Who are My Peers? Learner-Controlled Social Comparison in a Programming Course

Kamil Akhuseyinoglu[1], Aleksandra Klasnja Milicevic[2], and Peter Brusilovsky[1]

[1] University of Pittsburgh, PA, USA {kaa108,peterb}@pitt.edu
[2] University of Novi Sad, Novi Sad, Serbia akm@dmi.uns.ac.rs

Abstract. Studies of technology-enhanced learning (TEL) environments indicated that learner behavior could be affected (positively or negatively) by presenting information about their peer groups, such as peer in-system performance or course grades. Researchers explained these findings by the social comparison theory, competition, or by categorizing them as an impact of gamification features. Although the choice of individual peers is explored considerably in recent TEL research, the effect of learner control on peer-group selection received little attention. This paper attempts to extend prior work on learner-controlled social comparison by studying a novel fine-grained peer group selection interface in a TEL environment for learning Python programming. To achieve this goal, we analyzed system usage logs and questionnaire responses collected from multiple rounds of classroom studies. By observing student actions in selecting and refining their peer comparison cohort, we understand better whom the student perceives as their peers and how this perception changes during the course. We also explored the connection between their peer group choices and their engagement with learning content. Finally, we attempted to associate student choices in peer selection with several dimensions of individual differences.

Keywords: learner control · social comparison · open learner model · computer science education · self-regulated learning · online learning

1 Introduction

Over the last ten years, social comparison approaches have become an essential component of modern online learning tools. Researchers explored social comparison in various forms, such as leaderboards [21], comparative progress visualization [2], learning analytics dashboards [25], and socially-enhanced open learner modeling interfaces [5]. These social comparison approaches demonstrated their ability to increase learners' participation and contributions [26], help learners navigate more efficiently [16], and improve completion rates in MOOCs [8]. However, the studies on social comparison also demonstrated that it could provide no effect [8] or even negative effect for some groups of learners [19,23]. For example, high-performing learners were not affected by social comparison based on class average [8], while learners exposed to perfect peer performance exhibited declined success and increased drop rate [23]. These findings suggested that mismatches in selecting peer comparison groups could neutralize or negate the positive impact of social comparison. On the other hand, social psychology research states that comparison to similar peers strengthens the positive effect of social comparison [6].

To address the need for a proper peer group selection in social comparison, recent research explored the value of learner control over social comparison features, i.e., allowing learners to choose their peer comparison group [1]. While existing research reported positive results, the explored learner control options were quite limited: Instead of comparing themselves to the whole class, learners could choose the upper or lower part of the class as their peer groups. This paper explores the value of a more advanced interface for fine-grained learner control over social comparison in a Technology-Enhanced Learning (TEL) environment for learning Python programming. This interface allows a learner to choose precisely a segment of the class as the peer comparison group. As an added value, the freedom of choice provided by this interface offers an opportunity to examine how learners identify a segment of a class as their comparison peers. Then, we investigated how these comparison preferences relate to engagement and which factors cause variance in peer-group selections, such as achievement goals and social comparison orientation.

2 Social Comparison in Python Grids

We explored learner-controlled social comparison in a *practice system* designed for Python programming called Python Grids (PG) [1]. For this study, the PG interface was augmented with fine-grained learner-controlled social comparison features. This section reviews the components of the PG: content access interface with learner-controlled social comparison features and the set of available interactive learning tools.

2.1 The Content Access Interface

In the PG, an Open Social Learner Modeling (OSLM) interface [20] (Figure 1[B-D]) provides access to a set of Python learning content. The interface helps students decide what they need to work on and how much they need to practice freely. In this context, the ability to track personal and peer progress becomes critical to encourage students to practice more and guide them to the most relevant practice content. This ability is the core component of this interface.

The columns of the OSLM grid (Figure 1B) organize the learning content into 15 topics. The rows in the grid visualize the topic-by-topic progress of the student and the comparison peer group while making it easy to compare them to one another. The first row of the grid summarizes the topic-level progress of a learner using a green color of different density. The third row displays an aggregated average progress level of students in the selected comparison peer

