

**When Learning Together Goes Wrong:  
Negative Peer Effects in Online Learning**

Shohei Yamamoto  
Shuma Iwatani  
Koki Shimazu

April 2024

The Institute of Social and Economic Research  
Osaka University  
6-1 Mihogaoka, Ibaraki, Osaka 567-0047, Japan

**When Learning Together Goes Wrong: Negative Peer Effects in Online Learning**

Shohei Yamamoto<sup>1,\*</sup>, Shuma Iwatani<sup>2</sup>, Koki Shimazu<sup>3</sup>

<sup>1</sup> Graduate School of International Corporate Strategy (ICS), Hitotsubashi University, 2-1-2 Hitotsubashi, Chiyoda-ku, Tokyo 101-8439, Japan. Email: [syamamoto@ics.hub.hit-u.ac.jp](mailto:syamamoto@ics.hub.hit-u.ac.jp)

Phone: +81342123004 Fax: +81342123006

\*Corresponding author

<sup>2</sup> Faculty of Sociology, Kwansei Gakuin University, 1-155 Uegaharaichibancho, Nishinomiya, Hyogo, Japan 662-0891, Email: [iwatani.shuma@kwansei.ac.jp](mailto:iwatani.shuma@kwansei.ac.jp)

<sup>3</sup> Taktoia&Co., Daishin Building 501, 2-9 Kanda Nishikicho, Chiyoda-ku, Tokyo 101-0054, Japan. Email: [kokishimazu@taktopia.com](mailto:kokishimazu@taktopia.com)

**Acknowledgements:** The authors thank Atsushi Mizumoto, Charles Noussair, Daniel Banki, Daniel Navarro-Martinez, Daniel Read, Elif Tosun, Jordi Quoidbach, Noboyuki Hanaki, Takeshi Nishimura, and Yukihiko Funaki for helpful comments and suggestions. This work was supported by the Joint Usage/Research Center, the Institute of Social and Economic Research (ISER), Japan Society for the Promotion of Science (grant 21K13257) and Sasakawa Scientific Research Grant (grant 2022-1032).

**Declarations of interest:** None.

**When Learning Together Goes Wrong: Negative Peer Effects in Online Learning****Abstract**

This research examined the impacts of peer skill levels and perseverance through two experiments resembling online learning platforms. Study 1 recruited current English learners, while Study 2 involved participants who had not engaged in studying for more than six months. The results in both experiments revealed negative rather than positive peer effects. The participants ceased studying earlier and displayed reduced performance when learning with peers possessing lower perseverance, compared to when studying alone. This pattern was observed for similarly-skilled peers in Study 1 and higher-skilled peers in Study 2. Further analysis indicated that the negative peer effects predominantly originated from participants with lower levels of motivation. Additionally, it was shown that social proximity could foster positive effects when peers possess similar skills and higher perseverance levels. Our findings suggest that the strategic pairing of learners with appropriate partners is crucial for diminishing negative peer effects and enhancing positive peer influences.

*Keywords:* peer effects, perseverance, performance, online learning

### **When Learning Together Goes Wrong: Negative Peer Effects in Online Learning**

The advent of the digital age has revolutionized various sectors, with the education industry being a prominent beneficiary. The emergence and subsequent ubiquity of online learning platforms have provided an opportunity to transcend geographical boundaries and time constraints, offering a new dimension to the traditional learning methods. These platforms catering to a vast array of learning needs ranging from programming (e.g., Codecademy) and language (e.g., Duolingo) to obtaining certifications such as CFP (e.g., Udemy), have experienced explosive growth in recent years.

The online learning market, known for its limitless potential and extensive reach, has been expanding at an unprecedented rate. As of 2022, the industry was valued at \$198.2 billion, and is projected to reach a value of 602.0 billion by 2030, according to the market research company Vantage Market Research, thereby reinforcing its growing popularity and acceptance worldwide. The rising number of users engaging in these platforms attests to this growth. For instance, language learning platform Duolingo reported 49.5 million monthly active users in 2022.

Despite the surging popularity and the seemingly limitless opportunities offered by online learning platforms such as Coursera, Skillshare, LinkedIn Learning, as well as massive open online courses, a recurring issue that plagues the sector is the high attrition rate (Eriksson et al., 2017; Narayanasamy & Elçi, 2020). A considerable portion of their users tend to quit midway. Indeed, research suggests that student attrition rates on online courses are generally higher than those in traditional classroom-based courses (Levy, 2007; Tello, 2007).

The trend of users abandoning their courses before completion undermines the potential of these platforms and poses a significant challenge to realizing the full potential of online learning. To tackle the problem, some platforms have introduced social features for learners to engage more. As an example, Duolingo set up a feature of “leaderboards”; a function of monitoring how much

learners progress each week, compared to other learners around the world. The users can ascend the leaderboard through active engagement with the platform and can gauge their standings relative to users they follow. Duolingo claims that learners who follow other users on Duolingo are 5.6 times more likely to finish their course (Zabell, 2023). Another intriguing feature to enhance users' online learning experience is "Friends Quest." Every week, users are randomly paired up with one of their Duolingo friends. They are assigned a challenge such as completing a certain number of lessons. Once they complete the challenge in five days, they can win a reward in Duolingo.

As illustrated in the examples above, online learning platforms have been striving to strengthen the engagement of users by encouraging the users to interact with other users. Numerous previous papers have demonstrated that productivity or duration of work are positively influenced by the presence of peers or collaboration with others. Particularly, collaboration and monitoring appear to contribute positive peer effects from the results of several research pieces. Kandel & Lazear (1992) and Baron & Kreps (1999) found that mutual monitoring and peer pressure can lead to increased productivity and Falk & Ichino (2006) have demonstrated that individuals tend to spend more time on tasks when working in pairs than when working alone, highlighting the potential role of peer effects in improving work engagement. Furthermore, Mas & Moretti (2009) found beneficial productivity spillovers when highly productive staff members were incorporated into a work shift.

One critical factor influencing peer effects seems to be the abilities of the peers. Some research suggest that better peers can indeed be beneficial (Hoxby & Weingarth, 2006). For example, being observed by higher-skilled peers — quantified in terms of the average number of items scanned per second over a ten-minute period — can increase workers' productivity due to concerns about self-image (Mas & Moretti, 2009; Bursztyn & Jensen, 2017).

On the contrary, other studies have demonstrated that having higher-achieving peers can sometimes have no influence or even be detrimental. For instance, Van Veldhuizen, Oosterbeek and

Sonnemans (2018) found that being observed by a more productive peer did not increase individual's output. Additionally, Feld and Zöllitz (2017) noted that low-achieving students might experience adverse effects when in the company of high-achieving peers. Gill & Prowse (2012) found that individuals are likely to decrease their effort when faced with a competitor's high performance. This decrease is attributed to disappointment aversion, where agents experience psychological loss due to deviations from their expectations.

Adding to the complexity of the matter, a review conducted by Sacerdote (2014) found that roughly half of the analyses did not uncover any statistically significant peer effects from classmates' background ability. This suggests that the influence of peer ability on peer effects is a topic that warrants further exploration and nuanced understanding.

Another important determinant of peer effects appears to be the level of peers' perseverance. Research indicates that students who are grouped with more persistent peers achieve higher scores in exams (Golsteyn et al., 2021). Similarly, Gerhards & Gravert (2020) found observing peer behavior significantly enhanced participants' perseverance in a task. Another study has showed that positive peer effects are observed when paired with a peer demonstrating high levels of self-control, while negative peer effects manifest when paired with a peer exhibiting low self-control (Battaglini et al., 2017). Furthermore, sharing information of peers' exerting perseverance can contribute positive peer effects (Buechel et al., 2014). Buechel et al. (2018) examined the influence of peer's perseverance on performance in a controlled laboratory setting. The results suggested that the magnitude of peer effects was significantly influenced by how successfully achieving individuals communicated their accomplishments. Positive outcomes were noted when these individuals conveyed the potential success of others, whereas negative outcomes emerged when they framed their success as exceptional.

Drawing on the aforementioned studies, our study aimed to explore the impact of peer skill levels and perseverance on peer effects, and to examine how these two factors jointly influence learner performance and perseverance. Our experiment was designed to mirror realistic online learning environments possessing intrinsic value, while controlling for other variables. Notably, to the best of our knowledge, no studies have yet explored these combined effects on learning outcomes.

We formulate hypotheses regarding the influence of skill levels among peers, specifically how peers of similar and high skill levels affect behavior. First of all, we consider the situation when individuals possess skill levels similar to those of their peers. We hypothesize that, in this situation, the duration of time committed to studying is likely to be more closely aligned with that of their peers, as opposed to when studying is undertaken independently. As a result, the impact of these peers' level of perseverance becomes more pronounced. This hypothesis finds its grounding in the findings of Rosaz et al. (2016), who observed that the quitting time for tasks tends to align more closely with that of their partners through communication.

Rosaz et al. (2016) posited that this effect is due to a reduction in social distances between the partners. This concept of reduced social distances may also elucidate the findings of Bandiera et al. (2010). Their study revealed that less proficient workers were more productive when a friend was working nearby, whereas no such productivity boost was observed when the nearby individual was not a friend. This social closeness appears to be associated with the pairing of peers with similar skills, as evidenced by that students with similar skill levels are significantly more likely to choose similar others as friends and advisors (Lomi et al., 2011).

We further explored the peer effect in situations where partners possess higher skill levels, noting that the outcomes could be either positive or negative, as the existing literature does not offer clear predictions. Hoxby and Weingarth (2005) demonstrated the positive peer effects from high-skilled peers, implying that one's study time may more closely align with that of these peers. Additionally,

research indicates that being observed by a higher-skilled peer increases worker productivity due to concerns about self-image (Mas & Moretti, 2009; Bursztyn & Jensen, 2017). In contrast, Villeval (2020) posited that negative peer effects could occur when individuals perceive their lower-than-expected skill level from their high skilled partners, which could foster feelings of demoralization or shame.

