learning outcomes, and (3) gameplay. Prior work has shown that increased task interdependence leads to an improvement in *partner interactions* through an increase in communication, helping behavior, and information sharing [8, 25]. An increase of helping behaviors has been shown to improve *learning outcomes* [40]. Additionally, increased task interdependence might make the presence of other group members more salient and improve learning as research by Okita and colleagues [41] has shown that the mere belief in taking socially relevant action improves learning. Finally, *learning outcomes* and *gameplay* might be improved as increased task interdependence might has been shown to lead to increased motivation in completing a task through a social commitment effect: A lack of performance not only impacts one's own performance but also that of one's group members. For example Drnyei [13] argued that cooperative goal structures create a motivational system that positively influences learning outcomes. In line with this [39] argued that social engagement promotes learning motivation. Based on this prior work we offer the hypothesis (H1a) that increased task interdependence leads to an improvement of collaborative language learning along the dimensions of partner interactions, learning outcomes, and gameplay.

Despite the many studies that highlight the advantages of increased task interdependence, it is also plausible that increased task interdependence worsens collaborative language learning along the three dimensions. For example, increasing task interdependence might lead to worsened partner interactions. learning outcomes, and gameplay due to an increase in the number of interruptions through other group members [20]. Further, as time on task has been identified as a key determinant of learning outcomes [22], the additional level of coordination required for tasks with high task interdependence might directly interfere with learning outcomes because less of the available time is spent for language learning in favor of group communication and coordination. In sum, this work suggests an alternative hypothesis (H1b) that increased task interdependence leads to a worsening of collaborative language learning along the dimensions of partner interactions, learning outcomes, and gameplay.

Study design

To examine how the type of collaboration influences language learning we studied the impact of task interdependence on learning experience and outcomes in a between-participants study design with two conditions: high task interdependence vs. low task interdependence. Participants were randomly assigned to conditions. In the high interdependence condition, participants saw a subtly modified task description for each quest that asked the players to assist their partners with the task through the following statement: "help your partner complete the quest" (Figure 4). Research by Okita and colleagues [41] has shown that the mere belief in taking socially relevant action improves learning and we wanted to see if such a subtle change could lead to a measurable change in collaborative langage learning. In addition to the modification of the quest task description the game required participants in the high interdependence conditions that both users complete each task before the pair could proceed. Once both of the players had completed all of the quests in a level, the game automatically transitioned to the next level. In the low interdependence condition, players were able to continue as soon as they had completed all of the quests in a level. Once the player had finished all the tasks, a message box appeared that allowed the player to continue. The message box was placed so that it did not interfere with the gameplay. This way, the player could choose to stay and help the partner or continue alone.

Participants

We recruited 48 participants through fliers, face-to-face interactions, and our university's research system. 42 participants were included in our final analyses. Four participants (two pairs) were excluded from the final analysis due to technical difficulties and two more participants (one pair) were excluded because they did not follow the study procedures. Participants ranged from 19 to 30 years of age (M = 21.9, SD = 2.3) with 43.5 percent males and 56.5 percent females. Participants were compensated with either $10 or course credit. Participants were paired in groups of 2 and randomly assigned to one of the two conditions, resulting in 20 participants the low interdependence and 22 participants in the high independence condition. All participants had studied at least one second language before.

Experiment procedure

We scheduled participants to arrive in pairs of two at one of two laboratory rooms in order to avoid participants seeing each other, and thereby affecting our interaction measures. Upon arrival the participants were brought into a small room, sat in front of a computer, and were given a brief overview of the experiment. After giving consent to participate in our study, participants were asked to complete a brief demographic survey and were given a short language pretest that scored their ability to translate simple words and sentences. Immediately after this, the experimenter introduced and started the game. First, the experimenter asked each participant to complete a short (roughly 10 minute) individual tutorial intended to familiarize participants with the gameplay and controls. The experimenter also introduced participants to the chat interface that would allow participants to communicate throughout the tutorial and the main game experience. As the first participant completed the tutorial a screen appeared asking to wait for the partner to complete the tutorial. Participants could still chat during this time.

Immediately after the tutorial the main game started and participants were asked to complete a series of quests within the remainder of the 25 minutes. Dependent on the experimental condition participants were either presented with the high-interdependence or low-interdependence quest descriptions. The lengh of the game was designed such that no participant was able to complete all of the challenges in the given time.