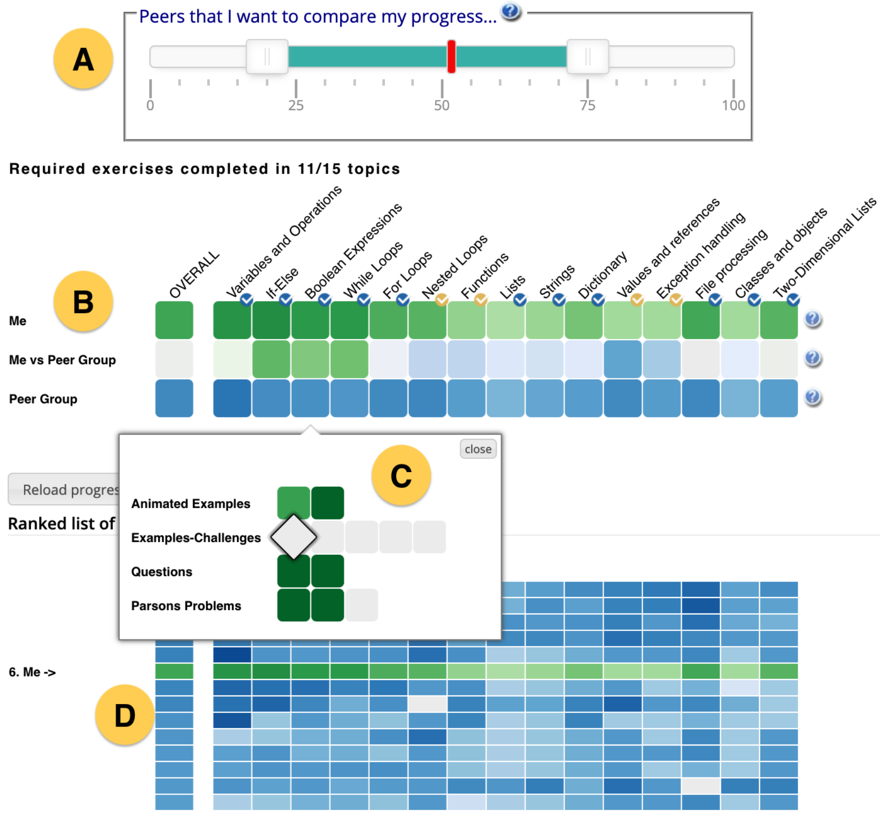

Fig. 1. The PG interface with fine-grained controllable social comparison features (A), OSLM grid (B), a set of learning activities (C), and anonymized ranked list (D).

group (Figure 1A) using a blue color. The middle *comparison* row presents the progress difference between the learner and the currently selected peer group. The green-colored topics in the middle row represent the topics where the learner is ahead of the comparison group. In contrast, the blue-colored topics show the topics where the comparison group is ahead of the student. In all cases, the darker color indicates a higher level of progress (or progress difference) for that topic. By clicking on a specific topic column, a student accesses the learning content available for this topic. Similar to the topic-level progress visualization, the PG also visualizes content-level progress using the green color density (Figure 1C). The progress of a topic or content is computed as the ratio of completed activities associated with the topic or content.

2.2 Learner-Control over Social Comparison

In our recent study [1], we explored some options for learner control, but these options were limited, i.e., a learner could compare herself to the upper or lower half of the class in addition to viewing the average progress of the whole class and anonymized ranked list of learners in the class (Figure 1D). For the current study, we augmented the interface with fine-grained control of the peer comparison group through the comparison slider widget (Figure 1A). The 0-100 progress scale represents all students in a class ranked by their current total progress in the PG from a student with the lowest progress (marked as 0) to the student with the highest progress (marked as 100). Within this range, each student could set the target comparison group using two sliders. The handles on the comparison slider define the minimum and maximum progress range of the comparison group within the class, i.e., the group that average progress is visualized by the bottom row of the PG interface (Figure 1B) and which is shown in the anonymized ranked list in detail (Figure 1D).

In the beginning, the peer group is placed in the middle of the class with the sliders set to the 25-75 range. At any time, the student can change the peer group by moving the handles or dragging the *cyan* colored segment between the handles (i.e., comparison group bar). After each change, the progress visualization in the PG interface and the ranked list are updated accordingly to show only students in the selected peer group. To help students in choosing the peer group, their own relative progress within the class is shown as a red cursor. Note that student progress is automatically displayed by the system and the position of the red cursor could move within the slider widget as the student standing in the class changes. In contrast, the selection of the peer group, i.e., the position of sliders, is fully controlled by students. Altogether, this interface offers students full freedom in deciding who their comparison peers are, i.e., how wide the group is, how far from the bottom of the class it starts, how close to the top of the class it ends, and how it is positioned in relation to student's own progress ranking.