We recruited Japanese residents and conducted novel experiments. Unlike several previous studies that experimentally measured peer effects using real effort tasks with little or no intrinsic value (e.g., Beugnot et al., 2019; Buechel et al., 2014, 2018; Georganas et al., 2015; Gerhards & Gravert, 2020), we provided participants with the opportunities to learn English vocabulary online which resemble a typical English online learning environment. Japan serves as a suitable context for this research due to its large market for English language education.<sup>1</sup> We chose English vocabulary as the focus of the experiment because of its likely intrinsic value to such Japanese participants and its ease of implementation. In the experiment, we compared the amount of effort exerted in studying and test scores between the participants studying alone (single condition) and the others studying with a partner (pair condition). This approach allowed us to gauge the influence of peer effects on participants' perseverance (the amount of effort) as well as their performance. Participants were paired with pre-recruited learning partners. The partners' skill levels and perseverance (defined as the duration for which they continue to learn English vocabulary) were measured prior to our experiments. Based on their pre-recorded outputs, we manipulated partners' levels of skills and perseverance, rendering these variables exogenous; thus, we can clearly estimate their impacts for

---

<sup>1</sup> The immense interest in English proficiency within the country is evident in the participation rates for the Eiken (an English proficiency exam inaugurated in 1963 within the country) and Test of English for International Communication (TOEIC). In 2022 alone, over 4 million individuals took the Eiken test, while more than 2 million participated in the TOEIC exam (including “TOEIC Listening & Reading Test” and “TOEIC Speaking & Writing Tests”), figures that underscore the significant demand for English language proficiency in Japan.

each effect and the interaction effect. Consequently, such online learning configurations offer considerable practical implications because our design resembles many online learning platforms. Our novel design, which can be applied to many other online learning studies, enables us to clearly estimate causal effects from peers.

We conducted two studies. In Study 1, we recruited participants who were currently studying English. In Study 2, we recruited participants who had not engaged in studying more than six months. The participants in Study 2 were less-skilled and did not enjoy studying compared to the participants in Study 1. In both experiments, we observed persistent negative peer effects stemming from partners with lower levels of perseverance. If partners abandoned learning midway, both the number of words studied and the test scores were lower compared to the outcomes when studying independently. This pattern held true when participants were paired with peers of similar skill levels (Study 1) and with higher-skilled peers (Study 2). Interestingly, when segmenting the sample based on the level of intrinsic motivation for learning the English language, the findings indicated that the negative peer effects predominantly originated from participants with lower levels of motivation.

Less intrinsically motivated participants suffered the negative peer effects to a greater degree. Although we did not find overall positive peer effects, closer psychological distances were positively correlated with the number of studied words and the test scores if the partner had similar skills and higher perseverance.

### **Study 1: Influence of Partner Skills and Perseverance on English Vocabulary Learning**

Study 1 was designed to test peer effects from learning partners on an English vocabulary learning setting. We investigated how the levels of skills and perseverance of learning partners influenced the peer effects.

Before conducting the main study, we separately conducted a study to recruit potential partners who were matched with the participants in the main study. We recruited four potential partners,

each possessing different characteristic factors: two distinguished by gender (male and female) and two distinguished by their level of English skills (high and middle).<sup>2</sup> We asked them to study a total of 100 English words. Additionally, we instructed them to send emojis from time to time to other participants to motivate their learning. We used the data of high-skilled female and middle-skilled female as potential partners in the main studies.<sup>3</sup>

## Method

**Participants.** In Study 1, we recruited 400 participants through Lancers, Inc. (<https://www.lancers.jp>). We excluded 26 participants who claimed to know more than 400 words in the pre-learning part, which we will detail later, as their vocabulary level exceeded the scope of this experiment. Consequently, we obtained a demographically diverse sample of 373 participants in total (48% female,  $M_{age} = 38.8$  years, age range: 18-72 years). Of these participants, 52% are fully employed, 14% hold part-time positions, 5% are students, and the remaining 29% fall into other categories. The criteria of our participants were as follows; (1) Japanese residents, (2) currently studying English and (3) having English skills less than the C1 level (less than a TOEIC score of 950). The participants received a fixed fee of 600 yen. The median duration of the experiment was around 48 minutes, but it varied depending on how many English words they wanted to learn.

**Design and procedure.** The participants were assigned to pair or single conditions. The participants in the pair condition had learning partners (whom we had recruited before conducting this experiment as mentioned above), whereas the participants in the single condition did not have

---

<sup>2</sup> Test of English for International Communication (TOEIC) score of 955 is considered as high English skills and middle TOEIC score of 615 is considered as middle English skills

<sup>3</sup> We used female as potential partners in this experiment. This is because participants felt their peers more beneficial in learning English when paired with female partners, compared to when paired with male partners, according to the questionnaire in the pretest.

any partners. Furthermore, the characteristics of partners were manipulated in the pair condition: English skills (high/similar to the participant), and their study time (quitter/finisher). Therefore, we have 5 conditions in total.

Participants took part in the five parts of the experiment: pre-learning part, pre-study questionnaire and pairing announcement part, mandatory learning part, voluntary learning part and test part (see Table 1 for a summary of the procedure). In the pre-learning part, participants were presented with English words and they simply answered whether they knew the words (they chose the option between “I know the word” and “I did not know the word at all”). They were instructed to choose “I know the word” even if they had some familiarity with the word). We selected 500 English words for the participants from a vocabulary book (Shimazu, 2021). The level of the selected words ranged from B1 to C2 levels in Common European Framework of Reference for Languages.

After the number of words that they did not know reached 100, they proceeded to the pre-study questionnaire and pairing announcement part. Here, they completed a demographic survey that inquired about various attributes, including age, gender, education, marital status, and employment status. The survey also asked about their primary motivations for learning English as well as their self-assessed level of English proficiency.<sup>4</sup> Additionally, they were instructed to choose their own avatar like the ones in typical smartphone apps.

Insert Table 1 about here.

Immediately following the questions, participants in the pair conditions were notified that they were paired with their learning partner (such information was not provided to the participants in the

---

<sup>4</sup> The options for the reasons for learning English include: for work, for travel, for communicate with international people, to help foreigners in need, to gain common sense, because it is cool, to enjoy foreign movies and other media, for the sake of children, and others. The English proficiency levels range from beginner (1) to expert (5). The participants who selected 5 in this question were not allowed to continue the experiment as per the instructions.

single condition). They were also informed of the partner's characteristics, including gender, name (comprised of randomly generated characters), and English proficiency skills (See Figure 1 for the screenshot). At this point, we implemented our first manipulation of the partners' skill while avoiding deception problems. Participants in the similar-skilled condition were informed that the partners answered the same number of words correctly as they did for the 10 questions in the pre-learning part. On the other hand, participants in the high-skilled condition were informed that the partners answered 3 words more correctly than they did for the 10 questions. We deliberately picked 10 questions in the pre-learning part. For example, we selected 10 questions where the participants in the high-skilled condition correctly answered 5 questions, and 10 questions where their partner correctly answered 8 questions.

Insert Figure 1 about here.

Subsequently, they proceeded to the mandatory learning part. Following the methodology in Rosaz et al. (2016), this part was designed to let participants to familiarize the procedure, as well as the level of difficulty and duration of our experimental task. Here, they must learn 5 words out of 100 words that they answered "I did not know the word at all" in the pre-learning part. In this part, questions were presented wherein an English word appeared above five Japanese translation options (Figure 2).<sup>5</sup> Participants were asked to select one of the options that correctly translated the given English word within 10 seconds. After that, the correct answers were shown and they were asked to remember it for 5 seconds. This process was repeated five times, resulting in the study of five words in this part. There was no break while doing this part. If they did not choose any option within the time period, they had a second chance to answer the same question within 10 seconds. Failure to

---

<sup>5</sup> Multiple-choice questions are a prevalent method for assessing vocabulary proficiency. Presenting answer choices in the first language, which is Japanese in our study, is efficient (Nation, 2001).

choose an option during this additional opportunity resulted in the termination of their participation in the study.

Upon completing the mandatory learning part, they went to the voluntary learning part. The procedure here was akin to the mandatory learning part, but with notable distinctions. Specifically, participants had the option to cease their learning at any point after answering one question. They could study up to 95 additional words. To ensure comprehension of these conditions, participants were required to answer a comprehension question, clarifying their understanding that they could discontinue the learning task at their discretion after answering the first question. After they correctly answered this question, they were allowed to start learning.

In this part, the procedure differed depending on whether participants were in the single or paired conditions. Those in the single condition continued their learning alone, as in the mandatory section. On the other hand, participants in the paired condition engaged in learning with information of their partners. They received emojis from their partner (refer to Figure 2) and had the opportunity to send emojis to their partner. This procedure aimed to foster a sense of learning with their partners.

At this stage, we introduced the second manipulation concerning partners' perseverance level. Participants in the quitter condition were notified that their partner had ceased studying after six questions. Conversely, those in the finisher condition were informed that the partner continued to study questions.

After the voluntary learning part, participants transitioned to the test part. This part consisted of a vocabulary test featuring 30 words that participants had previously identified as "I did not know the word at all" in the pre-learning part. Following the test, participants in the paired conditions responded to survey questions concerning their perceptions of their partners. The survey questions encompassed several aspects by asking following questions: How was the amount of words learned

by the partner (1: very little – 10: very much)?; How do you think your partner's level of English vocabulary compares to yours? (1: much lower – 7: much higher)?; How did the amount of words learned by the partner's influence the amount of words learned by you (1: decreased a lot – 7: increased a lot)?; Do you think it is beneficial to have the partner on your learning (0: not beneficial at all – 10: very beneficial)? and Inclusion of Other in Self Scale (IOS; Aron et al., 1992) measuring how close the participants felt with their partner (1: no overlap – 7: most overlap). Lastly, we asked how fun learning English words in this application was (with an 11-point scale).

Insert Figure 2 about here.

## Results

Table 2 shows the selected characteristics of the participants for each treatment. The variable *known\_words* represents the number of words participants knew in the pre-learning part, *eng\_level* denotes their self-assessed level of English proficiency, and *female* indicates the proportion of female participants. The variables between the treatments are not statistically significant indicating that the randomization works well (all p-values above 0.1).

Insert Table 2 about here.

The manipulations regarding the skills and perseverance of the partners in this study were successful. Participants in the high-skilled condition believed their partner had significantly higher vocabulary levels ( $M = 5.39$ ,  $SD = 1.24$ ) compared to those in the similar-skilled partners ( $M = 4.25$ ,  $SD = 1.02$ ;  $F_{(1,301)} = 76.50$ ,  $p < .001$ ,  $\eta^2 = .20$ ). Participants in the finisher conditions perceived their partner studied more ( $M = 7.02$ ,  $SD = 2.26$ ) compared to those in the quitter conditions ( $M = 3.68$ ,  $SD = 2.42$ ;  $F_{(1,301)} = 153.77$ ,  $p < .001$ ,  $\eta^2 = .34$ ).

### ***Graphical analysis with ANOVA***

The top of Figure 3 illustrates that the number of studied words in the voluntarily part across different conditions. Participants seem to have their inherent motivation to engage with English

vocabulary learning. They learned 38.31 words on average in the voluntary learning part, which was significantly larger than 1 ( $t(372) = 18.67, p < .001$ ). In addition, the average enjoyment rating for this experiment was 7.52 (on a scale from 0 to 10), suggesting a high level of enjoyment.

