Climbing the Ladder or Falling Behind: The Role of Leaderboard Composition in User Engagement

Agnieszka Kloc[*], Rodrigo Belo[†], Ting Li[‡]

October, 2024

Abstract

Leaderboards are designed to keep users engaged through competition. Despite their widespread implementation in educational environments, empirical findings on their effectiveness are mixed. While some research points toward enhancements in self-regulation, motivation, and performance through gamified competition, others highlight potential drawbacks such as reduced self-esteem and diminished intrinsic motivation. This study investigates the impact of leaderboard composition on user engagement through a randomized field experiment and a follow-up online experiment. We find that users exhibit higher engagement when competing against those with different scores rather than similar ones. The effect varies with individual competitiveness levels. In a follow-up experiment, we tested three competition mechanisms: social presence, performance feedback, and social comparison. Results indicated that performance feedback is the main factor driving leaderboards' positive effects. Our study implies that a universal leaderboard approach is ineffective. Instead, leaderboards should be customized based on competitive conditions and the competitiveness of an average platform user.

Keywords: Leaderboards, User Engagement, Gamification, Competition, Randomized Field Experiment.

1 Introduction

Leaderboards, gamification elements that rank users based on performance, are often implemented in online platforms to induce competition and enhance user engagement. For instance, in educational platforms, leaderboards can address typical online learning challenges such as diminished

[*]Agnieszka Kloc, Rotterdam School of Management, kloc@rsm.nl.
[†]Rodrigo Belo, Nova School of Business and Economics, Universidade Nova de Lisboa, rodrigo.belo@novasbe.pt.
[‡]Ting Li, Rotterdam School of Management, tli@rsm.nl.

student engagement and motivation (Hew & Cheung, 2014), high levels of procrastination (Huang et al., 2021), and low self-discipline (Banerjee & Duflo, 2014). However, while some studies in information systems, psychology, and human-computer interaction have identified positive effects of gamified competition — improved self-regulation, increased motivation, and enhanced cognitive performance (e.g., Santhanam et al., 2016; Burguillo, 2010) — other research presents a contrasting view. Competition can sometimes negatively impact self-esteem and undermine intrinsic motivation (Hanus & Fox, 2015; Chan & Briers, 2019; Nebel, Beege, et al., 2016).

To explain this inconsistency, researchers have focused on understanding how leaderboard design influences competitive conditions and their effectiveness in engaging users (e.g., Nebel, Schneider, & Rey, 2016; Bai et al., 2020). The prevailing assumption is that users are competitive and respond positively to elements that enable competition. However, competitiveness is a trait that varies widely among individuals (Baldauf et al., 2014). Recent research underscores the complexity of the interaction between leaderboard design and user characteristics (Hydari et al., 2023; Ho et al., 2023), also suggesting that the mechanisms through which leaderboards impact user engagement remain unclear. While many investigations emphasize competition, emerging evidence indicates that factors like social comparison may also significantly influence behavior (Bojd et al., 2022). Our study seeks to bridge these gaps by addressing the following questions: *How does leaderboard composition affect user engagement?* and *What mechanisms drive the influence of leaderboards on user engagement?*

To address these questions, we conducted a randomized field experiment with a European online platform for students preparing for their high school final exams. Participants were randomly assigned to either a control group with no exposure to leaderboards, or treatment groups with different leaderboard configurations. Specifically, users were placed on leaderboards with peers showing either similar (low dispersion) or dissimilar (high dispersion) behavior to them prior to being assigned to the leaderboard. Counterintuitively, our findings reveal that leaderboards with high score dispersion (i.e., composed of users with dissimilar behavior) led to *increased* engagement compared to those with low score dispersion. Moreover, users were more engaged when there was a significant gap between their scores and those of users immediately above or below them. These results challenge conventional beliefs regarding the impact of similarity on engagement and demonstrate that competitive dynamics can be far more intricate.

To further explore the underlying mechanisms, we conducted a follow-up online experiment that isolated components of competition: social comparison, performance feedback, and social presence. Participants were randomly exposed either to the same treatments as in the field experiment or to treatments targeting specific component mechanisms (e.g., only performance feedback), allowing us to isolate their impacts on participant behavior. Our analysis showed the positive effects of competition on engagement were primarily driven by performance feedback, whereas the negative effects commonly associated with social comparison were not observed. Furthermore, social presence did not significantly affect the outcomes. These findings suggest that the competitive elements of leaderboards are not uniformly beneficial or detrimental, but their impact is mediated by specific underlying mechanisms.

Finally, we examine individual differences in competitiveness and their impact on engagement. We find that users with a higher competitive drive engaged more with leaderboards featuring greater similarity and competition intensity, while those with lower competitiveness responded better to lower similarity and intensity. This highlights the importance of tailoring leaderboard design to the competitive nature of the platform's users.

The findings provide a deeper understanding of how gamification elements affect user engagement, demonstrating that greater dissimilarity among competitors can actually foster higher engagement. This research contributes to the literature on gamified competition by shedding light on the role of leaderboard composition, competitive intensity, and performance feedback in shaping user behavior in online user engagement in online platforms (S. Song et al., 2021; Leung et al., 2022; Bojd et al., 2022). It extends existing knowledge by identifying the role of competition intensity and its mixed effectiveness in fostering engagement in digital environments. Mediation analysis highlights performance feedback as the primary positive mechanism behind the effectiveness of leaderboards. Additionally, our findings offer valuable insights into individual differences in responses to competition, highlighting the importance of understanding user competitiveness to design optimal leaderboards.

2 Related Literature

Effectiveness of Leaderboards First, our paper adds to the literature on the effectiveness of leaderboards in non-game environments. There is substantial evidence demonstrating the positive impact of gamified competitions on learning engagement (e.g., Amo et al., 2020; Ding et al., 2018) and cognitive task performance (e.g., Tsay et al., 2018). However, recent studies in information systems indicate that the effects of leaderboards can vary significantly depending on the specific tasks and user groups involved. For example, Bojd et al. (2022) analyzed the use of leaderboards in health and fitness challenges, finding that leaderboards with more active participants were more effective for diet challenges but less so for exercise challenges. Similarly, Hydari et al. (2023) examined Fitbit data from U.S. university undergraduates, comparing step counts between leaderboard participants and non-participants. While they observed that leaderboard participants generally walked more, their sub-sample analyses revealed heterogeneous effects: smaller leaderboards were associated with higher step counts for sedentary individuals but with lower counts for highly active individuals.

In learning contexts, leaderboards are commonly used to address challenges such as reduced student engagement and motivation (Hew & Cheung, 2014), high levels of procrastination (Huang et al., 2021), and low self-discipline (Banerjee & Duflo, 2014). They aim to create a competitive environment that encourages participation in learning activities (Landers & Landers, 2014; Ding et al., 2018). However, some research indicates that leaderboards can negatively impact learning processes and outcomes (e.g., Hanus & Fox, 2015; Philpott & Son, 2022). These studies often suggest that the decline is due to a shift from intrinsic to extrinsic motivation, where participants only improve their performance and engagement up to a certain point before plateauing. However, much of the existing research is limited by endogeneity and identification issues, relying on small-sample observational data or quasi-experimental designs. We conduct a randomized field experiment in a real-world online learning platform with thousands of users to provide more definitive insights. Additionally, these studies typically examine multiple gamification elements simultaneously (e.g., leaderboards and badges), which can obscure the specific impact of each element. In this paper, we focus exclusively on the leaderboard, which allows for a more definite answer to how effective leaderboards are on their own.

Leaderboard Design Second, our paper extends the literature on the elements of leaderboard design that influence competition intensity. Whereas previous studies have manipulated competition intensity by varying the number of competitors (Garcia & Tor, 2009) or the display of ranks (Bai et al., 2021; S. Song et al., 2021), we focus on *leaderboard composition*, i.e., who competes against whom.

The literature has shown that the perception of competition intensity can significantly impact how users engage with leaderboards. For example, in a lab experiment, Nebel et al. (2017) found that when most of opponents' scores were set higher than the focal user's score, participants increased effort but decreased learning. This suggests under higher competition intensity, users invest more effort in the mechanics of the competition and less in the learning process. A limitation of this study is that the focal user is never significantly ahead of other competitors, which limits the generalizability of the results, as users may react differently when placed in different positions. In our study, we measure the effects of competition intensity on engagement when the focal user is placed in different, non-simulated positions in the leaderboard, including in the middle and at the top.

Santhanam et al. (2016) found that participants who thought they were matched with an equally skilled competitor and ended in a tie (no distance between competitors), reported the highest levels of engagement. On the other hand, those who were made to believe they were paired with a less skilled competitor (larger distance) and won, reported higher self-efficacy (confidence in accomplishing goals) and achieved better learning outcomes. As the experiment relied on competing in pairs and not in a leaderboard, it is difficult to speculate how matching multiple people at once would affect their engagement. Users in a leaderboard are exposed to multiple competitors simultaneously, which creates more complex social dynamics as participants can focus on different competitors at different times.

Nebel, Beege, et al. (2016) manipulated perceptions of discrepancy with a desired standard, which was set as being at the top of the leaderboard. They found that when individuals were ranked mid-leaderboard, indicating a higher discrepancy, they reported increased competitive effort and improved learning outcomes, although the manipulation did not alter their perception of competition.

Similar to these studies, we manipulate the composition of the competing group by varying the similarity of participants in the leaderboard, i.e., the dispersion of scores and the distance to competitors. We aim to understand how these design features influence the perception of competition intensity and, consequently, user engagement. We focus on the effects of leaderboard composition on engagement in a real-world online learning platform, where users are exposed to multiple competitors simultaneously.

User Heterogeneity Third, our study contributes to the existing literature on the role of personal characteristics in responses to leaderboards. Much of the current research suggests that variations in response to leaderboards are rooted in how individuals react to competition. Generally, men tend to respond positively, while women often display reluctance in engaging with competitive scenarios, which adversely affects their subsequent performance within competitions (Niederle & Vesterlund, 2007, 2011). In contrast, Ho et al. (2023) find that women respond positively to leaderboards, although to a lesser extent than to coupons. Additionally, men sustain higher engagement levels in the post-treatment period, whereas women's overall engagement declines.