2.3 Learning Activities

Once students decide to practice on a specific topic in the PG, they can "open" a topic and examine the available learning activities by clicking on the topic column. In each topic, the PG provided access to two types of *examples* and two types of *problems* for learning Python programming. Figure 1C shows the practice contents available for the topic of *Boolean Expressions*. Content items are shown as squares organized by the four content types. Example content types include *Animated examples* and *Examples-Challenges*. Animated examples [24] provide interactive visualization of the code execution. Examples-Challenges [15] consist of a single worked example that allows students to examine a solution to a coding problem and one or more "challenge" activities that ask students to find the missing code lines from a set of options. *Questions* and *Parson's prob-*

Table 1. Summary of practice with the learning content (N=122).

	Mean (%)	SD	Med
Number of sessions	8.66 (-)	6.26	7
Unique content accesses	99.5 (41%)	65.5	89.5
Unique questions and Parsons attempted	41.1 (51%)	24.0	41.5
Unique challenges, animated examples worked examples attempted/viewed	58.4 (36%)	47.7	53

lems are the problem types. Questions [17] are parameterized exercises that test student comprehension of program execution by asking to predict the output of a given program. Finally, Parson's problems [22] are code construction exercises in which students must arrange code lines in the correct order. In this study, students accessed 243 unique content: 81 problems (47 questions and 34 Parsons problems) and 162 examples (39 animated examples, 52 worked examples, and 71 challenges).

3 Research Methodology

3.1 Study Context

We conducted the study with 174 undergraduate students during multiple offerings of an introductory programming course at a large Australian university. The course was delivered online during the study due to the Covid-19 pandemic. The course does not assume any previous programming experience and covers programming fundamentals, including input and output, decision structures, loops, functions, data structures, file I/O, exceptions, and object-oriented programming concepts. One coordinator and two other instructors taught the course using the same syllabus, course materials, and grading policy. The passing grade is 50%, which students must collect through assignments (30%), a project (40%), and class participation (30%). By solving one Question and one Parson's problem for each of the 15 topics in the PG, i.e., 30 problems (37% of the problems in the system), students could receive up to 10% practice credit as a part of the class participation. The practice with the example content types was not counted for the credit. The blue checkmarks on each topic column in Figure 1B highlight the topics where the student fulfilled the credit requirement.

3.2 Data Collection

We collected data from four course offerings where we kept the PG the same. There were no significant differences between course offerings in learners' practice behavior in the PPG, including overall engagement and usage of the social comparison control features (e.g., the number of problem-solving attempts and peer group changes). Thus, we combined data from these offerings into a single dataset that includes system usage logs, performance measures, and individual learner differences collected through several standard instruments.

System Usage Logs: The system logs include detailed time-stamped records of practice with all learning activities including attempts to Parson's problems, questions, and challenges, viewing animated and worked examples (see Table 1). The logs also contain social comparison actions such as peer group changes and ranking list views (see Table 2). The system continuously recorded the current state of social comparison preferences, such as the orientation of the comparison group bar and the learner's current rank in the class (i.e., red cursor position).

Performance measures: In the first week of the class, we administered a pretest and several instruments focused on individual differences. The pretest had ten problems related to various Python programming concepts. Due to minimal participation in the post-test, we only considered course grades as the final performance measure.

Instruments: The social comparison orientation (only the ability factor) was measured by the Iowa-Netherlands Comparison Orientation Measure (INCOM) [14], and the achievement goal orientation framework [12] was applied to measure achievement orientations. Researchers demonstrated that both questionnaires are inter-connected in interpreting students' social comparison choices [4]. In this study, we administered these questionnaires mainly to explore their possible link to the comparison preferences observed in the PG.

In analysis, we used the logs from students who attempted at least one learning activity in the practice system. We only used students who gave their consent for the research study and received a final course grade (i.e., did not withdraw from the course). In total, we analyzed the logs of 122 students.

For questionnaire-based analysis, we filtered out students who selected the same option in all items and responded very quickly (in less than 4 minutes – 1st quartile is used as the threshold). After the initial filtering process, we analyzed the internal consistency of each scale and included the items with a factor loading of at least 0.5 on the appropriate subscale. For the achievement goal orientation, we found three valid constructs: (1) mastery approach (Cronbach's $\alpha = .61$), (2) mastery avoidance ($\alpha = .77$), and (3) performance orientation ($\alpha = .78$) (both performance avoidance and approach items loaded on the same factor). Further, we validated the social comparison orientation (ability factor) items ($\alpha = .62$). As a final step, we calculated a scale score by calculating the mean scores of the selected items related to a subscale and used these scores in our analysis. Not all students participated in the pretest and questionnaire. As a result, we only used students with the complete data for specific analyses[3].