Next, we found the main effect of experimental condition on the number of their studied words was significant (analysis of variance (ANOVA):  $F_{(4,368)} = 3.04, p = .02, \eta^2 = .03$ ). We further conducted multiple comparisons with Holm's sequentially rejective Bonferroni procedure. As shown in Figure 3, there were negative effects from the partners who abandoned their learning efforts in the voluntary learning part midway (quitters). Participants in the similar-quitter condition ( $M = 25.32, SD = 32.13$ ) studied a smaller number of words compared to those in the single condition ( $M = 42.36, SD = 39.37$ ), indicating the negative peer effect ( $t(368) = 2.68, p = .05$ ). The number in the similar-quitter condition was also significantly lower than the similar-finisher ( $M = 43.33, SD = 40.62; t(368) = 2.91, p = .04$ ) and high-finisher conditions ( $M = 43.62, SD = 40.45; t(368) = 2.88, p = .04$ ). There was no statistically significant difference between those in single and high-quitter condition ( $M = 37.08, SD = 37.72; t(368) = 0.84, p > .99$ ). On the other hand, positive effects stemming from peers who completed the entire learning tasks were not found; the number in the single condition was not different from that in either the similar-finisher ( $t(368) = 0.16, p > .99$ ) or high-finisher condition ( $t(368) = 0.20, p > .99$ ).

The bottom of Figure 3 displays that the number of correct answers in the test part across different conditions. The primary outcomes mirrored those of the analysis pertaining to the number of studied words. The results of ANOVA revealed the statistically significant main effect of experimental condition on the number of their correct answers ( $F_{(4,368)} = 3.86, p = .004, \eta^2 = .04$ ). Negative peer effects were observed, while positive peer effects were not. Specifically, the similar-quitter condition yielded a lower average number of correct answers ( $M = 16.66, SD = 5.71$ ) compared to the single condition ( $M = 19.59, SD = 5.65; t(368) = 2.88, p = .03$ ), although no

significant differences were found between the high-quitter ( $M = 18.92$ ,  $SD = 6.66$ ) and single conditions ( $t(368) = 0.66$ ,  $p > .99$ ). This pattern of results confirms that negative peer effects emanate exclusively from similarly skilled partners. Additionally, a notable difference was also observed between the similar-quitter and high-finisher conditions ( $t(368) = 3.74$ ,  $p = .002$ ), implying that participants paired with quitters performed worse than those paired with finishers. Further, the absence of positive peer effects was confirmed, as the number of correct answers in the similar-finisher condition ( $M = 18.80$ ,  $SD = 6.39$ ) and in the high-finisher condition ( $M = 20.45$ ,  $SD = 5.92$ ) did not statistically differ from the single condition ( $t(368) = 0.79$ ,  $p > .99$ ;  $t(368) = 0.84$ ,  $p > .99$ )

Insert Figure 3 about here.

### ***Regression analysis***

Next, we conducted regression analysis to further confirm the above analysis. First, we conducted an OLS regression of the number of words studied on treatment dummy variables, using the single condition as the baseline (see column 1 at the top of Table 3). Second, we incorporated demographic variables such as age, gender, and college—defined by a dummy variable that takes the value of 1 if the participant has a college degree—as well as a variable *known\_words* for the number of known words in the pre-learning part (see column 2). We repeated the analysis using Tobit regression to account for the constraint that participants could not study fewer than 1 or more than 95 words and correctly answer fewer than 0 or more than 30 words (see columns 3 and 4). The results remained consistent with the prior analyses conducted above, revealing that the similar-quitter condition had a significantly negative effect across all four regression models. The other treatment dummy variables of finishers had no significant impacts, indicating that the partners with higher perseverance had no impact on peer effects.

We also conducted the same regression analysis regarding the number of correct words in the test part (the bottom of Table 3). Most of the main findings were same as the results of former analysis regarding the number of studied words. The similar-quitter condition had a significantly negative effect on the number across the four regressions.

Insert Table 3 about here.

In additional analysis, we examined the correlation between learning outcomes and IOS, describing the psychological distance between the participants and their partners. We expected the positive correlation in cases where the partners completed the tasks (finishers) and a negative correlation when partners quit the tasks midway (quitters) because larger peer effects were found from people with closer social distance in the previous literature (e.g., Bandiera et al., 2010).

We found a moderate level of correlation between the number of studied words and IOS:  $r = -0.26$  ( $p = .03$ ) in the similar-quitter condition;  $r = -0.22$  ( $p = .05$ ) in the high-quitter condition;  $r = 0.21$  ( $p = .06$ ) in the similar-finisher condition; and  $r = -0.15$  ( $p = .20$ ) in the high-finisher condition. A similar correlation was observed between the number of correct words and IOS.<sup>6</sup> We regressed the number of studied words on the treatment dummy variables (the similar-quitter condition as the baseline treatment), IOS and their interaction variables.

The results, presented in the top of Table 4, consistently show significant positive coefficients for the similar-finisher and IOS interaction term, indicating a more positive correlation between IOS and the number of studied words in the similar-finisher condition compared to the similar-quitter condition. These results held across four regression models: without control variables (1st column), with control variables and the variable *known\_words* (2nd column), Tobit regression without

---

<sup>6</sup> The results are as follows: In the similar-quitter condition,  $r = -0.21$  ( $p = .07$ ); in the high-quitter condition,  $r = -0.14$  ( $p = .20$ ); in the similar-finisher condition,  $r = 0.15$  ( $p = .19$ ); and in the high-finisher condition,  $r = -0.11$  ( $p = .37$ ).

control variables (3rd column), and Tobit regression with control variables and the variable *known\_words* (4th column).

The bottom of Table 4 shows the same regressions in terms of the number of correct words. Again, we found the significant positive coefficients of the similar-finisher and IOS interaction term in the regressions.

### Discussion

In summary, our findings decisively demonstrate a negative impact from partners with low levels of perseverance, particularly when the skill levels were similar to the participants. This negative influence was not evident from higher-skilled partners. Contrarily, we found no evidence of the positive effects from partners with higher perseverance levels.

Our analysis regarding IOS suggests that lower social proximity accounts for the observed overall negative peer effects. Prior research has investigated the relationship between peer effects and social distance (Bandiera et al., 2010; Bicchieri et al., 2022). Specifically, Bicchieri et al. (2022) underscore the pivotal role of social proximity (e.g., higher IOS) on enhancing positive peer effects in norm compliance behaviors. They demonstrated that without social proximity, there's a decline in norm adherence; however, in the presence of social proximity, norm erosion is averted as participants react to both compliant and non-compliant peers' behaviors. This implies that the solely negative peer effects observed in our primary analysis could stem from insufficient social proximity toward partners. This idea is further supported by our data showing that the mean IOS score in paired conditions was a mere 2.19 out of 7, indicating a low average level of felt social proximity to partners.

Moreover, our analysis indicated the significance of skill similarity of partners in fostering positive peer effects. Our data indicated some evidence of a positive association between IOS and the number of words studied in the similar-finisher condition. This suggests that having partners of

similar skill levels, combined with social proximity, is pivotal for positive peer influences. This interpretation is corroborated by research showing the presence of positive peer effects among participants of similar abilities (Booij et al., 2017; Carrell et al., 2013).

Insert Table 4 about here.

### **Study 2: Sample with Lower Motivation in English Vocabulary Learning**

In Study 1, we identified negative peer effects exclusively when the partner's skills were similar to those of the participants. Despite this, no overarching positive peer effects were detected. We entertained the possibility that the failure to identify positive peer effects in Study 1 was due to a ceiling effect, suggesting that there was little room for already motivated participants to increase their study efforts.

Thus, we engaged participants with a lower baseline of motivation to delve deeper into the effect of learning partner skills and perseverance in Study 2.

**Participants.** We recruited 400 participants for our experiment via Lancers. We excluded 3 participants declaring to know more than 400 words, as in Study 1, because their vocabulary level was too high for this experiment. As a result, we obtained a demographically diverse sample of 397 participants in total (43% female,  $M_{age} = 40.66$  years, age range: 18-76 years). Of these participants, 47% are fully employed, 15% hold part-time positions, 2% are students, and the remaining 37% fall into other categories. The requirements for participation in this study were identical to those in Study 1, with the only distinction being that the participants must not have engaged in English language study for more than six months.<sup>7</sup> The participants received a fixed fee

---

<sup>7</sup> Not engaging English study refers to not taking English lessons or classes, not studying independently with English materials or apps, not watching English videos or podcasts, not learning by watching English movies or TV dramas, not reading English books or blogs, and not placing oneself in an environment where English is spoken.

of 600 yen if they completed all the parts of the experiment. The median duration of the experiment was around 39 minutes, but it varied depending on how many English words they wanted to learn.

**Design and procedure.** The procedure was exactly the same as in Study 1 except for one question. Right before the question asking how fun the study was, we added the intrinsic motivation scale (Schmidt et al., 1996). Incorporating this scale served to validate the enrollment of participants with weak motivation to learn English. The scale was a 6-point scale (1: Strongly disagree – 6: Strongly agree). The scale included four questions such as “I enjoy learning English very much”.<sup>8</sup> We took the average value of these questions to create the variable “motivation” ( $\alpha = 0.74$ ).

**Results and discussion.** Table 5 shows the selected characteristics of the participants for each treatment including variables *known\_words*, *eng\_level*, age, gender as in Study 1. The variables between the treatments were not statistically significant indicating the randomization worked well (all p-values above 0.1).<sup>9</sup> The mean value of motivation was 2.90 (the minimum possible value is 1 and the maximum possible value is 6), which was less than the middle point of 3.5 ( $t (396) = 13.81$ ;  $p < 0.001$ ), suggesting their motivation toward learning English was not high.

Insert Table 5 about here.

The manipulations regarding the skills and perseverance of the partners in this study were successful. Participants in the high-skilled condition believed their partner had significantly higher vocabulary levels ( $M = 5.51$ ,  $SD = 1.41$ ) compared to those in the similar-skilled partners ( $M = 4.07$ ,  $SD = 1.12$ ;  $F (1,319) = 101.26$ ,  $p < .001$ ,  $\eta^2 = .24$ ). Participants in the finisher conditions

---

<sup>8</sup> We excluded the one question “I wish I could learn English in an easier way, without going to class” from the original list of Schmidt et al. (1996) because not all the participants were students.