Competitiveness, defined as the ”dispositional preference to compete with others in achievement settings” (To et al., 2020), is another individual trait that significantly moderates the impact of leaderboards. Those with higher levels of competitiveness exhibit increased engagement when leaderboards are present, dedicating more time to each attempt (Amo et al., 2020). According to Höllig et al. (2020), the willingness to engage with a gamified system is primarily driven by the expectation of enjoyment, with competitiveness indirectly influencing this through heightened perceptions of enjoyment. Additionally, other unexplored traits and characteristics might also play a role (Amo et al., 2020; Hydari et al., 2023). For example, Theriault et al. (2021) propose that extroversion could serve as an additional moderator. However, our understanding of how competitive structures interact with users' characteristics remains limited. In this study, we measure and evaluate gender, competitiveness, task preparation, and goal orientation as potential moderators of the effects of leaderboards.

Why Do Leaderboards Work? Fourth, we study the underlying mechanisms behind the effectiveness of leaderboards in influencing user behavior. We are the first to gather, measure, and test a set of different potential leaderboard mechanisms. While leaderboards are commonly associated with the mechanism of competition, they serve three distinct functions that collectively contribute to creating a competitive environment. First, leaderboards enhance social presence by simulating the presence of others in an online environment. Second, they provide performance feedback by offering insights into individual performance. Last, leaderboards facilitate social comparison by enabling users to gauge their performance against others. To understand the impact of each function separately and in conjunction, we conducted an online experiment exposing participants to these dimensions individually.

Social presence has been studied in IS for its role in establishing trust in virtual environments (Srivastava & Chandra, 2018) and influencing attitudes toward customer service agents (Yang Gao et al., 2023). There is also evidence of a positive correlation between social presence and both learning engagement (Miao & Ma, 2022) and active learning (Molinillo et al., 2018). Nevertheless, the literature has not yet looked into the relative contribution of social presence in the context of leaderboards.

Performance feedback is widely recognized as a motivational tool in various contexts, such as crowdsourcing (Huang et al., 2019), pro-environmental initiatives (Karlin et al., 2015), intellectual tasks (Harackiewicz, 1979), and education (Hattie & Timperley, 2007). While it is generally perceived as motivating and beneficial for enhancing performance, feedback can also have negative consequences. For instance, when it shifts focus away from the task towards other self-related goals like impression management, it may lead to disengagement (Kluger & DeNisi, 1996). Recent reviews have highlighted pronounced variations in the effects of feedback depending on how it is provided (Wisniewski et al., 2020). As a result, the provision of performance feedback, particularly through leaderboards, may result in either positive or negative outcomes.

Social Comparison is a psychological process often linked to negative impacts on self-esteem (e.g., Vogel et al., 2014) and well-being (e.g., Park & Baek, 2018). However, social comparison has also been utilized as a mechanism to promote positive behavioral changes, such as encouraging contributions to online communities (Chen et al., 2010), promoting sustainable living practices (Bruchmann et al., 2021), or increasing engagement in online courses (Davis et al., 2017). In the context of leaderboards, Bojd et al. (2022) suggested that users may engage in more social comparison when there are fewer participants on the leaderboard and when extrinsic motivation is prominent (as seen in challenges related to dieting, for example). This heightened social comparison may lead to increased anxiety and, in turn, lower performance. Additionally, Hydari et al. (2023) proposed that smaller leaderboards could intensify social comparison, benefiting sedentary individuals while potentially hindering highly active individuals. As a result, we anticipate that social comparison may act as a mediator for the negative leaderboard effects on user engagement.

To disentangle which mechanisms mediate the effects of leaderboards, we run an online experiment to measure the reported social presence, performance feedback, and social comparison levels, and test them as potential mechanisms of the leaderboard effects alongside the mechanism of competition.

3 Empirical Context and Field Experiment

We use data from a randomized field experiment designed and implemented in collaboration with an online platform that helps high school students prepare for their final national exams. The company caters to the population of Dutch students and is one of the most popular online educational platforms in the country, having over six thousand active users weekly in the month preceding the exams. On the platform, users get access to a variety of revision materials, such as topic summaries, old exams, or interactive quizzes. The quizzes allow students to test their knowledge and achieve a grade estimate from a given exam (e.g. biology). The platform can be accessed either via desktop or mobile application.

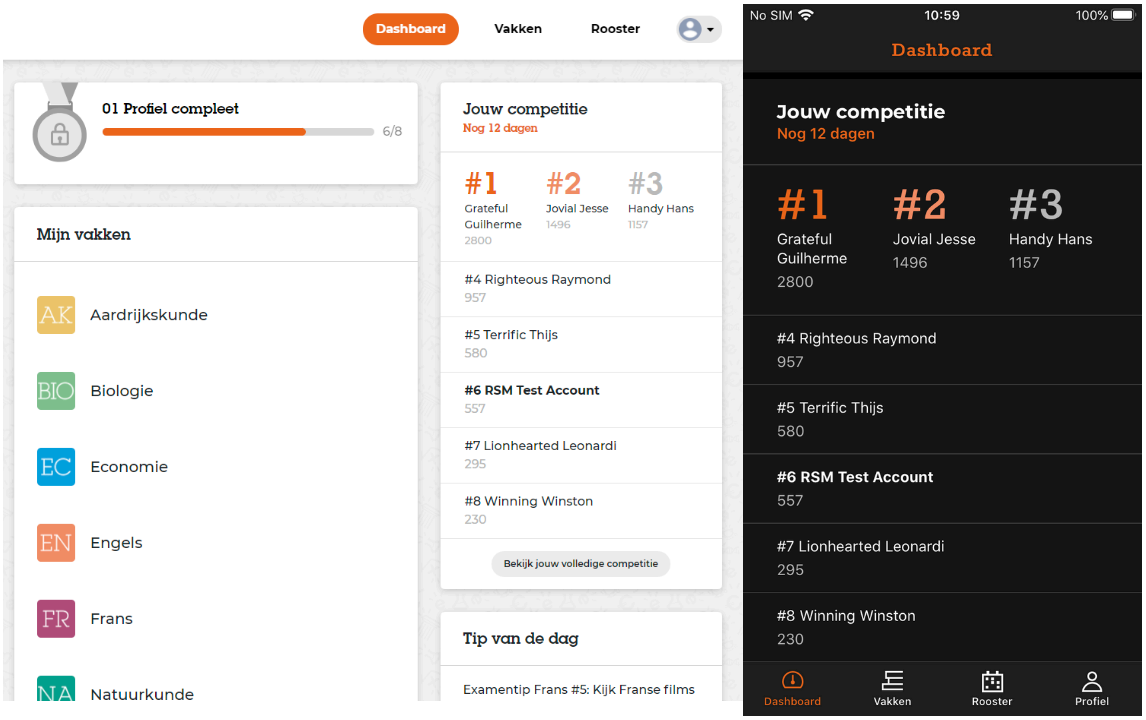
Before the exams of the academic year 2021/2022, the platform implemented a new design element, a leaderboard, to encourage students to spend more time studying on the platform and promoting usage continuance, i.e., repeated visits after the first log-in. Together with the platform, we designed a randomized field experiment in which we manipulated user allocation to leaderboards.

3.1 Experimental Design

The manipulation took place between March 16, 2022, and April 30, 2022. We observe all users registered during this period from the moment of their registration until May 27, 2022, the date of the last exam. Upon registration, each user was randomly assigned either to the treatment group — i.e., designated to be allocated to a leaderboard, with 2/3 probability — or to the control group — i.e., designated not to be allocated to a leaderboard, with 1/3 probability. Users in the treatment group were further randomly assigned to one of the two leaderboard compositions: leaderboards with high user similarity (users with similar scores) or leaderboards with low user similarity (users with dissimilar scores). The users in the control group did not receive any leaderboard-related information. For users assigned to a leaderboard, the leaderboard the user belongs to is displayed at the top of the main screen in the mobile app or the dashboard in the desktop app (see Figure 1 for reference).

Actual leaderboard allocation happened not at the moment of registration, but after a user became eligible for the study. Users became eligible and allocated to a leaderboard three days after their registration provided they remained active on the platform during the experiment period beyond the registration itself. Assigning users to leaderboards at least three days after their registration allowed us to rank them by their level of engagement before treatment and to ensure variation in the dispersion of user activity across leaderboards: every leaderboard was assembled such that either all users were drawn from the same activity quartile (high-similarity leaderboards) or evenly from all four quartiles (low-similarity leaderboards). Within a quartile, users were selected randomly, and each leaderboard was filled until it reached its maximum capacity of 20 participants. The starting leaderboard score was calculated using the time spent on the platform

Figure 1: Screenshot of the Desktop and Mobile App Main Screen. *Jouw Competitie* = Your Leaderboard. *Mijn Vakken* = My Subjects.



and points earned from quizzes up to the allocation moment. A user could receive 10 points for every minute spent on the platform and from 40 up to 640 points for every completed quiz (depending on the number of correct answers). To account for differences in the length of platform usage, we normalized the score by the number of days since the user registered.

3.2 Data and Sample Construction

Our final complete sample consists of 3,762 eligible users. Table 1 shows summary statistics at the leaderboard level. A total of 3,670 users were assigned to 188 leaderboards, out of which 2,491 belong to our final sample.[1] The average leaderboard consists of 19.5 users, with a minimum of 7 and a maximum of 20. As expected, and according to the treatment, leaderboards with users

[1] While our sample includes all users eligible for the study, not all of them were assigned to a leaderboard due to being allocated to the control group. In addition, not all users allocated to a leaderboard are part of our sample. This is because, at the moment the leaderboard functionality became available, some users were already active on the platform for a longer period, and although they were considered eligible to be assigned to a leaderboard by the platform, they were not eligible for the study as they had registered before the manipulation period.

drawn from the same activity quartile display a lower standard deviation in starting scores, when compared to those with users drawn uniformly from all quartiles ($sd_{low-similarity} = 416.36$ points vs. $sd_{high-similarity} = 216.22$ points, t-stat = 4.39, p-value < 0.01).

Table 1: Summary Statistics: Leaderboard Activity (Leaderboard Level).

Variable	N	Mean	Std. Dev.	Min	Pctl. 25	Pctl. 75	Max
Avg. Points Gained by User per Day	188	21.42	16.32	1.48	9.61	28.28	97.2
Total Number of Points Gained	188	12083.86	9474.76	890	5425	15665	58320
Total Hours of Engagement	188	19.43	15.85	1.42	8.45	25.51	97.2
Users Assigned	188	19.52	1.56	7	19	20	20

We compiled a daily panel featuring 3,762 eligible users. For each user, we collected information regarding their educational track (scientific, general, or vocational), their time spent using various platform features, and their performance on revision quizzes. Additionally, we encouraged users to voluntarily provide information about their gender, favorite school subjects, perceived level of exam preparation, and their goals for using the platform (referred to as 'student type'). Table 2 presents the descriptive statistics for the key demographic variables.