3.3 Data Analysis Methods

In regression analysis, we checked regression assumptions, including multicollinearity, by calculating the *Variance Inflation Factors (VIF)* and ensuring none of the features had $\sqrt{VIF} > 2$. Then, we performed a backward step-wise feature selection process. We reported regression model results with the features selected by this process. For linear mixed-effects models, we added learner identifier as a random effect which also resolves the non-independence issue of our session-based data. We shared the results of mixed-effects models after confirming that the model fitted better than a random-effect only model using the likelihood ratio test. For count data predictions (e.g., number of learning activities), we used Poisson regression.

3.4 Labeling the Social Comparison Preferences

Researchers have explored the direction of social comparison, i.e., upward and downward comparison (comparing with someone better or worse), to understand

[3] We had complete data for 53 students (43%), including system logs, course grades, pretest, achievement orientation, and social comparison orientation scores.

Table 2. Summary of social comparison actions and preferences (N=113).

	M(SD)	Med	Absolute Upward	Absolute Downward	Relative Upward	Relative Downward
Peer group changes	5.7(6.4)	3.0	40%	35%	47%	52%
Ranked list views	4.3(6.6)	2.0	-	-	-	-

the potential effects of social comparison [3,7,10]. Following the prior work, we labeled learners' comparison group changes with a comparison direction to examine their comparison intentions in our analyses.

First, we performed the labeling by checking the absolute position of the selected comparison group on the 0-100 scale (the cyan segment between sliders in Figure 1A). For the *absolute labeling*, we used the index position of 50 as the fixed reference point, and we labeled the comparison group obtained after each change of sliders by four comparison types: (1) Downward, (2) Upward, (3) Balanced, or (4) Average. Downward/Upward type means that the selected comparison group mainly (or entirely) contains students from the lower-half/higher-half of the class (students below/above the reference point value of 50). The balanced comparison corresponds to the case where the comparison group covers the lower and higher half of the class equally (e.g., the sliders set to the 30-70 range). Lastly, the average type covers the case where the student selected the whole class as the comparison group (i.e., the sliders set to the 0-100 range).

Second, we used the relative position of the comparison group to students' current rank in the class (shown as a red cursor in Figure 1A) to represent the comparison direction more reasonably. We summarized learners' comparison group selection with a single scalar value for *relative labeling*. This value corresponds to the distance of the learner's current position (i.e., the red cursor) to the mid-point of the selected comparison group (i.e., the cyan segment), and we called this value mid-distance. If this value is below 0, the student's position was lower than the most (or all) of the students in the selected comparison group, i.e., performing a relatively upward social comparison. If it is above 0, the student's position was higher than the most (or all) of the comparison group, indicating a relatively downward social comparison. By using the mid-distance value, we classified each group change as (1) Downward, (2) Upward, (3), or (3) Balanced. This case has no average type since we considered the learner's current position.

4 Results

The focus of our analyses is twofold. First, we want to examine learners' interactions with the social comparison control interface and understand the social comparison preferences they expressed through this interface. Second, we want to examine the association between these preferences and engagement with the practice system. To assure that engagement with the practice system is valuable for learning, we start our analyses by examining the connection between engagement and course performance.

4.1 Engagement with the Python Grids and Course Performance

As shown in Table 1, students extensively used all content types. Notably, they solved significantly more *problems* (Parsons and questions) than the criteria for obtaining the full practice credit (i.e., solving 30 problems) ($t(121) = 5.11, p < .001$), and 71% of them (N=87) exceeded this threshold. In addition, students practiced with 36% ($M = 58.4$) of the *example* content types, although they were not counted for credit. This data indicated that the students considered the Python Grids (PG) valuable for their learning rather than just a source of credit points.