<sup>9</sup> As we expected, the degree of intrinsic motivation did not differ between treatments (ANOVA:  $F (4, 392) = 0.73$ ,  $p = .57$ ), indicating the variable *motivation* described stable personality trait.

perceived their partner studied more ( $M = 6.70, SD = 2.38$ ) compared to those in the quitter conditions ( $M = 3.75, SD = 2.66; F_{(1,319)} = 108.09, p < .001, \eta^2 = .25$ ).

The characteristics of the sample were different between Studies 1 and 2 as expected (Table 6). The variable *known\_words* was smaller in Study 2 ( $t(615) = 11.66, p < 0.001$ ) and the level of English proficiency was lower in Study 2 ( $t(674) = 9.12, p < 0.001$ ). The participants in Study 2 enjoyed the experiment less ( $t(759) = 4.93, p < 0.001$ ). These findings implied the participants in Study 2 were less motivated to learn English.

Insert Table 6 about here.

### ***Graphical analysis with ANOVA***

The top of Figure 4 shows that the number of studied words in the voluntarily part between the experimental conditions. We found a statistically significant main effect of the conditions on the number of their studied words was significant (analysis of variance (ANOVA):  $F_{(4, 392)} = 3.22, p = .01, \eta^2 = .03$ ). We further conducted multiple comparisons with Holm's sequentially rejective Bonferroni procedure. As can be seen in Figure 4, there appears negative effects from the partners who abandoned their learning efforts in the voluntary learning part midway (quitters) due to a noticeable difference of 13.53 words between that in high-quitter and single conditions.

Nonetheless, the number in the high-quitter condition ( $M = 24.67, SD = 31.92$ ) and that in the similar-quitter condition ( $M = 31.73, SD = 34.22$ ) did not significantly differ from that in the single condition ( $M = 38.20, SD = 37.72; t(392) = 2.40, p = .10$  and  $t(392) = 1.10, p = .99$ ). Note that we did find that the number was significantly smaller in the high-quitter condition compared to the similar-finisher ( $t(392) = 2.74, p = .04$ ) or higher-finisher conditions ( $t(392) = 3.05, p = .02$ ), indicating the negative impacts from the high-skilled partners quitting midway compared to finishers. On the other hand, no positive peer effects were found as in Study 1. The number of studied words in both the similar-finisher ( $M = 40.35, SD = 38.67$ ) and the high-finisher condition

( $M = 41.81$ ,  $SD = 40.79$ ) was not different from the single condition ( $t (392) = 0.36$ ,  $p = .99$  and  $t (392) = 0.61$ ,  $p = .99$ ).

The bottom of Figure 4 displays the number of correct answers in the test part across different conditions. The primary outcomes mirrored those of the analysis pertaining to the number of studied words. The results of ANOVA revealed the statistically significant main effect of experimental condition on the number of their correct answers ( $F_{(4, 392)} = 2.62$ ,  $p = .03$ ). Figure 4 indicates negative effects, characterized by a noticeable difference of 1.84 words, constituting approximately 6% of all questions, between the high-quitter and single conditions. Yet, the number of correct words in the high-quitter condition ( $M = 15.31$ ,  $SD = 5.92$ ) as well as that in the similar-quitter condition ( $M = 16.64$ ,  $SD = 5.97$ ) did not significantly differ from that in the single condition ( $M = 17.14$ ,  $SD = 6.39$ ;  $t (392) = 1.97$ ,  $p = .30$  and  $t (392) = 0.52$ ,  $p = .99$ ). Importantly, lower performance was indeed observed when the high-quitter condition was compared to the high-finisher condition because the number of correct words in the high-quitter condition was lower compared to that in the high-finisher condition ( $t (392) = 2.85$ ,  $p = .05$ ). On the other hand, no positive peer effects were found as in Study 1; the number of correct answers in both the similar-finisher ( $M = 17.72$ ,  $SD = 5.89$ ) and the high-finisher condition ( $M = 17.96$ ,  $SD = 6.07$ ) was not statistically different from the single condition ( $t (392) = 0.58$ ,  $p = .99$  and  $t (392) = 0.83$ ,  $p = .99$ ).

Insert Figure 4 about here.

### **Regression analysis**

In our subsequent analysis, we conducted regression method to further validate our findings, following the same approach as in Study 1. Initially, we used OLS regression to model the number of studied words based on treatment dummy variables (see column 1 at the top of Table 7). Subsequently, we incorporated demographic variables such as age, gender, and educational level, along with the variable *known\_words* (see column 2). Finally, we included *motivation* as an

additional independent variable (see column 3). We repeated these three analyses using Tobit regression methods (see columns 4, 5, and 6). The coefficient of the high-quitter condition had a significantly negative effect on the number across all the above regressions. However, the other treatment dummy variables had no significant impacts, indicating no positive peer effects. The variable age and motivation had positive impacts on the number of studied words.

We also conducted a regression analysis regarding the number of correct words in the test part (the bottom of Table 7). The main results were the same with the results regarding the number of studied words. Specifically, the high-quitter condition had a significantly negative effect across all six models (the coefficient was marginally significant for only one model). Additionally, motivation positively influenced the number of correct words.

Insert Table 7 about here.

Next, we conducted analysis focusing on the motivation. The variable motivation was positively correlated with the number of studied words ( $r = 0.15; p = 0.003$ ) and correct words ( $r = 0.20; p < 0.001$ ).

To probe further, we partitioned the sample into two groups based on the median motivation value of 3. Tobit regression analyses were conducted to examine the number of studied words, utilizing treatment dummy variables, demographic variables, and *known\_words* as the explanatory variables (Table 8). This analysis was carried out separately for the less intrinsically motivated sample (column 1) and the more motivated sample (column 2). A parallel approach was employed to analyze the number of correct words (columns 3 and 4). The coefficients for *high\_quitter* and *similar\_quitter* were significantly negative among the less motivated sample, while they were not statistically significant among the more motivated sample. This result suggests that negative peer effects primarily emanate from the less motivated group. In terms of the number of correct words within the less motivated sample, the coefficient for *high\_quitter* was significantly negative ( $p <$

0.001), although the coefficient for *similar\_quitter* was not statistically significant. Similar to the results regarding the number of studied words, negative effects on performance were not observed in the more motivated sample, as the coefficients for both *high\_quitter* and *similar\_quitter* were not statistically significant.

Insert Table 8 about here.

In alignment with the approach in Study 1, we examined the relationship between social proximity, as measured by IOS scale and learning outcomes. We found a moderate level of correlation between the number of studied words and IOS:  $r = -0.17$  ( $p = .13$ ) in the similar-quitter condition;  $r = -0.02$  ( $p = .88$ ) in the high-quitter condition;  $r = 0.23$  ( $p = .05$ ) in the similar-finisher condition; and  $r = 0.12$  ( $p = .28$ ) in the high-finisher condition. A similar correlation was observed between the number of correct words and IOS.<sup>10</sup> To further investigate these relationships, we regressed the number of studied words on the treatment dummy variables (with the similar-quitter condition as the baseline treatment), IOS and their interaction variables. The structure of these four regressions, presented at the top of Table 9, paralleled our previous analyses. We consistently found significant positive coefficients for the interaction term between *similar-finisher* and IOS across all four regressions: without control variables (1st column), with control variables and the variable *known\_words* (2nd column), Tobit regression without control variables (3rd column), and Tobit regression with control variables and the variable *known\_words* (4th column). This suggests more positive correlation between IOS and the number of studied words in the similar-finisher condition, compared to the similar-quitter condition across the four regressions. The bottom of Table 9 displayed analogous regressions for the number of correct words. Again, we found the significant positive coefficients for the interaction term between *similar-finisher* and IOS.

---

<sup>10</sup> The results are as follows:  $r = -0.13$  ( $p = .28$ ) in the similar-quitter condition;  $r = 0.11$  ( $p = .31$ ) in the high-quitter condition;  $r = 0.27$  ( $p = .02$ ) in the similar-finisher condition; and  $r = 0.13$  ( $p = .25$ ) in the high-finisher condition

In summary, Study 2 successfully recruited less motivated and less skilled participants, and it consistently demonstrated a negative impact from partners with lower levels of perseverance. Notably, these negative effects were observed from higher-skilled partners, rather than from those with similar skill levels. This result differs from the findings of Study 1, which demonstrated negative effects from similar-skilled partners. Further, our nuanced analysis revealed that the participants with lower intrinsic motivation towards English learning were primarily affected by these negative effects. In line with Study 1, we did not find evidence supporting a positive impact from partners exhibiting higher levels of perseverance. Also, we observed a positive association between IOS and the number of words studied in the similar-finisher condition as in Study 1.

Insert Table 9 about here.

### **General discussion and Conclusions**

We conducted two experiments, both resembling a typical online English learning environment to scrutinize the influence of learning partners' skills and perseverance on peer effects. Our sample consists of a diverse demographic of Japanese individuals, ranging from older, employed participants to younger students; therefore, the results should demonstrate substantial external validity.

Furthermore, the two types of the manipulated characteristics of the learning partners are exogenous, allowing for an accurate estimation of the causal effects rooted in these characteristics. Investigating peer effects presents inherent challenges, primarily due to identification problems associated with measuring peer effects (Manski, 1993; Sacerdote, 2014).<sup>11</sup>

---

<sup>11</sup> For example, when examining the impact of the other classmates' average score on an individual student's score, the student's score and the average score are determined simultaneously. Thus, there's a two-way relationship between them: the student's score influences the class average, and the class average may also influence the student's score. Estimating the precise degree of peer influence presents a significant challenge due to the inherent endogeneity.

In Study 1, the recruited participants were actively engaged in studying English. In Study 2, we retained the experimental structure but enlisted participants who had refrained from English study for a minimum of six months. A consistent finding across both studies was the negative impact of partners characterized by lower perseverance levels. In Study 1, this negative influence was attributed to partners of similar skill levels, whereas in Study 2, it emanated from higher-skilled partners. However, across both investigations, we did not encounter evidence of positive impacts from partners with higher levels of perseverance.

Deepening our analysis in Study 2 to examine the intrinsic motivation towards English learning, we discerned that the less intrinsically motivated sample was the primary contributor to these negative effects. This observation aligns with a previous survey study by Tanaka (2017). She identified the presence of negative peer effects, as well as the absence of positive peer effects, on one's motivation in a demotivating learning environment characterized by low intrinsic motivation and a pervasive low perceived competence in vocabulary learning.

Intriguingly, both studies provided some evidence suggesting that social proximity measured by IOS could foster positive effects when partners possess similar skills and higher perseverance levels. The results align with earlier research demonstrating pronounced peer effects among individuals sharing closer social distance (Bandiera et al., 2010), and the enhanced peer effects observed when students of similar abilities were grouped together (Booij et al., 2017; Carrell et al., 2013).