Table 2: Summary Statistics: Demographics and Activity.

Variable	N	Mean	Std. Dev.	Min	Pctl. 25	Pctl. 75	Max
Female	3762	0.73	0.44	0	0	1	1
General	3762	0.38	0.48	0	0	1	1
Scientific	3762	0.24	0.43	0	0	0	1
Vocational	3762	0.38	0.49	0	0	1	1
Achieve Future Goals	3762	0.37	0.48	0	0	1	1
Better than Friends	3762	0.08	0.27	0	0	0	1
Avoid Guilt	3762	0.11	0.32	0	0	0	1
For Parents	3762	0.23	0.42	0	0	0	1
Highest Grade	3762	0.37	0.48	0	0	1	1
Just Pass	3762	0.78	0.42	0	1	1	1
Exam Preparation	3762	3.48	2.06	0	2	5	10
Time Active (Minutes)	3762	124.79	221.65	6	22	126.75	3495
Login Days	3762	10.15	6.77	2	5	14	49
Learning (Minutes)	3762	94	186.82	0	12	88.75	3194
Quizzes (Minutes)	3762	2.58	16.47	0	0	0	617
Static Practice (Minutes)	3762	21.79	83.16	0	1	12	2580
Hardest Subject (Minutes)	2780	19.86	64.48	0	0	12	1243
N Questions Answered	3762	6.42	42.98	0	0	0	1434

The vast majority (73%) of our sample is female, with 24% of the platform users enrolled in the Scientific Education track (highest level), 38% General Education (middle level), and 38% Vocational Education (bottom level). This distribution is approximately representative of the general population of final exam takers, where 21% of users are pursuing Scientific Education, 30% are completing General Education and 49% are pursuing Vocational Education (CITO, 2021), with a slight over-representation of General Education users and an under-representation of Vocational Education users. The following six rows of the table correspond to student types that a user could describe oneself with (more than one option could be selected). The most popular type chosen by the platform users was "I just want to pass my exams" (78 % of users chose it), and the least popular was "I want to be better than my friends", claimed by 8% of users. Additionally, users are mostly not confident about their exam preparation when they join the platform. They assess their current preparation level on a scale of 1-10 at an average of 3.5.

Most of the accounts are created in two months preceding the first exam (March and April, see Figure 2). When it comes to activity, on average users are active for 10 days (out of a maximum of 80 observed days), during which they spend a total of 125 minutes on the platform. Out of the 125 minutes, the majority of the time is spent on learning activities (such as watching videos, reading summaries, and reviewing topic-related definitions). Additionally, 22 minutes are spent on non-interactive practice (reviewing old exams or exercise books), and the remaining time is spent on general, non-learning-related, activities. The peaks of platform traffic happen just before the exams, at the beginning of May (see Figure 3).

4 Results

4.1 Randomization Check

To confirm that the randomization was effective, we compared the treatment and control groups based on their demographic characteristics and pre-treatment behaviors. We performed a series of t-tests to identify any significant differences in these variables between the experimental groups. Time-varying variables — such as pre-treatment daily activity — were aggregated at the user level and then averaged across groups. There were no significant differences between the groups in any

Figure 2: Registrations over Time

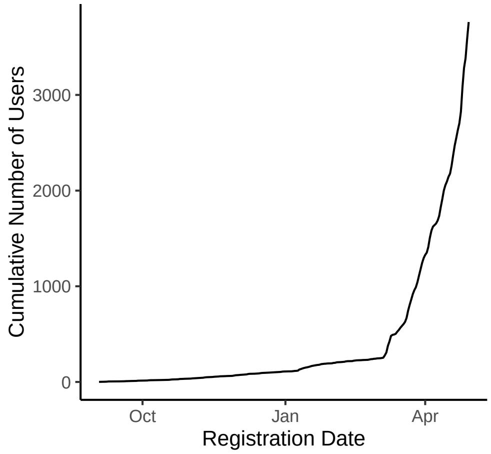
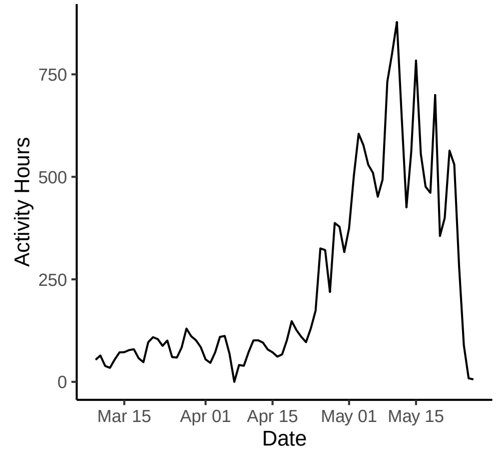


Figure 3: Activity over Time



demographic or activity-based variables, except for one registration goal related to guilt avoidance (t-value = 2.63, p-value = 0.01), which is to be expected, given the number of comparisons.[2]

4.2 Main Results

We start the analysis by looking into the effects of being assigned to a leaderboard on user engagement. We use a fixed-effects specification to estimate the effects of the treatment on the engagement variables:

$$\begin{aligned} \text{Engagement}_{it} &= \beta_1 \text{Leaderboard}_i + X_i \\ &+ tenure_{it} + w_t + \epsilon_{it} \end{aligned} \tag{1}$$

in which Engagement_{it} is the engagement variable for user i at time t, Leaderboard_i is a binary variable indicating whether a user was assigned to a leaderboard, X_i is a vector of user fixed effects of gender and education type, $tenure_{it}$ represents the user's platform tenure, w_t represents week fixed effects, and ϵ_{it} is the error term. Although fixed effects are not strictly necessary when analyzing the results of a randomized experiment, they can help obtain more precise estimates

[2]To determine whether the observed difference in this variable was genuine or due to multiple comparisons, we applied the Bonferroni correction. With 13 comparisons, the alpha level of 5% should be adjusted to $0.05/13 = 0.003$. The exact p-value for the "Avoid Feelings of Guilt" pairwise t-test was 0.0066, which is not significant under the corrected alpha level.

(Angrist & Pischke, 2009). Models without fixed effects yield similar results and are reported in Appendix B.

In line with previous literature (e.g., Kumar et al., 2010; Goli et al., 2022), we measure engagement by how frequently users log in and how much time they spend on the platform. Specifically, we use five different measures of engagement: (1) Login — a binary variable indicating whether a person has visited the platform on a given day, (2) Total Active Time — total time spent on the platform measured in minutes on a given day, (3 and 4) Learning/Practice Time — daily minutes spent on specific activities such as learning (e.g., reading summaries) or practice (e.g., reviewing questions from previous exams), and (5) Hardest Subject — daily minutes spent on learning and practicing activities related to the subject that users identified as most challenging.

Table 3 shows no significant effects of being allocated to a leaderboard on any of the engagement measurements. Our estimates are precise, with narrow confidence intervals. For example, being assigned to a leaderboard leads to less than 0.01 additional logins per day (out of an average of 10 daily logins), and to an increase in the total time spent on the platform by 0.01 standard deviations, with a 95% confidence interval of (-0.02, 0.04). These results are consistent with some of the literature reports where leaderboards have no discernible effects on behavioral outcomes (Schlömmer et al., 2021; Pedersen et al., 2017). Nevertheless, the literature suggests that the effects of leaderboards may be highly heterogeneous, and their effectiveness may depend on the user base and on the design of the leaderboard (e.g., Hydari et al., 2023; Ho et al., 2023; S. Song et al., 2021). We explore these two potential explanations in subsequent analyses.

4.3 Leaderboard Composition

Next, we analyze the composition of the leaderboard and test whether user similarity affects engagement. We operationalize user similarity with two metrics: leaderboard dispersion and competitor distance. Leaderboard dispersion is defined as the standard deviation of scores on the leaderboard.

Table 3: Panel Regression Results: Engagement as a Function of Experimental Group.

	Dependent variable:				
	Login	Total Active Time	Learning	Practice	Hardest Subject
	(1)	(2)	(3)	(4)	(5)
Leaderboard	0.008	0.009	0.008	0.002	0.007
	(0.006)	(0.016)	(0.014)	(0.020)	(0.015)
Time FE	Yes	Yes	Yes	Yes	Yes
Gender FE	Yes	Yes	Yes	Yes	Yes
Education FE	Yes	Yes	Yes	Yes	Yes
Observations	108,171	108,171	108,171	108,171	79,956
Adjusted R^2	0.046	0.017	0.015	0.005	0.003

Note: $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$
Standard errors clustered at user level. All continuous variables have been standardized by subtracting the value from the mean and dividing by the standard deviation.

A lower standard deviation at the start indicates more similar initial user engagement. Competitor distance is a local metric and measures the difference in points between a user's score and the scores of other nearby users on the leaderboard. These metrics provide insight into the level of competition a specific user faces at any given moment. The rationale for examining different compositions revolves around understanding how the distances between competitors influence perceptions of competition intensity and rivalry (Kilduff et al., 2010; Nebel, Schneider, & Rey, 2016; Bothner et al., 2007). These perceptions can, in turn, impact user outcomes (Santhanam et al., 2016).

Table 4 reveals that a one standard deviation increase in leaderboard score dispersion boosts user engagement by 9.2% of a standard deviation and platform visits by 3.5% of a standard deviation. This indicates that higher dispersion (or lower competition intensity) in leaderboard scores correlates with increased user engagement. To avoid endogeneity issues, the calculated leaderboard dispersion excludes the user's own score. In Table 5, we examine the distances to other competitors. Our findings consistently show that greater distances from competitors positively influence user engagement with the effect of a competitor's distance from below being much more pronounced (depending on the engagement metric, 4 to 6 times larger). This somewhat contradicts the economic theory of competitive crowding in tournaments, which suggests that having more competitors in closer proximity below (i.e., smaller distances to competitors below) should motivate an increase in performance (Bothner et al., 2007). However, it also aligns with the theory's assertion that crowding from below is more significant than from above, as individuals are more motivated to avoid losing their current rank than to improve it. This pattern may also be linked to user heterogeneity, the second potential factor behind the lack of leaderboard effects. If the platform's user base is generally less competitive, crowding from below could produce results that differ from traditional tournament theories, which typically assume highly competitive participants (e.g. NASCAR drivers).