To assess the relationship between the practice system usage and course performance, we regressed course grades on pretest scores, achievement goal subscale scores, and overall practice amount (i.e., percentage of uniquely accessed learning content). We found a statistically significant regression model ($F(5, 52 = 7.2)$, $adj.R^2 = .35$, $p < .001$) with pretest scores ($B = 5.3, p = .003$), system usage ($B = 15.9, p = .015$), mastery approach ($B = 5.6, p = .004$), and mastery avoidance ($B = 7.5, p < .001$) scores were positively associated with the grades. However, performance orientation was associated with lower course grades ($B = -5.0, p = .011$). Given these results, we observed that working with the practice system was positively associated with higher course grades while keeping prior knowledge and various individual differences constant.

4.2 Social Comparison Preferences

Students used social comparison controls noticeably on average, although the usage differed between students (see Table 2). Most students (83%) used the opportunity to change their comparison peer group at least once ($M = 5.7$). Similarly, 71% of the students viewed the anonymous ranked list at least once ($M = 4.3$). Also, there was a significant correlation between the number of ranking views and comparison group changes (r=.27, p=.002). Thus, we counted both actions as *social comparison events* in the rest of the analyses.

Following the comparison preference labeling process explained in Section 3.4, we could summarize learners' preferences in peer comparison group selection in detail (see Table 2). Out of 639 comparison group changes, 41% of changes were labeled upward, 35% downward, 12% average, and 12% balanced based on the *absolute labeling*. From the *relative labeling* prospect, students preferred downward comparison the most (52%), then upward comparison (47%). Only 1% of the changes were balanced. Thus, according to the absolute labeling, students preferred upward comparison the most. However, the dominance of upward comparison was not present in the relative labeling. This difference might originate from the fact that for high-performing students (e.g., a student at the 5th percentile), there is limited opportunity to perform an upward comparison due to the ceiling effect.

4.3 Social Comparison Events and Engagement

Throughout the semester, learners worked with the practice system in multiple sessions of varying duration and with different intensities. We hypothesized that if social comparison events (i.e., group change and ranking view) influence engagement, we should observe this effect on the total number of learning actions performed in a session (*num-act*), i.e., problem-solving attempts and example views. Thus, we classified all sessions (N=1057) into two types: those with at least one social comparison event occurred (27%) and those without (73%). Then, we compared the number of learning actions performed in these session types per student. We filtered out students who did not have both types of sessions for this analysis (N=93). We discovered that students practiced significantly more in sessions when they also performed a social comparison event ($M = 72, Med = 43$) compared to sessions without a comparison event ($M = 40, Med = 27$)($V = 3028, p < .001$). This observation holds for both the example and problem activity types. Moreover, students had a significantly higher chance to increase their in-system progress-based ranking as a result of their practice (19% progress difference) in sessions when they interacted with the social comparison controls ($t(92) = 5.54, p < .001$).

4.4 The Effect of Social Comparison Direction on Engagement

The results reported above revealed a positive association between the usage of social comparison controls and practice. However, this connection might depend on social comparison direction, namely upward or downward. This section assesses the effect of direction on learner engagement.

First, we analyzed the direction effect based on the *absolute* labeling. To perform such an evaluation, we considered learning sessions containing at least one comparison group change (N=146). This filtering was necessary to concentrate on sessions with explicit group change. We utilized the labeling process described in Section 3.4 and calculated the ratio of *upward* social comparison changes (*upward-ratio*) within a session. Then, we predicted the number of learning actions (*num-act*) performed in a session by fitting a linear mixed-effects model with the *upward-ratio* and session duration as fixed effects. We found significant positive effects of the *upward-ratio* ($B = .21, z = 8.5, p < .001$) and the session duration ($B = .95, z = 47.2, p < .001$) on *num-act*. We also found an opposite effect for the *downward* social comparison. These findings highlight the importance of comparison direction, namely *upward* social comparison, on engagement.

Second, we leveraged the relative labeling to examine the comparison direction. We used the *mid-distance* value (as described in Section 3.4) and calculated the mean of *mid-distance* for each learner session to represent the comparison direction. Using this mean value, we categorized a session either as an *upward* or *downward* comparison session, e.g., a session was labeled as upward when the mean *mid-distance* was below zero. Additionally, we categorized each session as a *lower* or *higher* standing session by computing the mean of learner position