Once the 25 minutes were over, participants completed a post-task questionnaire. The post-task questionnaire included a language post-test with the same vocabulary and sentence translation tasks as in the language pre-test and several questions to assess participants' partner interactions, learning outcomes, and gameplay. Finally experimenters conducted semi-structured interviews with the participants after which participants were thanked for their time, and paid or given class credit.

Measures

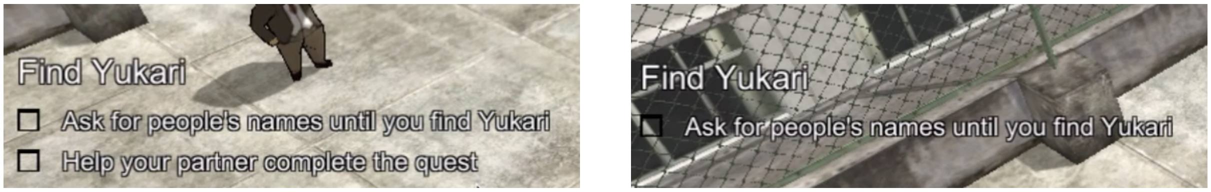

Figure 4. Experimental conditions. A key design question in collaborative learning is *interdependence*, or the degree to which learners must rely on one another to proceed. High interdependence may encourage users to help each other more, but may also limit their progress and frustrate them if their partner is unable to finish the task as quickly as them. To explore the optimal level of interdependence in Crystallize, we created two different experimental conditions that varied the degree of interdependence. In the high interdependence condition (left), players must wait for their partner to complete the quest before proceeding, and are given the task of "Help your partner complete the quest." In the low interdependence condition (right), which prioritizes freedom, players only have to complete their own tasks and are then able to proceed.

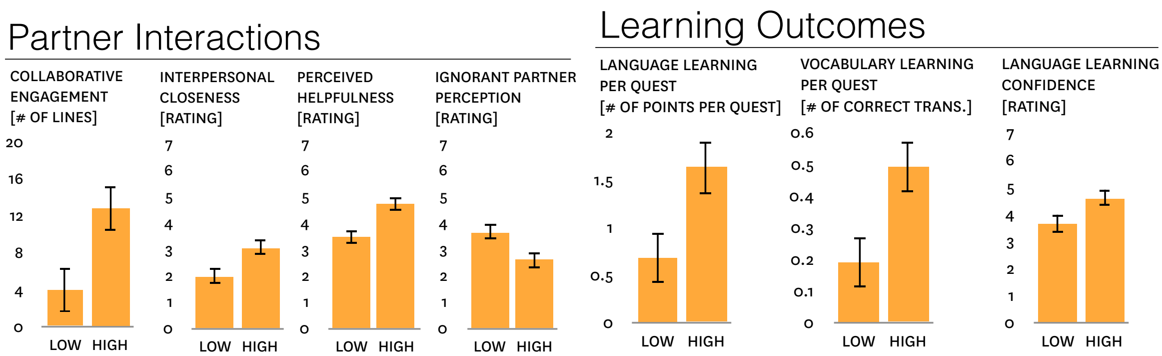

Figure 5. Partner interactions and learning outcomes. Each chart shows the differences between the low and high interdependence conditions. The four charts on the left examine partner interactions. These charts show that higher interdependence led to more collaboration (as measured by more lines typed in the chat box), greater feelings of closeness with the partner, a greater sense that the partner was helpful, and less perception of being ignored by the partner. The three charts on the right examine learning outcomes. We found that higher interdependence led to more vocabulary and grammar being learned per quest, more vocabulary being learned per quest, and greater feelings of confidence in one's own language learning abilities. Each error bar indicates +- 1 standard error.

We collected data through surveys, chat-logs and game-play logs to assess (1) participants' partner interactions and perceptions thereof, (2) subjective and objective learning outcomes, as well as (3) the depth and speed of gameplay.

Measures of Partner Interactions: Participants' partner interactions and perceptions thereof were assessed by measuring collaborative engagement, interpersonal closeness, perceived helpfulness, and ignorant partner perception. *Collaborative Engagement* was measured by counting the number of lines each person entered into the chat window during the 25 minutes of playing the game. *Interpersonal Closeness* was measured through the inclusion of the other in the self (IOS) scale, a widely used self report measure to assess the perception of an interpersonal relationship [5]. To create a measure of *Perceived Helpfulness* of interaction, participants rated on two 7-point Likert type scales to what degree they agreed that the interaction helped them and let them learn more during the game (1 = strongly disagree, 7 = strongly agree). The questions formed a reliable scale (Cronbachs = .87) and were averaged to create a measure of the perceived helpfulness of the interaction. Finally, to generate a measure of *Ignorant Partner Perception*, participants rated on three 7-point Likert type scales to what degree they agreed with the following statements about their partner: "ignored me", "moved on before I was ready", and "moved too fast" (1 = strongly disagree, 7 = strongly agree). The questions formed a reliable scale (Cronbachs = .79) and were averaged to create a measure of *Ignorant Partner Perception*.