Table 4: Panel Regression Results: Engagement as a Function of Leaderboard Competitive Dispersion.

	Dependent variable:				
	Login	Total Active Time	Learning	Practice	Hardest Subject
	(1)	(2)	(3)	(4)	(5)
Leaderboard Dispersion	0.035***	0.092***	0.081***	0.048***	0.051***
	(0.004)	(0.012)	(0.013)	(0.012)	(0.014)
PreTreatQuart Control	Yes	Yes	Yes	Yes	Yes
LdbPosit Control	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes
Gender FE	Yes	Yes	Yes	Yes	Yes
Education FE	Yes	Yes	Yes	Yes	Yes
Observations	65,919	65,919	65,919	65,919	49,403
Adjusted R^2	0.090	0.107	0.070	0.025	0.015

Note: *p<0.1; **p<0.05; ***p<0.01
Standard errors clustered at user level. All continuous variables have been standardized by subtracting the value from the mean and dividing by the standard deviation.

We further investigate whether this relationship holds when considering the distance to a competitor two (or more) positions above and below the focal user. For a competitor two positions below, the effect of points distance on total user activity remains significant ($\beta = 0.249$, $p < 0.01$), while the impact of a competitor two positions above disappears. Note that we always control for a user's leaderboard position to account for potential endogenous effects, considering that more engaged users are likely to end up in higher positions on the leaderboard and also at greater distances from other participants.

Overall, our results show that the distance to competitors below, rather than above, significantly

Table 5: Panel Regression Results: Engagement as a Function of Distances to Closest Competitors.

	Dependent variable:				
	Login	Total Active Time	Learning	Practice	Hardest Subject
	(1)	(2)	(3)	(4)	(5)
Distance to Competitor Above	0.013***	0.050***	0.047**	0.014	−0.002
	(0.005)	(0.017)	(0.023)	(0.017)	(0.015)
Distance to Competitor Below	0.058***	0.275***	0.277***	0.080**	0.218***
	(0.010)	(0.040)	(0.044)	(0.035)	(0.077)
PreTreatQuart Control	Yes	Yes	Yes	Yes	Yes
LdbPosit Control	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes
Gender FE	Yes	Yes	Yes	Yes	Yes
Education FE	Yes	Yes	Yes	Yes	Yes
Observations	59,915	59,915	59,915	59,915	44,947
Adjusted R^2	0.075	0.089	0.064	0.014	0.017

Note: $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$
Standard errors clustered at user level. All continuous variables have been standardized by subtracting the value from the mean and dividing by the standard deviation.

affects engagement. Additionally, the relationship between distance and engagement is positive: the farther away the competitors are, the more engaged the focal user becomes. To further understand these effects, we investigate how user heterogeneity influences leaderboard effectiveness.

4.4 Heterogeneity Analysis

The psychology literature discusses the heterogeneity in attitudes toward competition, with notable differences observed between females and males (Niederle & Vesterlund, 2007; Boudreau & Kaushik, 2023). Inspired by this, we examine whether the effects of leaderboards on our platform vary across different user groups. We consider five dimensions: gender, type of education, reported exam preparation, most difficult subject, and type of user. While most characteristics did not yield significant differences in how users responded to leaderboards, we uncovered a compelling finding related to competitively-oriented users—those who state the goal of "I want to be better than my friends" (Table 6). For this group, greater leaderboard dispersion and lower competition intensity actually led to lower engagement, contrary to the general trend where most users engage more under these conditions. Interestingly, these competitive users logged in more frequently when they were closer to their outperforming competitors and less frequently when they were farther away, highlighting a unique response pattern. Although this group is small (303 users), their distinct behavior suggests that our average treatment effects primarily reflect the responses of non-competitive users. This finding emphasizes the importance of tailoring leaderboard designs to accommodate different user motivations.

Table 6: Panel Regression Results: Engagement as a Function of Leaderboard Competitive Dispersion with Competitive User Type Moderator

	Dependent variable:				
	Login	Total Active Time	Learning	Practice	Hardest Subject
	(1)	(2)	(3)	(4)	(5)
Leaderboard Dispersion	0.037***	0.095***	0.083***	0.051***	0.051***
	(0.004)	(0.013)	(0.014)	(0.012)	(0.014)
Competitive User	0.022*	−0.006	0.005	−0.026	−0.031
	(0.011)	(0.028)	(0.029)	(0.025)	(0.031)
Dispersion x Competitive User	−0.026***	−0.047**	−0.032	−0.038**	−0.002
	(0.009)	(0.023)	(0.024)	(0.016)	(0.027)
PreTreatQuart Control	Yes	Yes	Yes	Yes	Yes
LdbPosit Control	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes
Gender FE	Yes	Yes	Yes	Yes	Yes
Education FE	Yes	Yes	Yes	Yes	Yes
Observations	65,919	65,919	65,919	65,919	49,403
Adjusted R^2	0.107	0.090	0.070	0.025	0.015

Note: *p<0.1; **p<0.05; ***p<0.01
Standard errors clustered at user level. All continuous variables have been standardized by subtracting the value from the mean and dividing by the standard deviation.

4.5 Alternative Measurements

We continue our analysis by exploring alternative measurements to ensure our results are robust across different specifications. First, we investigate the relationship between the distance to the first and last positions on the leaderboard and user engagement (Table 7). Consistent with our main findings, we observe that the distance to the first position has a negligible effect on user engagement, whereas the distance to the last position has a strong positive effect.

Next, we examine alternative measurements of leaderboard dispersion (Table 8). Specifically, we calculate the standard deviation of scores for users in neighboring positions. The results align with our primary dispersion measurement, showing that the more dispersed the scores of competitors directly above and below the focal user, the more engaged the user tends to be. Additionally, we evaluate the dispersion of scores in a relative leaderboard. While the full leaderboard consists of 20 users, the main dashboard typically displays a smaller set of scores: the top three positions, the focal user, and the two neighboring positions above and below. This relative leaderboard, being more immediately accessible, may have had a greater influence on the user compared to the full set of ranks, which required clicking the 'show full leaderboard' button. We find that an increase in the standard deviation of scores in this relative leaderboard similarly leads to an increase in daily user engagement.

Table 7: Panel Regression Results: Engagement as a Function of Distances to the First and Last Leaderboard Position.

	Dependent variable:				
	Login	Total Active Time	Learning	Practice	Hardest Subject
	(1)	(2)	(3)	(4)	(5)
Distance to First Position	0.020***	0.019***	0.018**	0.006	0.011
	(0.003)	(0.006)	(0.007)	(0.009)	(0.009)
Distance to Last Position	0.050***	0.298***	0.263***	0.163***	0.148***
	(0.005)	(0.022)	(0.028)	(0.046)	(0.042)
PreTreatQuart Control	Yes	Yes	Yes	Yes	Yes
LdbPosit Control	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes
Gender FE	Yes	Yes	Yes	Yes	Yes
Education FE	Yes	Yes	Yes	Yes	Yes
Observations	65,937	65,937	65,937	65,937	49,418
Adjusted R^2	0.127	0.110	0.098	0.037	0.022

Note: $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$
Standard errors clustered at user level. All continuous variables have been standardized by subtracting the value from the mean and dividing by the standard deviation.

Table 8: Panel Regression Results: Engagement as a Function of Dispersion of the Neighboring Positions and Relative Leaderboard.

	Dependent variable:									
	Login		Total Active Time		Learning		Practice		Hardest Subject	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Dispersion Position Above and Below	0.029***		0.127***		0.126***		0.037**		0.058**	
	(0.005)		(0.022)		(0.029)		(0.016)		(0.024)	
Dispersion of Local Leaderboard		0.027***		0.067***		0.060***		0.032***		0.038***
		(0.004)		(0.011)		(0.011)		(0.009)		(0.012)
PreTreatQuart Control	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
LdbPosit Control	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Gender FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Education FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	59,915	65,937	59,915	65,937	59,915	65,937	59,915	65,937	44,947	49,418
Adjusted R^2	0.068	0.088	0.088	0.105	0.057	0.068	0.013	0.024	0.013	0.015

Note: $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$
Standard errors clustered at user level. All continuous variables have been standardized by subtracting the value from the mean and dividing by the standard deviation.

5 Online Experiment

To better understand how leaderboards influence behavior, we conducted a follow-up online experiment designed to simulate our field study while addressing its limitations. In the original field experiment, we were unable to measure user perceptions and underlying behavioral mechanisms due to platform restrictions against implementing lengthy questionnaires out of concern for user experience. Additionally, to avoid deceiving users, we had to use real scores and opponents, which limited our ability to achieve perfect randomization in assigning participants to high and low-similarity leaderboards. The natural variability in scores and opponents across rounds further restricted our ability to effectively manipulate leaderboard similarity conditions. In contrast, the online experiment allows us to simulate opponents and apply perfectly randomized treatments, thereby overcoming these challenges and enabling us to more precisely test the impact of leaderboard user similarity on engagement.

5.1 Experimental Procedures and Measurement

In the online experiment, the treatment conditions either replicated those of the field experiment or focused on individual components of competition (e.g., performance feedback), allowing us to isolate and assess the behavioral mechanisms. Participants were randomly assigned to one of six groups: control, leaderboard with large or small score dispersion, social presence, performance feedback, or social comparison. In leaderboard treatments, opponents' scores were simulated based on the participant's score to create leaderboards with either large or small dispersion.

Participants were introduced to an anagram-solving task, chosen for its simplicity and cognitive challenge. Before the tasks, we collected data on anagram-solving skills, demographics, competitiveness, and goal orientation. Following this introductory part, each participant solved a set of 6 anagrams in 180 seconds. The tasks were interspersed with waiting periods during which participants were informed that they must wait for others to complete the tasks and for the results to be verified, mimicking real-world conditions and enhancing the realism of the environment. After each task, participants received their assigned treatment. This sequence was repeated three times. The survey concluded with questions measuring the mechanisms and whether participants used any additional tools. Key measurements are summarized in Table 9.