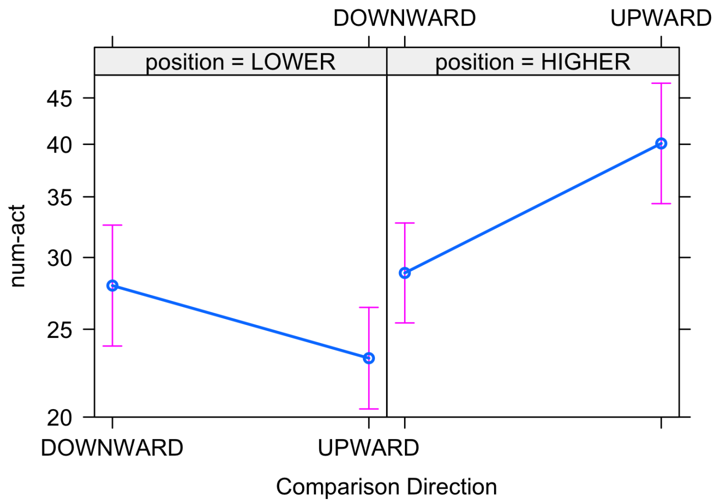

Fig. 2. Predicted number of unique learning activities (*num-act*) for the interaction term (direction*position). Purple bars denote 95% confidence interval.

index (on the 0-100 scale). For example, a higher standing session implies the learner is positioned in the higher half of the class (above 50) on average during that session. In this case, we considered sessions containing a comparison group change or the ones that come after the first comparison group change, not necessarily including another comparison change (N=765 sessions). This filtering was critical in observing the learner's explicit attitude in peer-group selection throughout multiple sessions, given that students could observe their positions without changing their comparison groups. We fitted a linear mixed model with the comparison *direction*, learner *position*, and session duration as fixed effects to predict *num-act* in a session. A significant interaction effect of position and direction was found ($B = .51, z = 7.77, p < .001$), along with a significant effect of session duration on *num-act*. As presented in the interaction effect plot (Figure 2), the results revealed that students performed more learning actions in lower standing sessions if they were engaged in downward comparison (num-act= 28) compared to upward comparison (num-act= 23). In contrast, if they were in the higher progress state, engaging with upward comparison (num-act= 40) was more effective than downward comparison (num-act= 29). To summarize, this detailed analysis revealed that engagement with the learning activities was associated with the direction of the social comparison and the progress standing of the student.

4.5 How to Explain Learners' Social Comparison Preferences?

We explored the social comparison preferences of learners in Section 4.2 to understand the frequency and type of comparison group changes, such as upward or downward comparison. However, in that section, we did not discuss the factors that might affect learners' choice in selecting their peer comparison group.

We started by checking which factors affect the size of the selected comparison group (i.e., having a more expansive comparison group bar in Figure 1A). A fitted linear mixed model revealed that the higher the learner's current position within the class, the wider the comparison bar is ($B = 3.26, t = 2.604, p = .010$). In addition, being closer to the end of the course was positively associated with choosing a larger comparison group ($B = 2.54, t = 1.992, p = .048$).

How did students increase the size of the comparison group? To modify the size and placement of a peer group, students could adjust either the left or right slider, and their use might be associated with different factors. To understand these factors, we fitted two separate mixed-effects models to predict the position of the left and right slider after controlling for the position of the opposite slider. Regression results indicated that the current standing of the learner in class was statistically significantly and positively associated only with the position of the right slider ($B = 2.91, t = 2.993, p = .003$). On the other hand, closeness to the end of the course was marginally and negatively correlated with only the left slider position ($B = -1.93, t = -1.824, p = .069$). As a result, we concluded that when students advanced in their standing within the class, they increased their comparison group size by adding stronger students (i.e., by moving the upper slider to the right). In addition, while approaching the end of the class, students added weaker students to their peer group by decreasing the position of the lower slider (i.e., moving it to the left).

We extended our analysis by connecting the comparison preferences with the collected self-reported instruments. Thus, we tried to predict the scalar value of *mid-distance* by using the collected instruments (see Section 3.2 for details). We fitted a linear mixed model on the session-based data (290 comparison group changes). The results indicated that there was a significant effect of *social comparison orientation score* ($B = -17.76, t = -4.774, p =< .001$) and *performance orientation score* ($B = 9.66, t = 2.757, p = .008$) on *mid-distance*. In other words, socially-oriented students preferred upward social comparison (given the sign of the regression coefficient) while performance oriented students favored downward social comparison. Following the previous analysis, we fitted another linear regression model to predict the size of the comparison group but could not find any significant model.

5 Discussion: Results in the Context of Related Work

In this paper, we report the results of several rounds of classroom studies to explore the effects of learner-controlled social comparison on learner engagement and performance in an online programming practice system. We observed that students used the system extensively throughout the semester and showed that their engagement with the system was positively correlated with the course grades. We also found a link between achievement goals and course performance, where mastery-oriented students finished the course with better grades [11].