### *Interpretations of negative peer effects*

Our findings corroborate an asymmetry in both positive and negative peer effects, with negative influences from peers appearing to be more potent. The identification of negative peer effects on individual perseverance, or the absence of positive influences, is not an isolated finding in this study. Battaglini et al. (2005), as well as Battaglini et al., (2017), have documented the negative

impacts stemming from peers exhibiting low self-control. Georganas et al. (2015) reported that neither observation of peers' behaviors nor being observed by peers results in positive peer effects.<sup>12</sup> This pattern has also been observed in various contexts such as cooperation, dishonesty, and prosocial behaviors (Bicchieri et al., 2022; Isler & Gächter, 2022; Rauhut, 2013; Thöni & Gächter, 2015). Thöni and Gächter (2015) found that when participants observed a peer exerting less (hypothetical) effort given a certain wage, they substantially decreased their own effort in a gift-exchange experiment in the lab. However, observing a peer choosing more effort did not consistently lead to an increase in their own effort. Isler and Gächter (2022) suggested that negative peer effects may arise when a peer, seen as representative of one's group, unexpectedly violates a norm.

In our study, strong negative peer effects on learning perceptions could manifest among participants whose partners unexpectedly quit learning early during the voluntary learning part. Although we found negative peer effects in both studies, the degree to which they manifested was dependent on the peer's skill level. Specifically, in Study 1, participants were not influenced by the high-quitter treatment, while in Study 2, they were affected by the high-quitter treatment. Such variations could not be sufficiently accounted for by the aforementioned studies, which have not explored how the magnitude of peer effects differs based on the skill level of peers.

---

<sup>12</sup> In contrast, Gerhards and Gravert (2020) noted that observing peer behavior significantly enhanced participants' perseverance in a task, as measured by their skipping behavior in an anagram task (word puzzle). In this task, skipping resulted in a monetary penalty, but waiting 90 seconds offered undertake an alternative task without any monetary cost. The divergence in findings from our study likely stems from fundamental differences in experimental design. Firstly, their study's anagram task inherently differs from our study's focus on English vocabulary learning in an app-like setting, especially in terms of intrinsic value. Moreover, the solvability of some anagram tasks varied, unlike the universal accessibility of English word study in our research. Secondly, monetary incentives motivated their participants, in contrast to the intrinsic motivation of our participants in learning English. Thirdly, the option to skip a task in their study complicates the interpretation of perseverance. Participants might have perceived peers' skipping behaviors not merely due to a lack of perseverance but probably as viewing the penalty as negligible, the wait as overly lengthy, or the peers less skilled. In contrast, quitting behaviors in our study more clearly indicate a loss of interest in learning.

Given the absence of definitive interpretations from existing theories and prior research, we offer speculative insights into our findings. In Study 2, the participants, who had less English knowledge, may have felt intimidated or demotivated when paired with higher-skilled partners. When such partners ceased their learning, participants might interpret this as an implicit message that the task was too challenging even for those with higher skill levels, thereby justifying their own decision to discontinue learning.

Furthermore, participants in Study 2 may have harbored underlying doubts about their English competence, perceiving the task excessively difficult. Witnessing a higher-skilled partner cease learning could reinforce these beliefs, (c.f., confirmation bias; Nickerson, 1998) thus contributing to their decision to cease learning.

Conversely, in Study 1, the participants, possessing a certain level of English knowledge, seem to have interpreted that “high-quitters”—those with even higher ability than themselves—stopped learning because they already had substantial knowledge. Drawing upon the findings of Isler and Gächter (2022), it seems plausible that negative peer effects failed to occur, possibly because the act of quitting was not sufficiently surprising to the participants.

### *Implications*

Our research has primarily focused on the peer effects in online English learning due to the ease of implementing such studies. However, our findings can be applied broadly, thanks to the vast array of online courses available in diverse domains, such as programming, art, science, business management, and even online fitness or mindfulness courses like yoga and meditation. The impact of peer effects in these various fields may diverge from our current findings, offering deeper insights into the nature of peer influences. Consequently, these areas of investigation will serve as the basis of our future research.

To address the issue of attrition, online learning platforms like Duolingo have proactively developed social features to promote users' engagement. Our study provides two practical suggestions for enhancing user-matching algorithms to optimize motivation and engagement. First, caution is advised when pairing users with partners who display low perseverance, especially for users who are less intrinsically motivated and hence more susceptible to negative peer influences. Second, our data indicated that increased engagement and motivation could be achieved by matching users with socially proximate partners who are similarly skilled and display higher perseverance levels. These user-matching strategies have the potential to be effective not only in online educational settings but also in various other collaborative scenarios.

## References

- Aron, A., Aron, E. N., & Smollan, D. (1992). Inclusion of other in the self scale and the structure of interpersonal closeness. *Journal of Personality and Social Psychology*, 63(4), 596–612.  
<https://doi.org/10.1037/0022-3514.63.4.596>
- Bandiera, O., Barankay, I., & Rasul, I. (2010). Social incentives in the workplace. *Review of Economic Studies*, 77(2), 417–458. <https://doi.org/10.1111/j.1467-937X.2009.00574.x>
- Baron, J. N., & Kreps, D. M. (1999). Consistent human resource practices. *California Management Review*, 41(3), 29–53. <https://doi.org/10.2307/41165996>
- Battaglini, M., Bénabou, R., & Tirole, J. (2005). Self-control in peer groups. *Journal of Economic Theory*, 123(2), 105–134. <https://doi.org/10.1016/j.jet.2005.04.001>
- Battaglini, M., Díaz, C., & Patacchini, E. (2017). Self-control and peer groups: An empirical analysis. *Journal of Economic Behavior & Organization*, 134, 240–254.  
<https://doi.org/10.1016/j.jebo.2016.12.018>
- Beugnot, J., Fortin, B., Lacroix, G., & Villeval, M. C. (2019). Gender and peer effects on performance in social networks. *European Economic Review*, 113, 207–224.  
<https://doi.org/10.1016/j.eurocorev.2019.01.002>
- Bicchieri, C., Dimant, E., Gächter, S., & Nosenzo, D. (2022). Social proximity and the erosion of norm compliance. *Games and Economic Behavior*, 132, 59–72.  
<https://doi.org/10.1016/j.geb.2021.11.012>
- Booij, A. S., Leuven, E., & Oosterbeek, H. (2017). Ability peer effects in university: Evidence from a randomized experiment. *The Review of Economic Studies*, 84(2), 547–578.  
<https://doi.org/10.1093/restud/rdw045>
- Buechel, B., Mechtenberg, L., & Petersen, J. (2014). Peer effects and students' self-control. *MPRA Paper*, Article 53658. <https://ideas.repec.org/p/pra/mprapa/53658.html>

- Buechel, B., Mechtenberg, L., & Petersen, J. (2018). If I can do it, so can you! Peer effects on perseverance. *Journal of Economic Behavior & Organization*, 155, 301–314.  
<https://doi.org/10.1016/j.jebo.2018.09.004>
- Bursztyn, L., & Jensen, R. (2017). Social image and economic behavior in the field: Identifying, understanding, and shaping social pressure. *Annual Review of Economics*, 9(1), 131–153.  
<https://doi.org/10.1146/annurev-economics-063016-103625>
- Carrell, S. E., Sacerdote, B. I., & West, J. E. (2013). From natural variation to optimal policy? The importance of endogenous peer group formation. *Econometrica*, 81(3), 855–882.  
<https://doi.org/10.3982/ECTA10168>
- Falk, A., & Ichino, A. (2006). Clean evidence on peer effects. *Journal of Labor Economics*, 24(1), 39–57. <https://doi.org/10.1086/497818>
- Feld, J., & Zöllitz, U. (2017). Understanding peer effects: On the nature, estimation, and channels of peer effects. *Journal of Labor Economics*, 35(2), 387–428. <https://doi.org/10.1086/689472>
- Georganas, S., Tonin, M., & Vlassopoulos, M. (2015). Peer pressure and productivity: The role of observing and being observed. *Journal of Economic Behavior & Organization*, 117, 223–232. <https://doi.org/10.1016/j.jebo.2015.06.014>
- Gerhards, L., & Gravert, C. (2020). Because of you I did not give up – Peer effects in perseverance. *Journal of Economic Psychology*, 81, 102316. <https://doi.org/10.1016/j.joep.2020.102316>
- Gill, D., & Prowse, V. (2012). A structural analysis of disappointment aversion in a real effort competition. *The American Economic Review*, 102(1), 469–503.
- Golsteyn, B. H. H., Non, A., & Zöllitz, U. (2021). The impact of peer personality on academic achievement. *Journal of Political Economy*, 129(4), 1052–1099.  
<https://doi.org/10.1086/712638>

Hoxby, C. M., & Weingarth, G. (2006). *Taking race out of the equation: School reassignment and the structure of peer effects.*

<https://citeseerx.ist.psu.edu/viewdoc/download?repid=rep1&type=pdf&doi=10.1.1.472.256>

1

Isler, O., & Gächter, S. (2022). Conforming with peers in honesty and cooperation. *Journal of*

*Economic Behavior & Organization, 195*, 75–86.

<https://doi.org/10.1016/j.jebo.2021.12.026>

Kandel, E., & Lazear, E. P. (1992). Peer pressure and partnerships. *Journal of Political Economy*,

*100*(4), 801–817. <https://doi.org/10.1086/261840>

Levy, Y. (2007). Comparing dropouts and persistence in e-learning courses. *Computers &*

*Education, 48*(2), 185–204. <https://doi.org/10.1016/j.compedu.2004.12.004>

Lomi, A., Snijders, T. A. B., Steglich, C. E. G., & Torlo, V. J. (2011). Why are some more peer

than others? Evidence from a longitudinal study of social networks and individual academic performance. *Social Science Research, 40*(6), 1506–1520.

<https://doi.org/10.1016/j.ssresearch.2011.06.010>

Manski, C. F. (1993). Identification of endogenous social effects: The reflection problem. *The*

*Review of Economic Studies, 60*(3), 531–542. <https://doi.org/10.2307/2298123>

Mas, A., & Moretti, E. (2009). Peers at work. *American Economic Review, 99*(1), 112–145.

<https://doi.org/10.1257/aer.99.1.112>

Narayanasamy, S. K., & Elçi, A. (2020). An effective prediction model for online course dropout

rate. *International Journal of Distance Education Technologies (IJDET), 18*(4), 94–110.