Table 9: Online Experiment: Main Measurements

Concept	Measurement	Description	Reference
Engagement	Score	avg words solved calculated for two post-treatment tasks	—
	Speed	avg seconds remaining from the time limit (180 s. per task) calculated for two post-treatment tasks	—
Competition Intensity		avg of 5-item survey instrument, included statements such as: "Other task participants were competitive about trying to do well"	Perceived Competition in Connelly et al. (2014)
Social Presence	Awareness of Others	avg of 4-item survey instrument, included statements such as: "I am aware of the other participants"	Social Presence two-component instrument in Kreijns et al. (2022)
	Proximity to Others	avg of 4-item survey instrument, included statements such as: "I constantly feel that the other participants are around"	
Performance Feedback		avg of 4-item survey instrument, included statements such as: "The feedback I received in-between the tasks helped me understand my performance"	Feedback Quality in Steelman et al. (2004)
Social Comparison		avg of 4-item survey instrument, included statements such as: "I have compared my task performance to others"	Social Comparison in Hanus & Fox (2015)

5.2 Data

The experiment was conducted through Prolific in November 2023. Of the 742 participants who completed the survey, we excluded 39 participants—four for failing both attention checks and 35 flagged as potential bots by reCAPTCHA bot detection—resulting in a final sample of 703 respondents. Table 10 presents the sample's summary statistics. Unlike the field study, where 70% of participants were female, our sample is gender-balanced. Approximately half of the participants identify as students, and nearly 80% rate their English proficiency as equivalent to that of native speakers. Participants' perceived preparation for the task (anagram solving) is higher than in the field sample but remains relatively low, with an average rating of 5 out of 10. A significant portion of respondents (45%) identify as predominantly knowledge-oriented, and 22% admit to using external tools (e.g., anagram solvers) during the task. Across the three task rounds, the average participant successfully solved 11.7 out of 18 words within 7.5 of the available 9 minutes.

Table 10: Summary Statistics: Demographics and Activity

Variable	N	Mean	Std. Dev.	Min	Pctl. 25	Pctl. 75	Max
Is Female	703	0.45	0.5	0	0	1	1
Is Student	703	0.47	0.5	0	0	1	1
Is EnglishC1C2	703	0.77	0.42	0	1	1	1
Task Preparation	703	5.5	1.87	0	4	7	10
Competitiveness	703	5.06	1.32	1	4.25	6	7
Perform Avoid Oriented	703	0.16	0.36	0	0	0	1
Perform Prove Oriented	703	0.11	0.31	0	0	0	1
Knowledge Oriented	703	0.45	0.5	0	0	1	1
Tools Used	703	0.22	0.42	0	0	0	1
Total Score	703	11.67	4.29	0	8	16	18
Total Time	703	456.61	100.42	113	387	540	608

With an ANOVA test, we confirm successful randomization with no significant pre-treatment differences (p-value above 0.05 for all available user characteristics). The manipulation check shows heightened competition in leaderboard groups, especially those with small dispersion/high intensity (pairwise comparison t-test between two leaderboard groups, $p = 0.022$). Although respondents recognized others' presence, they generally disagreed with feeling close to others (mean 2.99). Performance feedback was higher in leaderboard and feedback groups, and social comparison scores were notably higher in leaderboard groups than in social comparison group ($p \approx 0$). In conclusion, both the randomization and manipulation were successful.

5.3 Results

We start our analysis by assessing whether the field results replicate in the online context. The field study revealed that leaderboard presence had no observable effect on engagement. In Table 11, we use the same model specification and sample composition as in the field study, excluding the feedback, social presence, and social comparison groups, as these elements were not part of the original field experiment. Our analysis examines two engagement metrics available online: score and task-solving speed.

Exposure to a leaderboard significantly enhanced the speed of task completion. Participants in the leaderboard condition decreased their completion time (i.e., improved speed) by 35.6% of a standard deviation compared to the control group, even though it did not significantly impact

Table 11: OLS Regression Results: Performance as a Function of Experimental Group - Field Model

	Dependent variable:			
	Score		Speed	
	(1)	(2)	(3)	(4)
Leaderboard	0.053	0.067	0.324***	0.356***
	(0.112)	(0.087)	(0.117)	(0.093)
Score Before Treatment		0.628***		
		(0.042)		
Speed Before Treatment				0.668***
				(0.047)
Task Preparation		0.101**		0.009
		(0.042)		(0.045)
Constant	0.007	−0.0005	−0.133	−0.106
	(0.092)	(0.071)	(0.095)	(0.076)
Observations	352	352	352	352
Adjusted R^2	−0.002	0.404	0.019	0.380

Note: $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$
All continuous variables have been standardized by subtracting the value from the mean and dividing by the standard deviation.

task scores. On the one hand, this result is consistent with the field study, where we observed no significant leaderboard effects on engagement. On the other hand, the speed improvement is a new finding that was not present in the field study. The online sample is more competitive, as shown by their higher competitiveness scores, which might explain the positive effect of leaderboards on task speed. Additionally, the online study's task can be challenging, making it easier to improve speed rather than score. We conduct further analysis using the group allocation variable (with the control group as the baseline) and examine the effects across the entire online experimental sample. The detailed results of this analysis are presented in Table 12.

We observe that while different treatment assignments do not impact task scores, both types of leaderboard treatments—whether large or small dispersion—significantly improve speed. The treatment with small dispersion and high competition intensity has an effect that is twice as pronounced as the low-intensity treatment ($p < 0.01$). These findings contrast with the initial field study, which showed no notable effects or even negative outcomes related to competition intensity, such as score dispersion and competitor distance. Additionally, receiving performance feedback enhances task speed, although with a marginal statistical significance ($p < 0.1$). Neither the social comparison nor the social presence treatments show significant effects. The difference between the

Table 12: OLS Regression Results: Performance as a Function of Experimental Group

	Dependent variable:			
	Score		Speed	
	(1)	(2)	(3)	(4)
Treatment LLarge	0.051	0.101	0.208	0.218**
	(0.131)	(0.097)	(0.129)	(0.103)
Treatment LSmall	0.054	0.037	0.439***	0.492***
	(0.130)	(0.097)	(0.129)	(0.103)
Treatment SP	−0.094	−0.100	−0.090	−0.122
	(0.131)	(0.097)	(0.129)	(0.103)
Treatment PF	0.079	0.065	0.234*	0.196*
	(0.131)	(0.097)	(0.129)	(0.103)
Treatment SC	−0.133	−0.089	0.009	−0.104
	(0.130)	(0.097)	(0.129)	(0.103)
Score Before Treatment		0.641***		
		(0.028)		
Speed Before Treatment				0.589***
				(0.030)
Task Preparation		0.121***		0.055*
		(0.028)		(0.030)
Constant	0.007	−0.002	−0.133	−0.113
	(0.092)	(0.068)	(0.091)	(0.072)
Observations	703	703	703	703
Adjusted R^2	−0.001	0.448	0.025	0.381

Note: $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$
Groups: LLarge - Leaderboard Large Dispersion, *LSmall* - Leaderboard Small Dispersion, *SP* - Social Presence, *PF* - Performance Feedback, *SC* - Social Comparison. All continuous variables have been standardized by subtracting the value from the mean and dividing by the standard deviation.

field and online results could be attributed to varying user profiles: the field study involved students under exam pressure, whereas our online experiment appears to attract more competitively inclined individuals. Notably, the average competitiveness index is 5.25 out of 7, indicating participants tend to be competitive. Next, we revisit the heterogeneity analysis and explore the underlying mechanisms to better understand the reasons for this discrepancy.

5.4 Heterogeneity Analysis

Our field experiment revealed that smaller dispersion and higher competition intensity benefited competitively inclined users. Online, we noticed a significant interaction between small-dispersion treatment effects and the user's competitiveness index score (see Table 13, $\beta = 0.237$). More competitive users notably increased their task-solving speed when placed on a leaderboard with small score dispersion, compared to their less competitive counterparts. Sub-sample analysis further validates these findings, showing that the positive engagement effects of being in a small-dispersion/high-intensity leaderboard group were confined to highly competitive respondents ($\beta_{high-intensity} = 0.570$, $p < 0.01$), while non-competitive respondents showed no significant effects ($\beta_{high-intensity} = 0.040$, $p > 0.1$). Conversely, the positive effects of low-intensity leaderboards were exclusive to less competitive users ($\beta_{low-intensity} = 0.351$, $p < 0.05$), with no significant effects for competitive respondents ($\beta_{high-intensity} = 0.063$, $p > 0.1$). These outcomes highlight the discrepancy between field and online experiments, which stem from differences in the proportion of competitive users in the samples. Specifically, the field sample is dominated by non-competitively oriented users (as observed by the student type question), whereas the online sample is more balanced with respect to user competitiveness, with an average user scoring a 5 out of 7 on the scale of competitiveness.

Table 13: OLS Regression Results: Performance as a Function of Experimental Group - Competitiveness Heterogeneity

	Dependent variable:			
	Score		Speed	
	(1)	(2)	(3)	(4)
Leaderboard	0.052	0.098	0.210	0.216**
	(0.128)	(0.100)	(0.132)	(0.105)
Small Dispersion	0.039	−0.044	0.274**	0.305***
	(0.129)	(0.100)	(0.132)	(0.105)
Competitiveness Index Score	0.067	0.012	0.137**	0.013
	(0.065)	(0.053)	(0.067)	(0.056)
Score Before Treatment		0.621***		
		(0.042)		
Speed Before Treatment				0.661***
				(0.047)
Task Preparation		0.080*		−0.017
		(0.044)		(0.046)
Small Dispersion x Compet. Indx	0.208**	0.151*	0.182*	0.237***
	(0.104)	(0.081)	(0.107)	(0.085)
Constant	0.006	0.001	-0.137	-0.105
	(0.090)	(0.070)	(0.093)	(0.074)
Observations	352	352	352	352
Adjusted R^2	0.024	0.411	0.069	0.410

Note: $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$
All continuous variables have been standardized by subtracting the value from the mean and dividing by the standard deviation.

5.5 Mechanisms

In Tables 14 and 15, we present the results of our mechanism analysis. We begin by examining the direct effects of the mechanisms, utilizing either a model with the composite mechanism of competition intensity or models with the three distinct components of competition: social presence, performance feedback, and social comparison (Table 14). For social presence measurement, we calculate two separate factors, awareness of others and proximity to others (in line with the literature, see Kreijns et al., 2020, 2022).