Measures of Learning Outcomes: Subjective and objective learning outcomes were assessed by comparing scores on a translation test before and after the game and by measuring language learning confidence. We measured *Language Learning Per Quest* by scoring one point for each correct answer on the language pretest and post-test, calculating the difference between scores, and dividing the score by the number of completed quests. Both pretest and post-test consisted of 15 vocabulary questions, 5 sentences to be translated from Japanese into English and 5 sentences to be translated from English to Japanese. We also calculated the number of words learned per quest to measure *Vocabulary Learning Per Quest*. Of the 15 words measured, 11 could only be learned through quests. We calculated how many of these words players learned during the game, and divided this by the number of quests completed. Finally, we measured *Language Learning Confidence* by asking participants to what degree they agreed with the following statement: "I am confident that I can learn the Japanese language well" (1 = strongly disagree, 7 = strongly agree).

Measures of Gameplay: Speed and Depth of gameplay was assessed by measuring the number of quests people solved in the given time (*Engagement Speed*) and the number of words participants collected on average per quest (*Engagement Depth*). Once a player has collected the words necessary to complete a task, they can choose to proceed to the next challenge or continue to explore the environment and collect more words. Therefore, we take the number of words collected for a given quest to measure depth of engagement with the available content.

Findings about Task Interdependence

We conducted 2-condition (task interdependence high vs. low) Mixed Model ANOVAs to perform our analyses. Mixed Model ANOVAs were used due to the hierarchical structure of our data. Participants were nested into pairs. The demographic variables (age and gender) and time spent playing video games were set as control variables in all models. Table 2 provides an overview about the dependent variables as well as correlations between variables. As a manipulation check we asked participants to rate on a 7-point scale from "strongly disagree" to "strongly agree" the degree to which they believed that their success in the game depended on their partner. The manipulation check indicated that our manipulation was successful as participants on average believed to depend on their partner more in the high interdependence (M = 2.98, SE = 0.28) than in the low interdependence condition (M = 2.04, SE = 0.27, F [1, 19.77] = 5.35, p = .032).

Overall we found evidence in support of H1a over H1b. With some limitations, as task interdependence increased we saw an improvement in collaborative language learning across all thee dimensions. In the following sections we report in more depth about our findings regarding the three dimensions of (1) participants' partner interactions and perceptions thereof, (2) subjective and objective learning outcomes, as well as (3) the depth and speed of gameplay.

Interdependence improved partner interactions

As shown in Figure 5, participants on average wrote about three times as many lines of chat text in the high interdependence (M = 12.71, SE = 2.35) than in the low interdependence condition (M = 4.05, SE = 2.25, F [1, 19.57] = 7.00, p = .017). This difference in *Collaborative Engagement* was significant. Changing the level of task interdependence also affected *Interpersonal Closeness* (F [1, 18.53] = 4.64, p = .045). Participants in the high interdependence condition (M = 3.05, SE = 0.35) on average felt closer to their partner than in the low interdependence (M = 1.98, SE = 0.33) condition. Increasing the level of task interdependence from low to high improved the *Perceived Helpfulness of Interaction* during the game (F [1, 19.59] = 9.51, p = .006). Participants in the high interdependence condition (M = 4.68, SE = 0.27) perceived the interaction on average as more helpful for learning than in the low interdependence condition (M = 3.46, SE = 0.26). Finally, increased task interdependence led to lower ratings of *Ignorant Partner Perceptions* in the high interdependence (M = 2.61, SE = 0.25) in comparison to the low interdependence condition (M = 3.66, SE = 0.25, F[1, 17.59] = 7.62, p = .01).

Taken together, we found strong evidence that changing the structure of the task by varying the level of interdependence positively affected the quality of the interpersonal interactions and the relationship between learners as participants (1) chatted significantly more with their partner, (2) developed a significantly closer relationship over the 25 minutes with their partner, and (3) found the interaction to be significantly more helpful for learning.