<https://doi.org/10.4018/IJDET.2020100106>

Nation, I. S. P. (Ed.). (2001). The Cambridge Applied Linguistics Series. In *Learning vocabulary in*

*another language* (3rd ed., pp. ii–iv). Cambridge University Press.

- Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology*, 2(2), 175–220. <https://doi.org/10.1037/1089-2680.2.2.175>
- Rauhut, H. (2013). Beliefs about lying and spreading of dishonesty: Undetected lies and their constructive and destructive social dynamics in dice experiments. *PLOS ONE*, 8(11), e77878. <https://doi.org/10.1371/journal.pone.0077878>
- Rosaz, J., Slonim, R., & Villeval, M. C. (2016). Quitting and peer effects at work. *Labour Economics*, 39, 55–67. <https://doi.org/10.1016/j.labeco.2016.02.002>
- Sacerdote, B. (2014). Experimental and quasi-experimental analysis of peer effects: Two steps forward? *Annual Review of Economics*, 6(1), 253–272. <https://doi.org/10.1146/annurev-economics-071813-104217>
- Schmidt, R., Boraie, D., & Kassabgy, O. (1996). Foreign Language motivation: Internal structure and external connections. In *Language learning motivation: Pathways to the new century* (pp. 9–70). University of Hawaii, Second Language Teaching & Curriculum Center. <http://hdl.handle.net/10125/40769>
- Shimazu, K. (2021). *Bijuaru de oboeru IELTS kihon eitango*. Japan taimuzu shuppan.
- Tanaka, M. (2017). Examining EFL vocabulary learning motivation in a demotivating learning environment. *System*, 65, 130–138. <https://doi.org/10.1016/j.system.2017.01.010>
- Tello, S. F. (2007). An analysis of student persistence in online education: *International Journal of Information and Communication Technology Education*, 3(3), 47–62. <https://doi.org/10.4018/jicte.2007070105>
- Thöni, C., & Gächter, S. (2015). Peer effects and social preferences in voluntary cooperation: A theoretical and experimental analysis. *Journal of Economic Psychology*, 48, 72–88. <https://doi.org/10.1016/j.joep.2015.03.001>

- Villeval, M. C. (2020). Performance feedback and peer effects. In K. F. Zimmermann (Ed.), *Handbook of Labor, Human Resources and Population Economics* (pp. 1–38). Springer International Publishing. [https://doi.org/10.1007/978-3-319-57365-6\\_126-1](https://doi.org/10.1007/978-3-319-57365-6_126-1)
- Zabell, S. (2023, June 6). *4 reasons Duolingo is more fun with friends*. Duolingo Blog. <https://blog.duolingo.com/friends-social-features/>

Table 1

*Summary Procedure of the Experiment*

Part	Single condition		Pair condition			
			High-skilled		Similar-skilled	
	Finisher	Quitter	Finisher	Quitter		
1. Pre-learning	Answered if they recognized given English words (common).					
	Answered demographics and other related questions (common).					
2. Pre-study questionnaire and pairing announcement	No partner information provided.	<u>Received information on partner's English skills</u> Informed that partner answered 3 more words correctly.				Informed that partner answered the same number of words correctly.
3. Mandatory learning	Studied 5 out of 100 unfamiliar English words from pre-learning (common).				<u>Had the option to study English words with the partner.</u>	
4. Voluntary learning	Had the option to study English words alone. Could stop anytime after the first question.	<u>Could stop anytime after the first question.</u> Informed that their partner continued to study all questions (up to 95).				Informed that their partner continued to study all questions (up to 95). Informed that their partner stopped after the 6th question.
5. Test	Took a vocabulary test on 30 words they identified as unfamiliar during pre-learning (common).					

Table 2

Descriptive Statistics (Study 1)

	Similar-quitter	High-quitter	Similar-finisher	High-finisher	Single	Statistic test	p-value
known_words	80.00	86.81	90.48	81.01	83.94	$F_{(4,368)} = 0.27$	$p = .90$
	(70.82)	(70.10)	(88.42)	(65.17)	(59.70)		
eng_level	1.80	1.87	1.89	2.00	1.89	$F_{(4,368)} = 0.48$	$p = .75$
	(0.84)	(0.88)	(0.97)	(0.89)	(0.84)		
age	38.64	37.96	38.84	40.51	38.27	$F_{(4,368)} = 0.61$	$p = .66$
	(10.30)	(11.46)	(11.47)	(10.06)	(10.50)		
female	0.45	0.52	0.51	0.45	0.47	$\chi^2_{(4)} = 1.31$	$p = .86$
N	74	79	79	71	70		

Notes: Standard errors in parentheses. The variable *known\_words* represents the number of words participants knew in the pre-learning part; *eng\_level* denotes their self-assessed level of English proficiency; and *female* indicates the proportion of female participants.

Table 3

*Regression Results (Study 1): The Effect of Conditions on Perseverance (Top) and Performance (Bottom)*

	(1)	(2)	(3)	(4)
	OLS	OLS	Tobit	Tobit
similar_quitter	-17.033** (6.366)	-17.821** (6.236)	-25.589* (10.044)	-27.028** (9.779)
high_quitter	-5.281 (6.268)	-4.307 (6.146)	-5.733 (9.893)	-4.201 (9.627)
similar_finisher	0.972 (6.268)	0.660 (6.182)	0.649 (10.009)	0.560 (9.796)
high_finisher	1.263 (6.431)	0.329 (6.324)	0.811 (10.256)	-0.866 (10.015)
age		0.488** (0.182)		0.733* (0.289)
female		-5.661 (3.889)		-10.472 (6.129)
college		-5.143 (4.339)		-7.671 (6.849)
known_words		-0.087** (0.027)		-0.126** (0.043)
Constant	42.357*** (4.564)	37.209*** (8.917)	50.001*** (7.270)	42.832** (14.121)
<i>N</i>	373	373	373	373

	(1)	(2)	(3)	(4)
	OLS	OLS	Tobit	Tobit
similar_quitter	-2.924 ** (1.017)	-2.891 ** (1.014)	-2.996 ** (1.023)	-2.960 ** (1.014)
high_quitter	-0.662 (1.001)	-0.704 (0.999)	-0.703 (1.007)	-0.745 (1.000)
similar_finisher	-0.788 (1.001)	-0.739 (1.005)	-0.830 (1.007)	-0.770 (1.006)
high_finisher	0.865 (1.027)	0.754 (1.028)	0.830 (1.034)	0.718 (1.029)
age		0.026 (0.030)		0.024 (0.030)
female		-0.310 (0.632)		-0.321 (0.632)
college		0.692 (0.705)		0.743 (0.705)
known_words		0.008 (0.004)		0.009 (0.004)
Constant	19.586 *** (0.729)	17.557 *** (1.450)	19.658 *** (0.734)	17.643 *** (1.450)
<i>N</i>	373	373	373	373

Notes: The first two columns of the table report the results from ordinary least squares (OLS)

regressions, while the last two columns present the outcomes of Tobit regressions. The base condition of these analyses is the Single condition. Standard errors in parentheses; \*, \*\* and \*\*\* stand for statistical significance at the 5%, 1% and 0.1% level respectively.

Table 4

*Regression Results (Study 1): The Effect of Conditions and IOS on Perseverance (Top) and Performance (Bottom)*

	(1)	(2)	(3)	(4)
	OLS	OLS	Tobit	Tobit
high_quitter	9.487	7.099	24.339	21.742
	(14.452)	(14.208)	(22.440)	(21.882)
similar_finisher	-17.989	-24.337	-23.265	-32.710
	(14.422)	(14.161)	(22.781)	(22.196)
high_finisher	11.558	10.957	27.930	27.017
	(14.696)	(14.242)	(23.198)	(22.334)
IOS	-10.270	-12.309*	-10.475	-13.393
	(5.422)	(5.334)	(8.320)	(8.134)
high_quitter#IOS	0.952	3.455	-3.012	0.390
	(7.182)	(7.114)	(11.009)	(10.808)
similar_finisher#IOS	16.538**	19.635**	21.437*	26.243**
	(6.248)	(6.122)	(9.766)	(9.512)
high_finisher#IOS	5.173	5.730	1.605	2.346
	(6.530)	(6.364)	(10.154)	(9.832)
known_words		-0.081**		-0.117**
		(0.028)		(0.044)
Constant	44.338***	34.239*	44.070*	28.456
	(10.935)	(13.133)	(16.975)	(20.350)
Controls	No	Yes	No	Yes
<i>N</i>	303	303	303	303

	(1)	(2)	(3)	(4)
	OLS	OLS	Tobit	Tobit
high_quitter	1.423	1.940	1.513	2.066
	(2.387)	(2.415)	(2.376)	(2.388)
similar_finisher	-2.531	-2.464	-2.561	-2.473
	(2.382)	(2.407)	(2.370)	(2.379)
high_finisher	2.286	2.616	2.348	2.691
	(2.427)	(2.421)	(2.416)	(2.393)
IOS	-1.493	-1.346	-1.493	-1.335
	(0.895)	(0.907)	(0.891)	(0.896)
high_quitter#IOS	0.422	0.107	0.390	0.054
	(1.186)	(1.209)	(1.180)	(1.196)
similar_finisher#IOS	2.195*	2.148*	2.218*	2.162*
	(1.032)	(1.041)	(1.027)	(1.029)
high_finisher#IOS	0.969	0.725	0.959	0.705
	(1.078)	(1.082)	(1.073)	(1.069)
known_words		0.008		0.008
		(0.005)		(0.005)
Constant	19.426***	16.975***	19.426***	16.940***
	(1.806)	(2.233)	(1.797)	(2.206)
Controls	No	Yes	No	Yes
<i>N</i>	303	303	303	303

Notes: The first two columns of the table report the results from ordinary least squares (OLS)

regressions, while the last two columns present the outcomes of Tobit regressions. The base condition of these analyses is the Single condition. In the second and fourth columns, variables

controlled for include age, gender (noted as female), and attainment of a college degree or higher.

Standard errors in parentheses; \*, \*\* and \*\*\* stand for statistical significance at the 5%, 1% and 0.1% level respectively.