The unique design of the user-controlled social comparison interface also enabled us to explore the diverse learner preferences towards social comparison. Social comparison theory states that people want continuous improvement and assess their capabilities and opinions by comparing themselves to similar people [13]. Moreover, the performance-based reward system in education leads students to compare themselves socially [9]. Our analyses show that students paid considerable attention to social comparison features. We also observed a gradual change in their social preferences, which is consistent with the findings of Huguet et al. [18], who argued that social comparison is a dynamic process that changes over time. Our data also demonstrated that students tend to choose the upward social comparison (in absolute labeling) most frequently in a TEL environment, the tendency observed earlier in other contexts [10].

A deeper analysis of social comparison choices yielded more discoveries, which correlate with findings reported in the literature. First, we observed that students practiced significantly more and increased their in-system progress levels in sessions where they also self-assess their current state by interacting with the learner-controlled social comparison features. Researchers presented similar positive effects of social comparison [26,8]. We also highlighted that the direction of the comparison and the progress level of a learner impact the benefit of social comparison. We found that engaging with upward social comparison (in absolute labeling) was positively associated with enhanced practice intensity. Researchers argued that learners tend to perform upward comparison as a means of self-improvement when they also recognize that they can improve their standing [7,18]. Moreover, the progress state of a learner interacted with the comparison direction (in relative labeling) such that performing a comparison that is "matched" to their current state (i.e., performing upward comparison while being in the higher state) was more beneficial on engagement. This interaction could mean that the upward comparison might be beneficial only when students do not feel uneasy about being inferior [3]. We believe that the novel learner-controlled comparison features with OSLM features helped learners choose appropriate peer groups based on their standing, leading to increased engagement.

We concluded our analysis by exploring the factors affecting the comparison preferences. For example, we observed that students added high-performing students into their peer groups based on their standing within the class. Finally, we connected peer group preferences back to learners' differences and discovered that students with higher social comparison orientation favored upward social comparison, while performance-oriented students preferred downward comparison. This finding conforms to earlier observations where researchers found that the performance-avoidance group conducts downward comparison more [4].

6 Prospects and Limitations

Our work demonstrated that fine-grained learner controls on social comparison could increase the effect of social comparison by helping learners find the most appropriate peers. Moreover, we showed that these control features provide valuable insight into students' intentions in the peer-group selection and emerge as a practical technology for future studies. We want to explore learner control more broadly while addressing several limitations of this study in future work. We hope to augment our findings with qualitative analysis to understand how students think and feel while adjusting their comparison groups. Moreover, the online delivery of the programming course could impact students' comparison behavior. Even though we diligently verified our statistical findings, we conducted some of the analysis only with limited data. Also, the authors are conscious of the difference between causality and correlations, and more rigorous study designs are needed to investigate causal effects. Finally, although the system usage was encouraged slightly through course credits, our study might be susceptible to the self-selection bias since the majority of the system use was voluntary. We hope to address these limitations in our future work.

Acknowledgements This material is partially supported by the National Science Foundation under Grant No. 1822752.

References

1. Akhuseyinoglu, K., Barria-Pineda, J., Sosnovsky, S., Lamprecht, A.L., Guerra, J., Brusilovsky, P.: Exploring student-controlled social comparison. In: European Conference on Technology Enhanced Learning EC-TEL 2020. pp. 244–258. Springer International Publishing (2020)
2. Auvinen, T., Hakulinen, L., Malmi, L.: Increasing students' awareness of their behavior in online learning environments with visualizations and achievement badges. IEEE Transactions on Learning Technologies **8**(3), 261–273 (2018)
3. Blanton, H., Gibbons, F.X., Buunk, B.P., Kuyper, H.: When better-than-others compare upward: Choice of comparison and comparative evaluation as independent predictors of academic performance. Journal of Personality and Social Psychology **76**(3), 420–430 (1999)
4. Bounoua, L., Cury, F., Regner, I., Huguet, P., Barron, K.E., Elliot, A.J.: Motivated use of information about others: Linking the 2×2 achievement goal model to social comparison propensities and processes. British Journal of Social Psychology **51**(4), 626–641 (2012)
5. Brusilovsky, P., Somyürek, S., Guerra, J., Hosseini, R., Zadorozhny, V., Durlach, P.J.: Open social student modeling for personalized learning. IEEE Transactions on Emerging Topics in Computing **4**(3), 450–461 (2016)
6. Cialdini, R.B., Wosinska, W., Barrett, D.W., Butner, J., Gornik-Durose, M.: Compliance with a request in two cultures: The differential influence of social proof and commitment/consistency on collectivists and individualists. Personality and Social Psychology Bulletin **25**(10), 1242–1253 (1999)
7. Collins, R.L.: For better or worse: The impact of upward social comparison on self-evaluations. Psychological Bulletin **119**(1), 51–69 (1996)
8. Davis, D., Jivet, I., Kizilcec, R.F., Chen, G., Hauff, C., Houben, G.J.: Follow the successful crowd: raising mooc completion rates through social comparison at scale. In: 7th International Learning Analytics & Knowledge Conference. pp. 454–463 (2017)