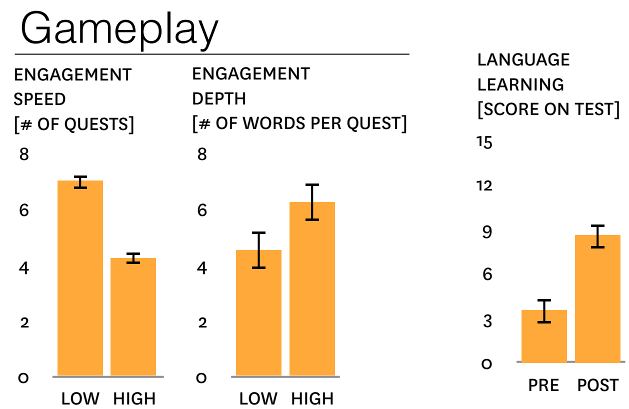

Figure 6. Gameplay and overall impact. Each chart shows the differences between the low and high interdependence conditions. The two charts on the left show that greater interdependence led to fewer quests being completed but more words being collected per quest (which indirectly measures how much players are exploring and thus engaging with the material). The chart on the right shows the overall learning gains from pretest to posttest, indicating that players learned an average of five words during the user study. The error bars indicate +- 1 standard error.

The more engaged interactions in the high interdependence conditions indicate that using gameplay mechanics to force collaboration can lead to better interactions. This finding is especially encouraging as many sources indicate that social ties are what motivate learners over longer periods of time.

Interdependence improved some learning outcomes

As shown in Figure 5, participants on average improved on the language test by almost one point per quest in the high interdependence (M = 1.66, SE = 0.30) in comparison to the low interdependence condition (M = 0.68, SE = 0.29, F [1, 12.19] = 5.39, p = .038. This difference in *Language Learning per Quest* was significant. When just considering the vocabulary learned and taking into account the words that were learned during the tutorial section we still found that learned more than two times as many words per quest in the high interdependence condition (M = 0.49, SE = 0.08) than in the low interdependence condition (M = 0.19, SE = 0.07, F [1, 20.41] = 8.23, p = .009). This difference in *Vocabulary Learning per Quest* was also significant. This means that although players in the high interdependence condition were exposed to less language content, they retained more of the content that they were exposed to.

Additionally we found a marginally significant difference in subjective learning outcomes as participants on average reported a higher Japanese *Language Learning Confidence* in the high interdependence (M = 4.60, SE = 0.33) than in the low interdependence condition (M = 3.64, SE = 0.32, F[1, 21.6] = 4.05, p = .06).

These findings suggest that increasing the level of task interdependence not only positively impacts the interpersonal interaction and relationship between partners but also objective and to some degree subjective language learning outcomes.

Variable	Partner Interactions 1	2	3	4	Learning Outcomes 5	6	7	Gameplay 8	9
Partner Interactions									
1. Collaborative engagement	–	0.53	0.40	-0.36*	0.56	0.60	0.19	-0.51*	0.48
2. Interpersonal closeness		–	0.45	-0.29*	0.14	0.18	0.08	-0.15*	0.15
3. Perceived helpfulness			–	-0.14*	0.10	0.16	0.21	-0.16*	0.09
4. Ignorant partner perception				–	-0.29*	-0.20*	-0.20*	0.18	-0.25*
Learning Outcomes									
5. Language learning per quest					–	0.84	0.23	-0.58*	0.65
6. Vocabulary learning per quest						–	0.29	-0.47*	0.44
7. Language learning confidence							–	-0.06*	0.17
Gameplay									
8. Engagement speed								–	-0.78*
9. Engagement depth									–
M	8.429	2.513	4.095	3.127	1.166	0.339	4.024	5.762	5.295
SD	8.577	1.282	1.453	1.341	1.048	0.304	1.490	2.721	2.659
Minimum	0.000	1.000	1.000	1.000	0.000	0.000	2.000	2.000	1.750
Maximum	39.000	7.000	6.500	6.667	5.000	1.330	7.000	14.000	12.500

Note. * $p < .01$ (All two-tailed tests.)

Table 2. Table of correlations.