Table 5

*Descriptive Statistics (Study 2)*

	Similar-quitter	High-quitter	Similar-finisher	High-finisher	single	Statistic test	p-value
known_words	37.97	33.48	31.15	35.79	32.11	$F_{(4,392)} = 0.30$	$p = .88$
	(43.34)	(38.40)	(57.88)	(43.12)	(40.23)		
eng_level	1.38	1.51	1.25	1.31	1.38	$F_{(4,392)} = 1.94$	$p = .10$
	(0.65)	(0.71)	(0.55)	(0.57)	(0.67)		
motivation	2.98	2.88	2.78	2.87	2.98	$F_{(4,392)} = 0.73$	$p = .57$
	(0.98)	(0.88)	(0.66)	(0.87)	(0.89)		
age	41.22	39.37	40.40	41.30	41.28	$F_{(4,392)} = 0.53$	$p = .71$
	(11.42)	(10.42)	(9.3)	(12.29)	(10.03)		
female	0.39	0.39	0.50	0.44	0.43	$\chi^2_{(4)} = 2.65$	$p = .62$
N	78	94	72	77	76		

Notes: Standard deviations in parentheses. The variable *known\_words* represents the number of words participants knew in the pre-learning part; *eng\_level* denotes their self-assessed level of English proficiency; *motivation* describes the average value of the intrinsic motivation scale (Schmidt et al., 1996); and *female* indicates the proportion of female participants.

Table 6

*Summary Statistics in Studies 1 and 2*

	<b>Study 1</b>	<b>Study 2</b>	<b>t-test</b>	<b>p-value</b>
known_words	84.60 (71.62)	34.13 (44.52)	11.66	$p < .001$
eng_level	1.89 (0.89)	1.38 (0.64)	9.12	$p < .001$
enjoy	7.51 (2.34)	6.70 (2.23)	4.93	$p < .001$
# of studied words (single treatment)	42.36 (39.37)	38.20 (37.72)	0.65	$p < .52$
# of correct words (single treatment)	19.59 (5.65)	17.14 (6.39)	2.45	$p < .02$

Notes: Standard deviations in parentheses. The variable *known\_words* represents the number of words participants knew in the pre-learning part; *eng\_level* denotes their self-assessed level of English proficiency; and *enjoy* measures the overall enjoyment participants reported for learning English words in the experiment.

Table 7

*Regression Results (Study 2): The Effect of Conditions on Perseverance (Top) and Performance (Bottom)*

	(1)	(2)	(3)	(4)	(5)	(6)
	OLS	OLS	OLS	Tobit	Tobit	Tobit
similar_quitter	-6.467	-6.448	-6.302	-9.625	-9.610	-9.049
	(5.895)	(5.850)	(5.783)	(8.655)	(8.540)	(8.426)
high_quitter	-13.527*	-12.786*	-12.044*	-18.856*	-17.699*	-16.295*
	(5.642)	(5.614)	(5.554)	(8.285)	(8.185)	(8.083)
similar_finisher	2.150	2.247	3.637	2.232	2.491	4.728
	(6.015)	(5.945)	(5.892)	(8.886)	(8.731)	(8.639)
high_finisher	3.608	3.543	4.411	4.606	4.528	5.945
	(5.913)	(5.842)	(5.781)	(8.786)	(8.633)	(8.529)
age		0.538**	0.539**		0.830***	0.830***
		(0.170)	(0.168)		(0.250)	(0.247)
female		5.748	4.966		8.414	7.321
		(3.702)	(3.668)		(5.436)	(5.374)
college		3.450	3.066		4.656	4.055
		(3.824)	(3.781)		(5.607)	(5.538)
known_words		-0.039	-0.061		-0.056	-0.088
		(0.042)	(0.042)		(0.061)	(0.061)
motivation			6.739**		10.004**	
			(2.119)		(3.143)	
Constant	38.197***	12.948	-5.948	43.866***	5.325	-22.937

	(4.195)	(8.638)	(10.402)	(6.198)	(12.659)	(15.367)
N	397	397	397	397	397	397
	(1)	(2)	(3)	(4)	(5)	(6)
	OLS	OLS	OLS	Tobit	Tobit	Tobit
similar_quitter	-0.504 (0.975)	-0.853 (0.962)	-0.824 (0.946)	-0.476 (0.978)	-0.827 (0.960)	-0.799 (0.943)
high_quitter	-1.836* (0.933)	-2.044* (0.923)	-1.899* (0.908)	-1.815 (0.936)	-2.025* (0.921)	-1.875* (0.905)
similar_finisher	0.577 (0.995)	0.654 (0.978)	0.927 (0.964)	0.553 (0.998)	0.629 (0.976)	0.906 (0.961)
high_finisher	0.816 (0.978)	0.681 (0.961)	0.851 (0.945)	0.846 (0.981)	0.708 (0.959)	0.877 (0.942)
age		0.053 (0.028)	0.053 (0.027)		0.051 (0.028)	0.052 (0.027)
female		0.085 (0.609)	-0.068 (0.600)		0.071 (0.608)	-0.086 (0.598)
college		1.764** (0.629)	1.688** (0.618)		1.755** (0.628)	1.681** (0.617)
known_words		0.015* (0.007)	0.010 (0.007)		0.015* (0.007)	0.010 (0.007)
motivation			1.324***			1.337***

			(0.347)			(0.346)
Constant	17.145***	13.537***	9.823***	17.145***	13.598***	9.847***
	(0.694)	(1.421)	(1.701)	(0.696)	(1.418)	(1.696)

---

N	397	397	397	397	397	397
---	-----	-----	-----	-----	-----	-----

Notes: The first three columns of the table report the results from ordinary least squares (OLS) regressions, while the last three columns present the outcomes of Tobit regressions. The base condition of these analyses is the Single condition. Standard errors in parentheses; \*, \*\* and \*\*\* stand for statistical significance at the 5%, 1% and 0.1% level respectively.

Table 8

*Tobit Regression Results by Motivation-divided Group (Study 2): The Effect of Conditions on Perseverance and Performance*

	(1)	(2)	(3)	(4)
	DV = Perseverance	DV = Perseverance	DV = Performance	DV = Performance
	Less motivated	More motivated	Less motivated	More motivated
similar_quitter	-32.992** (11.924)	6.982 (11.928)	-1.461 (1.395)	-0.335 (1.294)
high_quitter	-42.146*** (11.292)	2.777 (11.666)	-3.436* (1.313)	-0.385 (1.272)
similar_finisher	-12.693 (11.684)	12.802 (12.937)	0.628 (1.366)	0.784 (1.388)
high_finisher	-16.202 (12.149)	21.290 (11.918)	1.315 (1.405)	0.351 (1.278)
age	0.819* (0.327)	0.966** (0.368)	0.043 (0.039)	0.073 (0.039)
female	-3.239 (7.259)	19.100* (7.919)	-0.689 (0.857)	0.725 (0.847)
college	12.221 (7.475)	-5.397 (8.120)	1.850* (0.880)	1.472 (0.874)
known_words	0.042 (0.114)	-0.095 (0.074)	0.025 (0.013)	0.010 (0.008)
Constant	18.618 (17.074)	-7.640 (18.299)	13.595*** (2.006)	13.054*** (1.975)
N	193	204	193	204

Notes: Standard errors in parentheses; \*, \*\* and \*\*\* stand for statistical significance at the 5%, 1% and 0.1% level respectively. Columns 1 and 3 include only the less intrinsically motivated sample (*motivation* < 3), while columns 2 and 4 contain only the more motivated sample (*motivation*  $\geq$  3). The variable *known\_words* represents the number of words participants knew in the pre-learning part; and *college* denotes if the participant has a college degree.

Table 9

*Regression Results (Study 2): The Effect of Conditions and IOS on Perseverance (Top) and Performance (Bottom)*

	(1)	(2)	(3)	(4)
	OLS	OLS	Tobit	Tobit
high_quitter	-15.658	-14.582	-22.635	-20.613
	(10.663)	(10.576)	(15.340)	(15.097)
similar_finisher	-16.284	-15.082	-25.178	-22.917
	(11.533)	(11.510)	(16.897)	(16.734)
high_finisher	-7.990	-7.254	-13.723	-12.487
	(11.279)	(11.245)	(16.716)	(16.573)
IOS	-4.551	-4.635	-5.960	-5.995
	(3.194)	(3.162)	(4.586)	(4.507)
high_quitter#IOS	4.081	3.838	6.541	5.994
	(4.694)	(4.654)	(6.720)	(6.611)
similar_finisher#IOS	11.503*	10.898*	17.204*	16.115*
	(4.603)	(4.572)	(6.835)	(6.734)
high_finisher#IOS	8.175	7.788	12.496	11.852
	(4.358)	(4.332)	(6.494)	(6.415)
known_words		-0.044		-0.061
		(0.045)		(0.065)
Constant	41.182***	18.615	46.551***	12.448
	(7.790)	(11.307)	(11.237)	(16.347)
Controls	No	Yes	No	Yes

N	321	303	303	303
	(1)	(2)	(3)	(4)
	OLS	OLS	Tobit	Tobit
high_quitter	-3.582*	-3.486*	-3.600*	-3.501*
	(1.749)	(1.719)	(1.749)	(1.707)
similar_finisher	-2.869	-2.359	-2.964	-2.448
	(1.892)	(1.871)	(1.891)	(1.858)
high_finisher	-1.265	-0.409	-1.255	-0.407
	(1.851)	(1.828)	(1.850)	(1.816)
IOS	-0.579	-0.599	-0.597	-0.614
	(0.524)	(0.514)	(0.524)	(0.510)
high_quitter#IOS	1.154	1.189	1.158	1.191
	(0.770)	(0.756)	(0.770)	(0.751)
similar_finisher#IOS	1.815*	1.779*	1.836*	1.798*
	(0.755)	(0.743)	(0.754)	(0.738)
high_finisher#IOS	1.161	0.888	1.159	0.889
	(0.715)	(0.704)	(0.714)	(0.699)
known_words		0.018*		0.018*
		(0.007)		(0.007)
Constant	17.844***	13.950***	17.908***	14.070***
	(1.278)	(1.838)	(1.278)	(1.826)
Controls	No	Yes	No	Yes

<i>N</i>	321	321	321	321
----------	-----	-----	-----	-----

---

Notes: The first two columns of the table report the results from ordinary least squares (OLS) regressions, while the last two columns present the outcomes of Tobit regressions. The base condition of these analyses is the Single condition. In the second and fourth columns, variables controlled for include age, gender (noted as female), and attainment of a college degree or higher. Standard errors in parentheses; \*, \*\* and \*\*\* stand for statistical significance at the 5%, 1% and 0.1% level respectively.

User 'PTQOI84' is **female**. Here is PTQOI84's avatar.



Moreover, the algorithm of this English vocabulary app has determined that **PTQOI84's level of English vocabulary knowledge is the same as yours**.