Our findings indicate that competition intensity has a significant and positive effect on both engagement measurements (columns 2 and 6). When we decompose competition into its three component mechanisms, the positive effects of competition are shown to originate primarily from performance feedback, whereas the other mechanisms tend to exert mostly negative, albeit statistically insignificant, effects. To calculate the indirect, mediator effects, we employ the causal counterfactual framework developed by Imai et al. (2010) and employed by other IS scholars to analyze mediation (e.g., Li et al., 2022). The framework allows us to estimate both direct and indirect treatment effects. The indirect effect measures the impact of a mediator being changed from the value it would take under the control condition to the value it would have under the leaderboard treatment condition while keeping the treatment status fixed. The direct effect, on the other hand, calculates the impact of the treatment status being changed, while the mediator is held at the value that would occur under the no-treatment exposure. To calculate these effects, we estimate two equations, mediator and outcome:

$$E[M \mid a, c] = \beta_0 + \beta_1 a + \beta_2' c \tag{2}$$

$$\begin{aligned} E[Y \mid a, M, c] = \theta_0 + \theta_1 * a + \theta_2 * m \\ + \theta_3 * a * m + \theta_4' * c \end{aligned} \tag{3}$$

In Table 15, we observe that competition intensity serves as a mediator of engagement in terms of score, but not in terms of task completion speed ($p < 0.1$), where it exhibits a direct effect

Table 14: OLS Regression Results: Performance as a Function of Mechanisms

	Dependent variable:							
	Score				Speed			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Competition Intensity	0.099***	0.076***			0.100***	0.094***		
	(0.038)	(0.028)			(0.038)	(0.031)		
Score Before Treatment		0.641***		0.638***				
		(0.028)		(0.028)				
Speed Before Treatment						0.576***		0.574***
						(0.031)		(0.030)
Task Preparation		0.113***		0.110***		0.044		0.028
		(0.028)		(0.028)		(0.031)		(0.031)
Awareness of Others			0.030	0.009			−0.057	−0.017
			(0.044)	(0.033)			(0.044)	(0.036)
Proximity to Others			−0.043	−0.028			0.053	0.046
			(0.044)	(0.033)			(0.044)	(0.036)
Feedback			0.166***	0.143***			0.134***	0.110***
			(0.043)	(0.032)			(0.043)	(0.035)
Social Comparison			0.015	−0.030			0.074	0.061*
			(0.045)	(0.034)			(0.045)	(0.037)
Constant	−0.000	−0.000	−0.000	−0.000	−0.000	−0.000	−0.000	−0.000
	(0.038)	(0.028)	(0.037)	(0.028)	(0.038)	(0.034)	(0.030)	(0.030)
Observations	703	703	703	703	703	703	703	703
Adjusted R^2	0.008	0.451	0.024	0.460	0.009	0.347	0.030	0.362

Note: *p<0.1; **p<0.05; ***p<0.01
All continuous variables have been standardized by subtracting the value from the mean and dividing by the standard deviation.

instead. Additionally, we find that among the component mechanisms, feedback mediates the treatment's effect on both score and speed. The other two mechanisms—social presence and social comparison—might have direct effects on task completion speed, but they do not influence score performance nor do they mediate the effects (tables with detailed results are available upon request).

6 Conclusions

In this paper, we examine the impact of leaderboards as a gamification element on user engagement within an online learning platform. Contrary to expectations, we found no significant difference in engagement between users included in a leaderboard and those who were not. However, among users assigned to a leaderboard, those with greater score dispersion—indicating lower competition intensity—exhibited higher engagement levels. This effect was contingent upon the users' competitiveness. Specifically, non-competitive users responded positively to lower competition intensity, while competitive users showed the opposite trend. Furthermore, we found that the effects of the

Table 15: Mediation Analysis: Engagement as a Function of Treatment Mediated by Competition Intensity and Performance Feedback

Component	**Estimate**	**95% CI Lower**	**95% CI Upper**	**p-value**
		Mediator: Competition Intensity		
		DV: Score		
ACME	0.090	0.014	0.18	0.024**
ADE	-0.024	-0.197	0.14	0.820
Total Effect	0.067	-0.087	0.23	0.442
Prop. Mediated	0.724	-9.534	10.94	0.446
		DV: Speed		
ACME	0.0755	-0.0086	0.17	0.082*
ADE	0.2873	0.0858	0.47	0.002***
Total Effect	0.3628	0.1803	0.53	0.000***
Prop. Mediated	0.2002	-0.0248	0.60	0.082*
		Mediator: Performance Feedback		
		DV: Score		
ACME	0.243	0.086	0.40	0.002***
ADE	-0.174	-0.393	0.06	0.136
Total Effect	0.069	-0.089	0.23	0.420
Prop. Mediated	1.973	-24.533	30.21	0.422
		DV: Speed		
ACME	0.178	0.026	0.33	0.014**
ADE	0.181	-0.037	0.41	0.110
Total Effect	0.359	0.188	0.55	0.000***
Prop. Mediated	0.496	0.078	1.18	0.014**

Note: *p<0.1; **p<0.05; ***p<0.01
Quasi-Bayesian Confidence Intervals Based on 1000 Monte Carlo Simulations

leaderboards were mediated by the mechanisms of competition and performance feedback.

Our research contributes to the existing literature on the impact of leaderboards on engagement in online platforms. While earlier studies have underscored the leaderboards' potential to enhance learning and task performance, more recent research suggests that their effects are contingent on the design (Bai et al., 2020; H. Song et al., 2013). Our study addresses this variability by creating leaderboards with different user compositions and examining how different configurations affect user engagement on a learning platform. By examining leaderboards as a single design element and conducting a randomized field experiment in a real-world online platform, we provide valuable insights into the role of leaderboards in mitigating common platform challenges like declining user engagement. Furthermore, we explore the mechanisms underlying the effectiveness of leaderboards in influencing user behavior (Liu et al., 2017; Wu et al., 2015). We hypothesize that leaderboards serve three distinct functions: conveying social presence information, providing performance feedback, and offering social comparison cues. These functions collectively foster a competitive environment. Our experimental design isolates and examines the effects of each function, both individually and in combination. Finally, we corroborate existing research (To et al., 2020; Amo et al., 2020; Theriault et al., 2021), which emphasizes the importance of user characteristics in understanding the effects of leaderboards. Our findings confirm the heterogeneity of leaderboard effects across different tasks and user groups (Hydari et al., 2023; Bojd et al., 2022). In both experiments, individual traits such as competitiveness moderate the impact of leaderboards on outcomes, highlighting the necessity of tailoring gamification approaches to the characteristics of the platform's users.

Our study offers valuable guidance to online platforms working on developing effective gamified interventions. First, the effectiveness of leaderboards depends more on their design than their mere presence. Factors such as the structuring of the opponents and their (starting) scores significantly impact engagement. Second, personalizing leaderboard designs based on user traits, like competitiveness, is crucial, particularly on platforms with diverse user bases. Customizing competition intensity could improve satisfaction and engagement. Lastly, as performance feedback drives engagement, platforms should focus on providing clear, frequent, and meaningful feedback through leaderboards.

The limitations of this study provide interesting avenues for future research. First, our experiments do not account for the influence of social networks or the role of peer dynamics in competitive environments as the leaderboard participants do not know each other in person. Future research could explore how social connections moderate the influence of leaderboards and their design on user engagement. Second, we focus on one gamification design element, leaderboard, while multiple gamification elements are often implemented simultaneously. Future research could explore the effectiveness of combining leaderboards with a specific design (e.g. high score dispersion leaderboards) with other gamification features and their designs. Third, while our study tested leaderboards with different levels of score dispersion, there exist many other potentially noteworthy leaderboard design features, such as length of competition, point structure, or option of collaboration between participants that can further affect the effectiveness of leaderboards in stirring up user engagement.

References

Amo, L., Liao, R., Kishore, R., & Rao, H. R. (2020, November). Effects of structural and trait competitiveness stimulated by points and leaderboards on user engagement and performance growth: A natural experiment with gamification in an informal learning environment. *European Journal of Information Systems*, *29*(6), 704–730.

Angrist, J. D., & Pischke, J.-S. (2009). *Mostly harmless econometrics: An empiricist's companion.* Princeton University Press.

Bai, S., Hew, K. F., & Huang, B. (2020, June). Does gamification improve student learning outcome? Evidence from a meta-analysis and synthesis of qualitative data in educational contexts. *Educational Research Review*, *30*, 100322.

Bai, S., Hew, K. F., Sailer, M., & Jia, C. (2021, November). From top to bottom: How positions on different types of leaderboard may affect fully online student learning performance, intrinsic motivation, and course engagement. *Computers & Education*, *173*, 104297.

Baldauf, S. A., Engqvist, L., & Weissing, F. J. (2014). Diversifying evolution of competitiveness. *Nature Communications*, *5*(1), 5233.

Banerjee, A. V., & Duflo, E. (2014). (dis) organization and success in an economics mooc. *American Economic Review*, *104*(5), 514–18.

Bojd, B., Song, X., Tan, Y., & Yan, X. (2022, June). Gamified Challenges in Online Weight-Loss Communities. *Information Systems Research*, *33*(2), 718–736.

Bothner, M. S., Kang, J.-h., & Stuart, T. E. (2007, June). Competitive Crowding and Risk Taking in a Tournament: Evidence from NASCAR Racing. *Administrative Science Quarterly*, *52*(2), 208–247.

Boudreau, K. J., & Kaushik, N. (2023, January). Gender Differences in Responses to Competitive Organization? A Field Experiment on Differences Between STEM and Non-STEM Fields from an Internet-of-Things Platform. *Organization Science*, *34*(6), 2119–2142.

Bruchmann, K., Chue, S. M., Dillon, K., Lucas, J. K., Neumann, K., & Parque, C. (2021, September). Social Comparison Information Influences Intentions to Reduce Single-Use Plastic Water Bottle Consumption. *Frontiers in Psychology*, *12*, 612662.

Burguillo, J. C. (2010). Using game theory and competition-based learning to stimulate student motivation and performance. *Computers & Education*, *55*(2), 566–575.

Chan, E., & Briers, B. (2019, August). It's the End of the Competition: When Social Comparison Is Not Always Motivating for Goal Achievement. *Journal of Consumer Research*, *46*(2), 351–370.

Chen, Y., Harper, F. M., Konstan, J., & Li, S. X. (2010, September). Social Comparisons and Contributions to Online Communities: A Field Experiment on MovieLens. *American Economic Review*, *100*(4), 1358–1398.

CITO. (2021). *Verslag van de examencampagne 2021 voortgezet onderwijs* (Tech. Rep.). Cito Instituut voor Toetsontwikkeling. Retrieved from `https://cito.nl/media/mj2pveav/cito`$_{v}erslag_{e}xamencampagne_{2}021.pdf$

Connelly, C. E., Ford, D. P., Turel, O., Gallupe, B., & Zweig, D. (2014). 'i'm busy (and competitive)!'antecedents of knowledge sharing under pressure. *Knowledge Management Research & Practice*, *12*(1), 74–85.