Interdependence led to slower but more engaged gameplay

Engagement Speed: During the allotted time of 25 minutes, participants on average completed almost 3 quests less in the high interdependence (M = 4.25, SE = 0.51) than in the low interdependence condition (M = 7.00, SE = 0.50, F[1, 17.46] = 14.06, p = .002). However participants on average also collected more words per quest in the high interdependence (M = 6.29, SE = 0.73) than in the low interdependence condition (M = 4.53, SE = 0.70, F [1, 18.37] = 2.98, p = .10, a difference in *Engagement Depth* that was marginally significant. This indicates the possibility that while increased task interdependence led to slower gameplay, the depth of the engagement with the learning material increased. This explanation is also supported through our qualitative examination of the chat-logs. We found several instances of sentences level questions being discussed in the high interdependence conditions but this was rarely so in the low interdependence conditions. Discussing sentence structures over chat might have contributed to more reflection on the sentence and grammar structure, and therefore lead to better retention of vocabulary. Another explanation consistent with prior work by Okita and colleagues [41], is that when players see a correct sentence coming from another player, they are more likely pay attention to that sentence and learn it.

Chat log coding

To gain a more in depth understanding of the impact of the task interdependence manipulation on actual collaboration behavior we coded the complete set of chat logs according to the categories of Bales' widely adopted Interaction Process Analysis coding scheme [6]. A simplified description of the codes can be found in Table 3. Bales' categories were developed to analyze a group's problem solving process by distinguishing two general processes: Socio-emotional interaction behaviors (positive A; negative D) and task related interaction behaviors (Answers B; questions C). Three researchers coded the complete set of chat-logs independently using the four codes listed in table 3 and a fifth code we called "neutral" that didn't match any of the coding categories. One code was assigned for each line in the chat log. Analysis of inter-rater agreement revealed a Fleiss Kappa of 0.64, which constitutes a substantial level of agreement according to Landis and Koch [32].

For our quantitative analysis of the codes we excluded data from the chat logs that were generated during the tutorial segment as there were no differences between experimental conditions for the tutorial. To assess collaboration and helping behaviors we calculated the ratio of help-giving behaviors (code B) to help-seeking behaviors (code C). We found this ratio to be higher in the high interdependence logs (ratio: 3.7) than in the low interdependence logs (ratio: 1.6) indicating that our manipulation of task interdependence affects participants' collaborative behaviors during the game.

This increase in collaborative interactions and especially helping behaviors as task interdependence increased can be further illustrated through two transcripts shown below. The first chat-log excerpt from one of the groups in the high-interdependence conditions shows a mutual exchange of questions, suggestions, humor, and validation through 11 turns and over several minutes that was typical for several of the high-interdependence groups. In contrast, none of the interactions of the low-task interdependence groups showed interactions that went beyond three turns and the transcript below shows a typical exchange in which an interaction ends after an answer to a specific question has been given.

Code	Description
A	Positive Reaction: e.g. small talk, support
B	Task Attempted Answers: e.g. providing help
C	Task Questions: e.g. requesting help on the task
D	Negative Reaction: e.g. ignorance, rejection
N	Neutral: e.g. accidental key press, off-topic

Table 3. Coding scheme used for our analysis of the chat logs. Codes were derived by simplifying Bales 1950 [6] Interaction Process Analysis coding scheme.

High interdependence (P3 & P4)	Low interdependence (P43 & P44)
P3 12:02 *says I need to help you complete it too*	P43 6:05 *How do you say me too?*
P4 12:24 *ka goes last?*	P43 6:41 *Got it*
P3 12:34 *yea*	P44 6:43 *watashi mo*
P4 12:39 *hmm, i thought that was it*	P44 6:53 *nice!*
P3 12:40 *ka means a question*	
P3 12:49 *and gakusei is student*	
P3 12:56 *got it*	
P4 12:57 *ahh we both have to ask it*	
P4 13:04 *haha*	

Findings about the game as a whole

We also wanted to see how the game as a whole impacted learning and therefore conducted a t-test across both conditions comparing Japanese language skills as measured by the language pretest and posttest. We indeed found that participants performed significantly better on the Japanese language test after playing the game than before ($t(41) = 12.73$, $p < .01$). Participants on average scored 5 more points on the language test (out of available 25) after the game ($M = 8.38$, $SE = 1.07$) than before ($M = 3.28$, $SE = 1.12$), and looking at the individual results revealed that every participant gained at least one point.

Through our post-study interviews, we also gained insight into several aspects of gameplay. There was feedback from participants regarding the style of learning promoted by the game. The game encourages learning the meanings of words and phrases from context, and learning grammar implicitly. However, many participants requested features that would contradict this style of learning, such as explicit grammar learning and provided translations of sentences. We believe that these suggestions are largely due to prior experience in language courses. The question here is: should we compromise the learning style because of the prior learning preferences of the individual? Our research indicates that this question should be carefully considered when designing new language learning applications. If the methods are significantly different from the norm, we should consider educating users about the method before throwing them into it.