9. Dijkstra, P., Kuyper, H., van der Werf, G., Buunk, A.P., van der Zee, Y.G.: Social comparison in the classroom: A review. Review of Educational Research **78**(4), 828–879 (2008)
10. Dumas, F., Huguet, P., Monteil, J.M., Rastoul, C., Nezlek, J.B.: Social comparison in the classroom: Is there a tendency to compare upward in elementary school? Current Research in Social Psychology **10**(12), 166–186 (2005)
11. Elliot, A.J., Mcgregor, H., Gable, S.L.: Achievement goals, study strategies, and exam performance: A mediational analysis. Journal of Educational Psychology **91**, 549–563 (1999)
12. Elliot, A.J., Murayama, K.: On the measurement of achievement goals: Critique, illustration, and application. Journal of educational psychology **100**(3), 613 (2008)
13. Festinger, L.: A theory of social comparison processes. Human Relations **7**(2), 117–140 (1954)
14. Gibbons, F.X., Buunk, B.P.: Individual differences in social comparison: development of a scale of social comparison orientation. Journal of Personality and Social Psychology **76**(1), 129 (1999)
15. Hosseini, R., Akhuseyinoglu, K., Brusilovsky, P., Malmi, L., Pollari-Malmi, K., Schunn, C., Sirkiä, T.: Improving engagement in program construction examples for learning python programming. International Journal of Artificial Intelligence in Education **30**(2), 299–336 (2020)
16. Hsiao, I.H., Brusilovsky, P.: Guiding and motivating students through open social student modeling: Lessons learned. Teachers College Record **119**(3) (2017)
17. Hsiao, I.H., Brusilovsky, P., Sosnovsky, S.: Web-based parameterized questions for object-oriented programming. In: World Conference on E-Learning, E-Learn 2008. pp. 3728–3735 (2008)
18. Huguet, P., Dumas, F., Monteil, J.M., Genestoux, N.: Social comparison choices in the classroom: Further evidence for students' upward comparison tendency and its beneficial impact on performance. European Journal of Social Psychology **31**(5), 557–578 (2001)
19. Kitsantas, A., Zimmerman, B.J., Cleary, T.: The role of observation and emulation in the development of athletic self-regulation. Journal of Educational Psychology **92**(4), 811–817 (2000)
20. Loboda, T.D., Guerra, J., Hosseini, R., Brusilovsky, P.: Mastery Grids: An open source social educational progress visualization. In: European Conference on Technology Enhanced Learning, EC-TEL 2014. pp. 235–248. Springer (2014)
21. Morales-Trujillo, M.E., García-Mireles, G.A.: Gamification and SQL: An empirical study on student performance in a database course. ACM Transactions on Computing Education **21**(1) (2021)
22. Parsons, D., Haden, P.: Parson's programming puzzles: A fun and effective learning tool for first programming courses. In: Proceedings of the 8th Australasian Conference on Computing Education-Volume 52. pp. 157–163 (2006)
23. Rogers, T., Feller, A.: Discouraged by peer excellence: Exposure to exemplary peer performance causes quitting. Psychological Science **27**(3), 365–374 (2016)
24. Sirkiä, T.: Jsvee & kelmu: Creating and tailoring program animations for computing education. Journal of Software: Evolution and Process **30**(2) (2018)
25. Teasley, S.D.: Student Facing Dashboards: One Size Fits All? Technology, Knowledge and Learning **22**(3), 377–384 (2017)
26. Vassileva, J.: Toward social learning environments. IEEE Transaction on Learning Technologies **1**(4), 199–214 (2008)