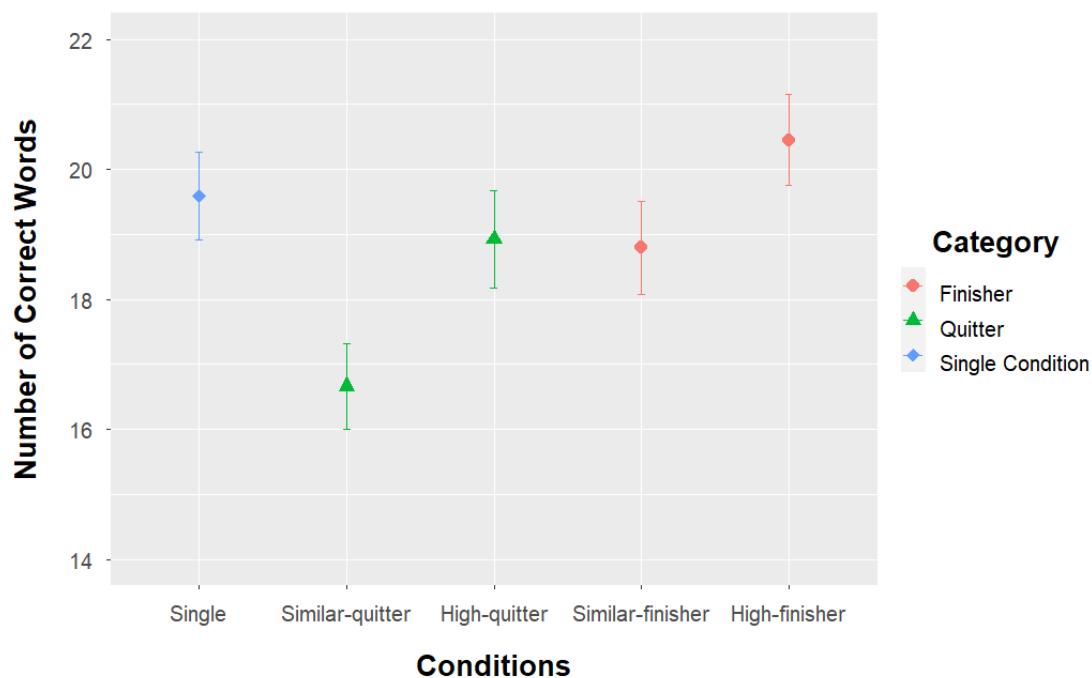
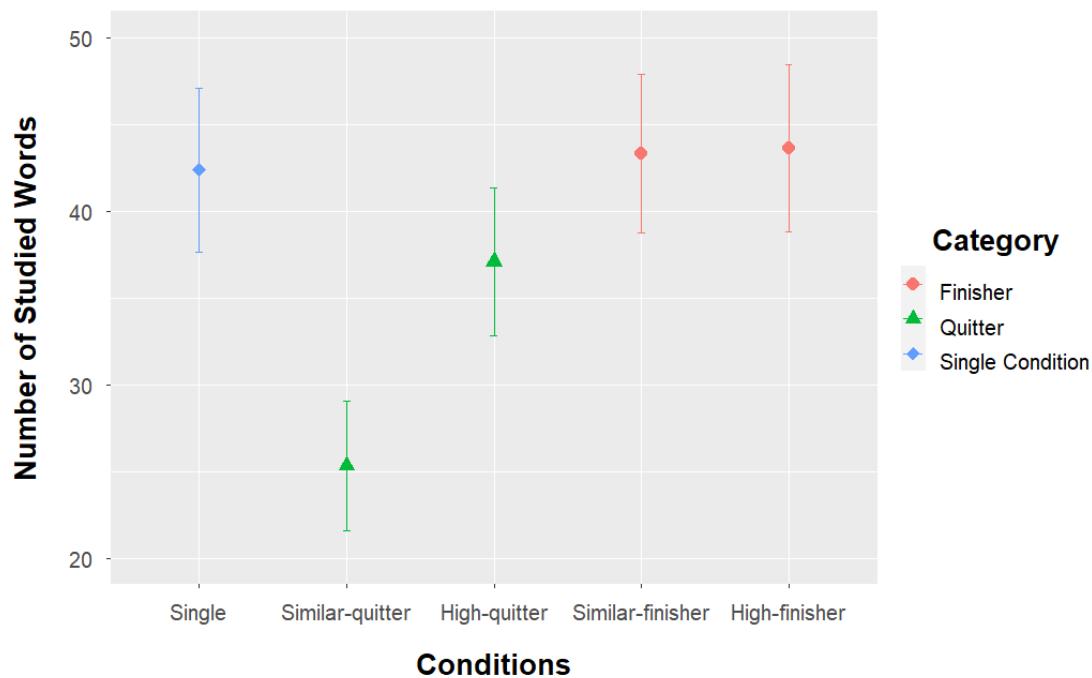
This is based on the English level responses provided earlier and the percentage of correct answers PTQOI84 gave in some questions in the Knowledge parts, which was about the same as yours. In other words, out of the 10 questions that both you and PTQOI84 answered in common, PTQOI84 got the same number of words correct as you did.

*Figure 1. The Information of the Partner in the Instructions (Similar-skilled Partner Condition).*



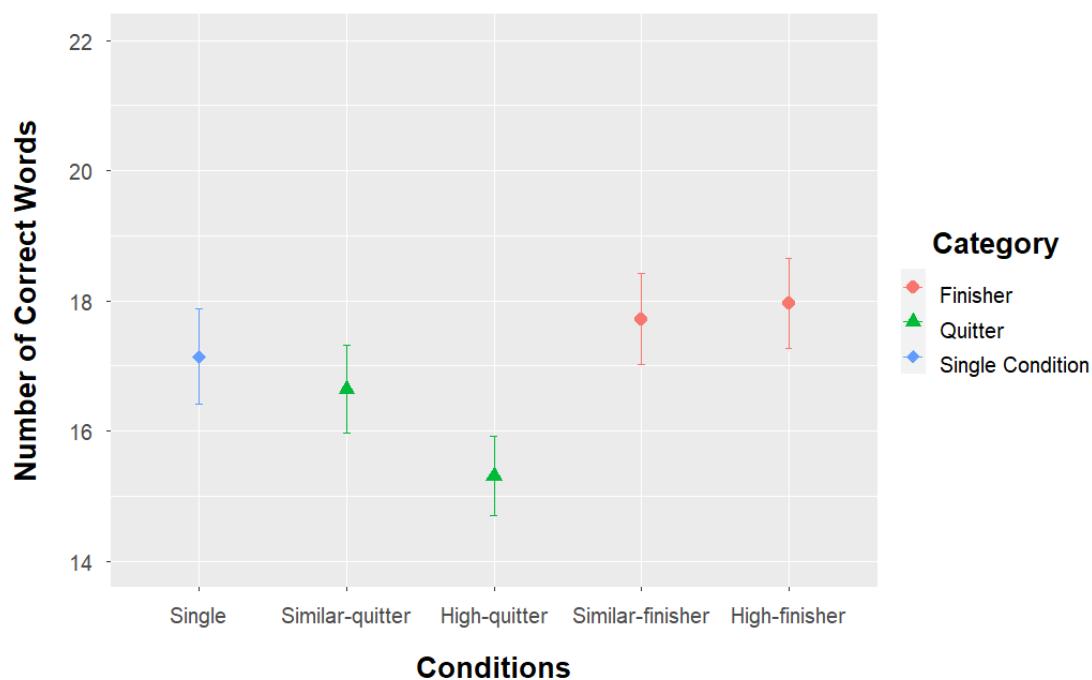
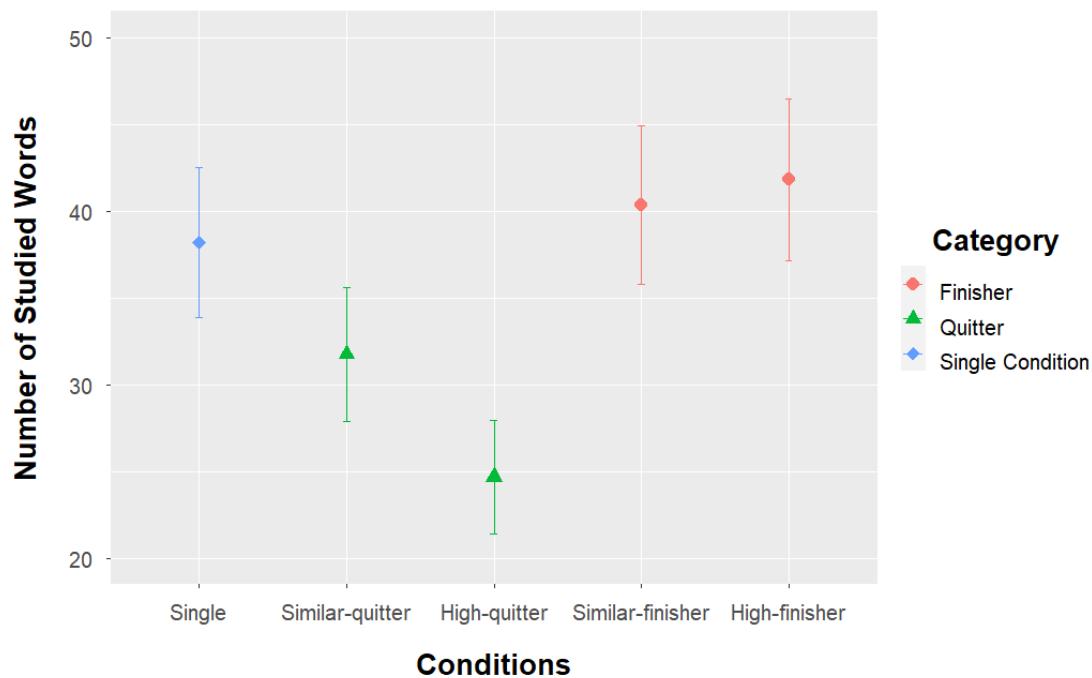
Figure 2. The Screenshot of an Example Question in the Voluntary Learning Part (Pair Condition).

In the upper right corner, the learning status of the partner (PTQOI84) and their sent emoji are displayed. Participants are required to select the correct Japanese translation of the presented English word from five given options.



*Figure 3.* The Number of Studied Words in the Voluntarily Learning Part (Top) and the Number of Correct Words in the Test Part (Bottom) for Each Treatment (Study 1).

Means are represented by three distinct shapes: a circle for Finisher, a triangle for Quitter, and a diamond for Single condition. Standard error bars are included.



*Figure 4.* The Number of Studied Words in the Voluntarily Learning Part (Top) and the Number of Correct Words in the Test Part (Bottom) for each Treatment (Study 2).

Means are represented by three distinct shapes: a circle for Finisher, a triangle for Quitter, and a diamond for Single condition. Standard error bars are included.