Nearly all of the participants requested a voice feature. Some theories of language learning indicate that recognition is an important intermediate step to production, which may indicate that voice is not as important as users believe for this stage of learning, but going forward we should seriously consider how and if voice should be implemented. If voice is not implemented, we should be careful to educate users on the rationale behind the exclusion of voice.

Limitations

Several limitations of our evaluation have to be considered. First, our findings are based on a very short (25 minutes) interaction with the game and it remains an open question how task interdependence affects long-term learning and motivations. Fortunately, our findings indicate that processes that are conducive to long-term maintenance of motivation in particular, such as relationship building, were impacted by changing task interdependence. Other researchers have found that these processes can improve long term learning outcomes. For example Krause et al. found that improving social interactions through games in MOOCs led to long term benefits in test scores and retention [29].

Another limitation of our study was the participant pool. We recruited students who required course credits or wanted to receive payments for an hour of participation in our study. Prior work found that the use of any incentives negatively impacts internal motivation [11]. Our use of incentives to participate in the study might have affect the generalizability of our findings to "out of lab" contexts. However, since all participants received the same incentives and the motivational influence is constant across conditions, this does not change the relevance our findings about the differences between the low and high task interdependence conditions. It remains an open question whether people who choose to play the game driven by the motivation to learn Japanese would have responded differently to our manipulation. However, given our encouraging findings even with participants who were not necessarily motivated to learn Japanese, we expect that the positive impact might even increase for students intrinsically motivated to learn Japanese. Future studies might also benefit from adding a measure of intrinsic motivation.

CONCLUSIONS

We designed *Crystallize* not only as a language learning platform but also as a research platform to study collaborative learning and uncover optimal design practices. Besides our findings on task interdependence and language learning, we demonstrated how *Crystallize* allows to examine the effects of varying game dynamics on language learning experience and outcomes. We showed how interface features can be manipulated, what measures that can be collected, and the kind of insights we can gain about collaborative language learning from systematically varying features of the game.

Reaching proficiency is a long process, and different aspects of the design may become important in a longer term study. For example, maintaining long term relationships between learners and sustaining motivation are both important considerations for designing for long-term language learning. Therefore, we plan to run a longitudinal study to examine how learning and perceptions of the game change over time.

Furthermore, we have only scratched the surface of leveraging situated learning for game-based language learning. There are still many aspects of real world interaction that are missing from *Crystallize*. In the future, we hope to design effective experiences that clearly demonstrate to learners not just *how* to say things in a foreign language, but *when and why* they should say them.

Finally, we hope to look more closely at how situated learning is qualitatively different from learning abstract linguistic constructs. There are many open questions in education about the definition of situated learning and when it should be used. Building on the design outlined in this paper, we believe we can answer some of these questions.