Davis, D., Jivet, I., Kizilcec, R. F., Chen, G., Hauff, C., & Houben, G.-J. (2017, March). Follow the successful crowd: Raising MOOC completion rates through social comparison at scale. In *Proceedings of the Seventh International Learning Analytics & Knowledge Conference* (pp. 454–463). ACM.

Ding, L., Er, E., & Orey, M. (2018). An exploratory study of student engagement in gamified online discussions. *Computers & Education*, *120*, 213–226.

Garcia, S. M., & Tor, A. (2009, July). The N-Effect: More Competitors, Less Competition. *Psychological Science*, *20*(7), 871–877.

Goli, A., Chintagunta, P. K., & Sriram, S. (2022, February). Effects of Payment on User Engagement in Online Courses. *Journal of Marketing Research*, *59*(1), 11–34.

Hanus, M. D., & Fox, J. (2015). Assessing the effects of gamification in the classroom: A longitudinal study on intrinsic motivation, social comparison, satisfaction, effort, and academic performance. *Computers & Education*, *80*, 152–161.

Harackiewicz, J. M. (1979). The effects of reward contingency and performance feedback on intrinsic motivation. *Journal of Personality and Social Psychology*, *37*(8), 1352.

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, *77*(1), 81–112.

Hew, K. F., & Cheung, W. S. (2014). Students' and instructors' use of massive open online courses (moocs): Motivations and challenges. *Educational Research Review*, *12*, 45–58.

Ho, Y.-J. I., Liu, S., & Wang, L. (2023, June). Fun Shopping: A Randomized Field Experiment on Gamification. *Information Systems Research*, *34*(2), 766–785.

Höllig, C. E., Tumasjan, A., & Welpe, I. M. (2020). Individualizing gamified systems: The role of trait competitiveness and leaderboard design. *Journal of Business Research*, *106*, 288–303.

Huang, N., Burtch, G., Gu, B., Hong, Y., Liang, C., Wang, K., … Yang, B. (2019, January). Motivating User-Generated Content with Performance Feedback: Evidence from Randomized Field Experiments. *Management Science*, *65*(1), 327–345.

Huang, N., Zhang, J., Burtch, G., Li, X., & Chen, P. (2021). Combating procrastination on massive online open courses via optimal calls to action. *Information Systems Research*, *32*(2), 301–317.

Hydari, M. Z., Adjerid, I., & Striegel, A. D. (2023, July). Health Wearables, Gamification, and Healthful Activity. *Management Science*, *69*(7), 3920–3938.

Imai, K., Keele, L., & Tingley, D. (2010). A general approach to causal mediation analysis. *Psychological Methods*, *15*(4), 309–334.

Karlin, B., Zinger, J. F., & Ford, R. (2015). The effects of feedback on energy conservation: A meta-analysis. *Psychological Bulletin*, *141*(6), 1205.

Kilduff, G. J., Elfenbein, H. A., & Staw, B. M. (2010). The psychology of rivalry: A relationally dependent analysis of competition. *Academy of Management Journal*, *53*(5), 943–969.

Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: a historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin*, *119*(2), 254.

Kreijns, K., Bijker, M., & Weidlich, J. (2020). A rasch analysis approach to the development and validation of a social presence measure. *Rasch measurement: applications in quantitative educational research*, 197–221.

Kreijns, K., Xu, K., & Weidlich, J. (2022, March). Social Presence: Conceptualization and Measurement. *Educational Psychology Review*, *34*(1), 139–170.

Kumar, V., Aksoy, L., Donkers, B., Venkatesan, R., Wiesel, T., & Tillmanns, S. (2010). Undervalued or overvalued customers: Capturing total customer engagement value. *Journal of Service Research*, *13*(3), 297–310.

Landers, R. N., & Landers, A. K. (2014). An empirical test of the theory of gamified learning: The effect of leaderboards on time-on-task and academic performance. *Simulation & Gaming*, *45*(6), 769–785.

Leung, A. C. M., Santhanam, R., Kwok, R. C.-W., & Yue, W. T. (2022, March). Could Gamification Designs Enhance Online Learning Through Personalization? Lessons from a Field Experiment. *Information Systems Research*, *34*(1), 27–49.

Li, X., Grahl, J., & Hinz, O. (2022, June). How Do Recommender Systems Lead to Consumer Purchases? A Causal Mediation Analysis of a Field Experiment. *Information Systems Research*, *33*(2), 620–637.

Liu, D., Santhanam, R., University of Oklahoma, Webster, J., & Queen's University. (2017, April). Toward Meaningful Engagement: A Framework for Design and Research of Gamified Information Systems. *MIS Quarterly*, *41*(4), 1011–1034.

Miao, J., & Ma, L. (2022). Students' online interaction, self-regulation, and learning engagement in higher education: The importance of social presence to online learning. *Frontiers in Psychology*, *13*, 815220.

Molinillo, S., Aguilar-Illescas, R., Anaya-Sánchez, R., & Vallespín-Arán, M. (2018). Exploring the impacts of interactions, social presence and emotional engagement on active collaborative learning in a social web-based environment. *Computers & Education*, *123*, 41–52.

Nebel, S., Beege, M., Schneider, S., & Rey, G. D. (2016, December). The higher the score, the higher the learning outcome? Heterogeneous impacts of leaderboards and choice within educational videogames. *Computers in Human Behavior*, *65*, 391–401.

Nebel, S., Schneider, S., Beege, M., & Rey, G. D. (2017). Leaderboards within educational videogames: The impact of difficulty, effort and gameplay. *Computers & Education*, *113*, 28–41.

Nebel, S., Schneider, S., & Rey, G. D. (2016). From duels to classroom competition: Social competition and learning in educational videogames within different group sizes. *Computers in Human Behavior*, *55*, 384–398.

Niederle, M., & Vesterlund, L. (2007). Do women shy away from competition? do men compete too much? *The Quarterly Journal of Economics*, *122*(3), 1067–1101.

Niederle, M., & Vesterlund, L. (2011). Gender and competition. *Annual Review of Economics*, *3*(1), 601–630.

Park, S. Y., & Baek, Y. M. (2018). Two faces of social comparison on facebook: The interplay between social comparison orientation, emotions, and psychological well-being. *Computers in Human Behavior*, *79*, 83–93.

Pedersen, M. K., Rasmussen, N. R., Sherson, J. F., & Basaiawmoit, R. V. (2017). Leaderboard effects on player performance in a citizen science game. *arXiv preprint arXiv:1707.03704*.

Philpott, A., & Son, J.-B. (2022). Leaderboards in an efl course: Student performance and motivation. *Computers & Education*, *190*, 104605.

Santhanam, R., Liu, D., & Shen, W. (2016). Research note—gamification of technology-mediated training: Not all competitions are the same. *Information Systems Research*, *27*(2), 453–465.

Schlömmer, M., Spieß, T., & Schlögl, S. (2021). Leaderboard positions and stress—experimental investigations into an element of gamification. *Sustainability*, *13*(12), 6608.

Song, H., Kim, J., Tenzek, K. E., & Lee, K. M. (2013). The effects of competition and competitiveness upon intrinsic motivation in exergames. *Computers in Human Behavior*, *29*(4), 1702–1708.

Song, S., Han, K., & Animesh, A. (2021). How Am I Doing? The Impact of Localized Leaderboards in Digital Platforms. In *ICIS 2021 Proceedings.*

Srivastava, S. C., & Chandra, S. (2018, September). Social Presence in Virtual World Collaboration: An Uncertainty Reduction Perspective Using a Mixed Methods Approach. *MIS Quarterly*, *42*(3), 779–803.

Steelman, L. A., Levy, P. E., & Snell, A. F. (2004). The feedback environment scale: Construct definition, measurement, and validation. *Educational and Psychological Measurement*, *64*(1), 165–184.

Theriault, M., Ruel, T., Léger, P.-M., & Plante, J.-F. (2021). Can competition though leaderboards lead to better engagement and learning of data science concepts? an experimental study. In *Amcis 2021 proceedings.*

To, C., Kilduff, G. J., & Rosikiewicz, B. L. (2020). When interpersonal competition helps and when it harms: An integration via challenge and threat. *Academy of Management Annals*, *14*(2), 908–934.

Tsay, C. H.-H., Kofinas, A., & Luo, J. (2018). Enhancing student learning experience with technology-mediated gamification: An empirical study. *Computers & Education*, *121*, 1–17.

Vogel, E. A., Rose, J. P., Roberts, L. R., & Eckles, K. (2014, October). Social comparison, social media, and self-esteem. *Psychology of Popular Media Culture*, *3*(4), 206–222.

Wisniewski, B., Zierer, K., & Hattie, J. (2020). The power of feedback revisited: A meta-analysis of educational feedback research. *Frontiers in Psychology*, *10*, 3087.

Wu, Y., Kankanhalli, A., & Huang, K.-w. (2015). Gamification in fitness apps: How do leaderboards influence exercise? In *Icis 2015 proceedings.*

Yang Gao, Huaxia Rui, & Shujing Sun. (2023, September). The Power of Identity Cues in Text-Based Customer Service: Evidence from Twitter. *MIS Quarterly*, *47*(3), 983–1014.