REFERENCES

1. Mark S Ackerman. 2000. The intellectual challenge of CSCW: the gap between social requirements and technical feasibility. *Human–Computer Interaction* 15, 2-3 (2000), 179–203.
2. Erik Andersen, Yun-En Liu, Richard Snider, Roy Szeto, and Zoran Popović. Placing a Value on Aesthetics in Online Casual Games. In *CHI 2011*. ACM Press.
3. Erik Andersen, Eleanor O'Rourke, Yun-En Liu, Richard Snider, Jeff Lowdermilk, David Truong, Seth Cooper, and Zoran Popović. The Impact of Tutorials on Games of Varying Complexity. In *CHI 2012*. ACM Press.
4. John R Anderson, Lynne M Reder, and Herbert A Simon. 1996. Situated learning and education. *Educational researcher* 25, 4 (1996), 5–11.
5. Aron E. N. & Smollan D. Aron, A. 1992. Inclusion of Other in the Self Scale and the structure of interpersonal closeness. *Journal of personality and social psychology* 63 (1992), 596–612. Issue 4.
6. Robert F Bales. 1950. A set of categories for the analysis of small group interaction. *American Sociological Review* (1950), 257–263.
7. Robert Bley-Vroman. 1989. What is the logical problem of foreign language learning. *Linguistic perspectives on second language acquisition* 4 (1989), 1–68.
8. Jeffrey L Crawford and Gordon A Haaland. 1972. Predecisional information seeking and subsequent conformity in the social influence process. *Journal of Personality and Social Psychology* 23, 1 (1972), 112.
9. Mihaly Csikszentmihalyi. 1990. *Flow: The Psychology of Optimal Experience*. Harper & Row Publishers, Inc., New York, NY, USA. 49 pages.
10. David Dearman and Khai Truong. Evaluating the Implicit Acquisition of Second Language Vocabulary Using a Live Wallpaper. In *CHI 2012*. ACM Press.
11. Edward L Deci, Richard Koestner, and Richard M Ryan. 2001. Extrinsic rewards and intrinsic motivation in education: Reconsidered once again. *Review of educational research* 71, 1 (2001), 1–27.
12. Paul Denny. The Effect of Virtual Achievements on Student Engagement. In *CHI 2013*. ACM Press.
13. Zoltán Dörnyei. 1997. Psychological processes in cooperative language learning: Group dynamics and motivation. *The Modern Language Journal* 81, 4 (1997), 482–493.
14. Nicolas Ducheneaut and Robert J Moore. The social side of gaming: a study of interaction patterns in a massively multiplayer online game. In *CSCW 2004*. ACM Press.
15. Nicolas Ducheneaut, Nicholas Yee, Eric Nickell, and Robert J Moore. Alone together?: exploring the social dynamics of massively multiplayer online games. In *CHI 2006*. ACM Press.
16. Darren Edge, Kai-Yin Cheng, Michael Whitney, Yao Qian, Zhijie Yan, and Frank Soong. Tip Tap Tones: Mobile Microtraining of Mandarin Sounds. In *MobileHCI 2012*. ACM.
17. Darren Edge, Stephen Fitchett, Michael Whitney, and James Landay. MemReflex: Adaptive Flashcards for Mobile Microlearning. In *MobileHCI 2012*. ACM.
18. Darren Edge, Elly Searle, Kevin Chiu, Jing Zhao, and James A Landay. MicroMandarin: mobile language learning in context. In *CHI 2011*. ACM Press.
19. Alan Firth and Johannes Wagner. 1998. SLA property: No trespassing! *The Modern Language Journal* 82, 1 (1998), 91–94.
20. Cyrus K Foroughi, Nicole E Werner, Erik T Nelson, and Deborah A Boehm-Davis. 2014. Do interruptions affect quality of work? *Human Factors: The Journal of the Human Factors and Ergonomics Society* 56, 7 (2014), 1262–1271.
21. James Paul Gee. 2001. Reading as situated language: A sociocognitive perspective. *Journal of adolescent & adult Literacy* (2001), 714–725.
22. Maribeth Gettinger. 1985. Time allocated and time spent relative to time needed for learning as determinants of achievement. *Journal of Educational Psychology* 77, 1 (1985), 3.
23. Robert Howland, Sachi Urano, and Junichi Hoshino. 2012. SanjigenJiten: computer assisted language learning system within a 3d game environment. In *Advances in Computer Entertainment*. Springer, 262–273.
24. Daniel Johnson, Peta Wyeth, Madison Clark, and Christopher Watling. Cooperative Game Play with Avatars and Agents: Differences in Brain Activity and the Experience of Play. In *CHI 2015*. ACM Press.
25. David W Johnson. 1973. Communication in conflict situations: A critical review of the research. *International Journal of Group Tensions* 3, 1 (1973), 46–67.
26. David W Johnson, Roger T Johnson, and Mary Beth Stanne. 2000. Cooperative learning methods: A meta-analysis. (2000).
27. René F. Kizilcec and Emily Schneider. 2015. Motivation As a Lens to Understand Online Learners: Toward Data-Driven Design with the OLEI Scale. *ACM ToCHI* 22, 2, Article 6 (March 2015).
28. Raph Koster. 2004. *Theory of fun for game design*. " O'Reilly Media, Inc.".

29. Markus Krause, Marc Mogalle, Henning Pohl, and Joseph Jay Williams. A Playful Game Changer: Fostering Student Retention in Online Education with Social Gamification. In *Learning @ Scale 2015*. ACM Press.

30. Chinmay Kulkarni, Julia Cambre, Yasmine Kotturi, Michael S Bernstein, and Scott Klemmer. Talkabout: Making Distance Matter with Small Groups in Massive Classes. In *CSCW 2015*.

31. Marjan Laal and Seyed Mohammad Ghodsi. 2012. Benefits of collaborative learning. *Procedia-Social and Behavioral Sciences* 31 (2012), 486–490.

32. J Richard Landis and Gary G Koch. 1977. The measurement of observer agreement for categorical data. *biometrics* (1977), 159–174.

33. Diane Larsen-Freeman. 2007. Reflecting on the cognitive–social debate in second language acquisition. *The Modern Language Journal* 91, s1 (2007), 773–787.

34. Martha Larson, Nitendra Rajput, Abhigyan Singh, and Saurabh Srivastava. I want to be Sachin Tendulkar!: a spoken english cricket game for rural students. In *CSCW 2013*. ACM Press.

35. Jonas Linderoth. 2005. Animated game pieces. Avatars as roles, tools and props. In *Aesthetics of Play Conference Proceedings*.

36. Derek Lomas, Kishan Patel, Jodi L. Forlizzi, and Kenneth R. Koedinger. Optimizing challenge in an educational game using large-scale design experiments. In *CHI 2013*.

37. Joseph Edward McGrath. 1984. *Groups: Interaction and performance*. Vol. 14.

38. Luis C. Moll. 1992. *Vygotsky and education: Instructional implications and applications of sociohistorical psychology*. Cambridge University Press.

39. Scott Nicholson. 2015. A recipe for meaningful gamification. In *Gamification in Education and Business*. Springer, 1–20.

40. David Nunan. 1992. *Collaborative language learning and teaching*. Cambridge University Press.

41. Sandra Y Okita, Jeremy Bailenson, and Daniel L Schwartz. 2007. The mere belief of social interaction improves learning. In *Proceedings of the Twenty-ninth Meeting of the Cognitive Science Society*.

42. Anne Marie Piper, Eileen O'Brien, Meredith Ringel Morris, and Terry Winograd. SIDES: a cooperative tabletop computer game for social skills development. In *CSCW 2006*. ACM Press.

43. Alexander Renkl, Robert K Atkinson, Uwe H Maier, and Richard Staley. 2002. From example study to problem solving: Smooth transitions help learning. *The Journal of Experimental Education* 70, 4 (2002), 293–315.

44. Jeremy Roschelle and Stephanie D Teasley. 1995. The construction of shared knowledge in collaborative problem solving. In *Computer supported collaborative learning*. Springer, 69–97.

45. Richard Saavedra, P Christopher Earley, and Linn Van Dyne. 1993. Complex interdependence in task-performing groups. *Journal of applied psychology* 78, 1 (1993), 61.

46. Chantal Savelsbergh, P Storm, and Ben Kuipers. 2008. Do Leadership behavior, Team stability and Task interdependence relate to Team learning? *Proceedings of the WAOP* (2008), 1–26.

47. Vivian I Schneider, Alice F Healy, and Lyle E Bourne Jr. 2002. What is learned under difficult conditions is hard to forget: Contextual interference effects in foreign vocabulary acquisition, retention, and transfer. *Journal of Memory and Language* 46, 2 (2002), 419–440.

48. Mark K Singley, Moninder Singh, Peter Fairweather, Robert Farrell, and Steven Swerling. Algebra jam: supporting teamwork and managing roles in a collaborative learning environment. In *CSCW 2000*. ACM Press.

49. Leonard Springer, Mary Elizabeth Stanne, and Samuel S Donovan. 1999. Effects of small-group learning on undergraduates in science, mathematics, engineering, and technology: A meta-analysis. *Review of educational research* 69, 1 (1999), 21–51.

50. Susan G. Straus. 1999. Testing a Typology of Tasks - An Empirical Validation of McGrath's (1984) Group Task Circumplex. *Small Group Research* 30, 2 (1999), 166–187.

51. Zachary O Toups, Andruid Kerne, and William A Hamilton. 2011. The Team Coordination Game: Zero-fidelity simulation abstracted from fire emergency response practice. *ToCHI* 18, 4 (2011), 23.

52. Endel Tulving and Donald M Thomson. 1973. Encoding specificity and retrieval processes in episodic memory. *Psychological review* 80, 5 (1973), 352.

53. Luis von Ahn. 2013. Duolingo: learn a language for free while helping to translate the web. In *IUI'13*. ACM, 1–2.

54. Nick Yee, Jeremy N Bailenson, Mark Urbanek, Francis Chang, and Dan Merget. 2007. The unbearable likeness of being digital: The persistence of nonverbal social norms in online virtual environments. *CyberPsychology & Behavior* 10, 1 (2007), 115–121.