A Field Experiment - Experimental Design Full Description

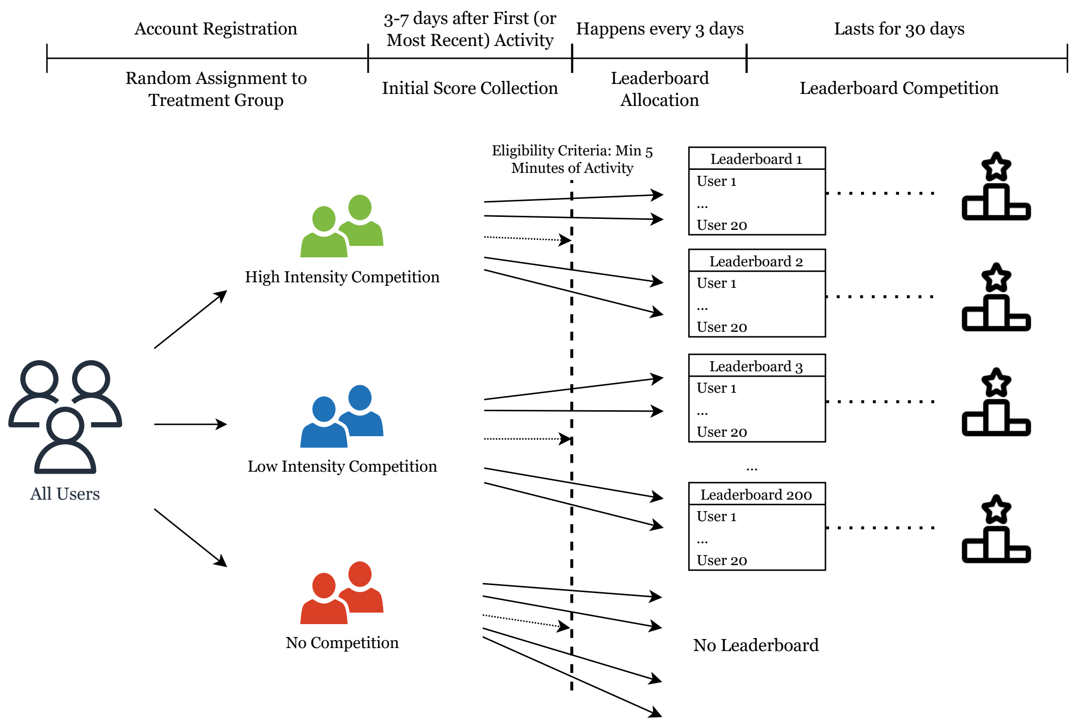


Figure 4: The Experimental Process.

At registration, each user was randomly assigned to one of three experimental groups: no competition, low competition intensity, or high competition intensity. In the *no competition* condition, users are not allocated to a leaderboard and are not ranked against others.

For the *low* and *high* competition intensity groups, we monitor users' engagement in their first three to seven days on the platform and use this information to allocate them to leaderboards that reflect the level of competition they were randomly assigned to. This procedure allows us to create the desired level of competition while maintaining the randomness of the treatment. To be considered for a leaderboard, new users must have registered at least 3 days ago and engaged with the platform within the past 7 days. We refer to these as *the leaderboard eligibility criteria*. In each leaderboard treatment group, multiple leaderboards of maximum 20 users are created in every allocation round.[3]

Every three days, an algorithm verifies whether each user who has not yet been allocated to a leaderboard meets these criteria. If they do, the user is assigned to a leaderboard that corresponds to their treatment group. Every user has an equal chance of receiving each of the three treatments, regardless of their initial engagement or any other individual characteristics. The decision to observe engagement within a brief period (3-7 days) balances the need for sufficient platform interaction observation and the avoidance of prolonged waiting times for treatment. This strategy aims to prevent potential loss of user interest before the treatment takes effect.

We create leaderboards in such a way that all users in a specific leaderboard have been randomly assigned to the same level of competition intensity. We operationalize the structure of competition by manipulating the similarity of users based on their level of engagement on the platform. Building on what Kilduff et al. (2010) hypothesized about the relationship between the intensity of rivalry and competitors' similarity, we assume that competition intensity will be higher in leaderboards where users are similar in their initial engagement levels. The underlying assumption is that users' initial levels of engagement remain relatively consistent during the competition—a user who was highly engaged in the first few days after creating an account will likely remain highly engaged compared to other users. In leaderboards with high competition intensity, users have similar levels of engagement, whereas in leaderboards with low competition intensity, users have disparate levels of engagement. For instance, in a low-intensity competitive structure, users are matched with others of varying engagement levels, as measured by their activity before allocation. Conversely, in a high-intensity competitive structure, users are paired with others who have similar engagement levels. Figure 4 depicts the complete experimental procedure.

The scores on the leaderboards are based on two elements: time engagement and performance.

[3]For example, if 74 users were eligible for assignment in one allocation round in the high intensity group, 4 leaderboards would be created.

A user earns 10 points for every minute in the platform, and between 40 and 640 points for each completed quiz, depending on the ratio of correct answers.[4]

B Main Models without Fixed Effects

Table 16: Panel Regression Results: Engagement as a Function of Experimental Group.

	Dependent variable:				
	Login	Total Active Time	Learning	Practice	Hardest Subject
	(1)	(2)	(3)	(4)	(5)
Leaderboard	0.011*	0.016	0.014	0.006	0.008
	(0.006)	(0.017)	(0.015)	(0.020)	(0.015)
Time FE	No	No	No	No	No
Gender FE	No	No	No	No	No
Education FE	No	No	No	No	No
Observations	108,171	108,171	108,171	108,171	79,956
Adjusted R^2	0.0001	0.00004	0.00003	−0.00000	0.00000

Note: $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$
Standard errors clustered at user level. All continuous variables have been standardized by subtracting the value from the mean and dividing by the standard deviation.

Table 17: Panel Regression Results: Engagement as a Function of Leaderboard Competitive Dispersion.

	Dependent variable:				
	Login	Total Active Time	Learning	Practice	Hardest Subject
	(1)	(2)	(3)	(4)	(5)
Leaderboard Dispersion	0.021***	0.073***	0.065***	0.039***	0.035***
	(0.003)	(0.010)	(0.011)	(0.009)	(0.012)
PreTreatQuart Control	Yes	Yes	Yes	Yes	Yes
LdbPosit Control	Yes	Yes	Yes	Yes	Yes
Time FE	No	No	No	No	No
Gender FE	No	No	No	No	No
Education FE	No	No	No	No	No
Observations	65,919	65,919	65,919	65,919	49,403
Adjusted R^2	0.060	0.077	0.059	0.020	0.011

Note: $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$
Standard errors clustered at user level. All continuous variables have been standardized by subtracting the value from the mean and dividing by the standard deviation.

[4] The score obtained in a quiz includes a correction for the guessing factor, meaning a student will receive points for quiz activity only after they reach a certain level of correct answers in relation to the total number of questions answered. Most quizzes consist of 40 multiple-choice questions.

Table 18: Panel Regression Results: Engagement as a Function of Leaderboard Competitive Distances.

	Dependent variable:				
	Login	Total Active Time	Learning	Practice	Hardest Subject
	(1)	(2)	(3)	(4)	(5)
Distance to Competitor Above	0.015***	0.056***	0.057**	0.010	0.013
	(0.005)	(0.018)	(0.024)	(0.018)	(0.015)
Distance to Competitor Below	0.025***	0.198***	0.181***	0.098***	0.089***
	(0.005)	(0.020)	(0.019)	(0.035)	(0.022)
PreTreatQuart Control	Yes	Yes	Yes	Yes	Yes
LdbPosit Control	Yes	Yes	Yes	Yes	Yes
Time FE	No	No	No	No	No
Gender FE	No	No	No	No	No
Education FE	No	No	No	No	No
Observations	63,450	63,450	63,450	63,450	47,555
Adjusted R^2	0.063	0.100	0.079	0.026	0.014

Note: *p<0.1; **p<0.05; ***p<0.01
Standard errors clustered at user level. All continuous variables have been standardized by subtracting the value from the mean and dividing by the standard deviation.

Table 19: Panel Regression Results: Engagement as a Function of Leaderboard Competitive Dispersion with Competitive User Type Moderator

	Dependent variable:				
	Login	Total Active Time	Learning	Practice	Hardest Subject
	(1)	(2)	(3)	(4)	(5)
Leaderboard Dispersion	0.023***	0.077***	0.068***	0.042***	0.036***
	(0.004)	(0.011)	(0.012)	(0.010)	(0.013)
Better Than Friends	0.024**	−0.006	0.002	−0.022	−0.031
	(0.012)	(0.029)	(0.030)	(0.025)	(0.031)
Dispersion x Better Than Friends	−0.035***	−0.055**	−0.040	−0.041***	−0.005
	(0.009)	(0.023)	(0.025)	(0.016)	(0.026)
PreTreatQuart Control	Yes	Yes	Yes	Yes	Yes
LdbPosit Control	Yes	Yes	Yes	Yes	Yes
Time FE	No	No	No	No	No
Gender FE	No	No	No	No	No
Education FE	No	No	No	No	No
Observations	65,919	65,919	65,919	65,919	49,403
Adjusted R^2	0.061	0.077	0.059	0.020	0.011

Note: *p<0.1; **p<0.05; ***p<0.01
Standard errors clustered at user level. All continuous variables have been standardized by subtracting the value from the mean and dividing by the standard deviation.

Table 20: Panel Regression Results: Engagement as a Function of Distances to the First and Last Leaderboard Position.

	Dependent variable:				
	Login	Total Active Time	Learning	Practice	Hardest Subject
	(1)	(2)	(3)	(4)	(5)
Distance to First Position	0.009***	−0.005	−0.003	−0.005	−0.004
	(0.003)	(0.005)	(0.006)	(0.009)	(0.008)
Distance to Last Position	0.040***	0.275***	0.244***	0.150***	0.133***
	(0.005)	(0.022)	(0.027)	(0.044)	(0.040)
PreTreatQuart Control	Yes	Yes	Yes	Yes	Yes
LdbPosit Control	Yes	Yes	Yes	Yes	Yes
Time FE	No	No	No	No	No
Gender FE	No	No	No	No	No
Education FE	No	No	No	No	No
Observations	65,937	65,937	65,937	65,937	49,418
Adjusted R^2	0.064	0.113	0.086	0.032	0.018

Note: $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$
Standard errors clustered at user level. All continuous variables have been standardized by subtracting the value from the mean and dividing by the standard deviation.

Table 21: Panel Regression Results: Engagement as a Function of Dispersion of the Neighboring Positions and Relative Leaderboard.

	Dependent variable:									
	Login		Total Active Time		Learning		Practice		Hardest Subject	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Dispersion Position Above and Below	0.024***		0.120***		0.120***		0.034**		0.053**	
	(0.005)		(0.021)		(0.028)		(0.015)		(0.023)	
Dispersion of Local Leaderboard		0.017***		0.058***		0.053***		0.028***		0.030***
		(0.003)		(0.009)		(0.010)		(0.007)		(0.011)
PreTreatQuart Control	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
LdbPosit Control	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	No	No	No	No	No	No	No	No	No	No
Gender FE	No	No	No	No	No	No	No	No	No	No
Education FE	No	No	No	No	No	No	No	No	No	No
Observations	59,915	65,919	59,915	65,919	59,915	65,919	59,915	65,919	44,947	49,403
Adjusted R^2	0.044	0.060	0.056	0.075	0.048	0.058	0.010	0.020	0.010	0.011

Note: $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$
Standard errors clustered at user level. All continuous variables have been standardized by subtracting the value from the mean and dividing by the standard deviation.