Exploring Student-Controlled Social Comparison

Kamil Akhuseyinoglu[1]([✉]) [iD], Jordan Barria-Pineda[1] [iD], Sergey Sosnovsky[2] [iD],
Anna-Lena Lamprecht[2] [iD], Julio Guerra[3] [iD], and Peter Brusilovsky[1] [iD]

[1] University of Pittsburgh, Pittsburgh, PA, USA
{kaa108,jab464,peterb}@pitt.edu
[2] Utrecht University, Utrecht, The Netherlands
{s.a.sosnovsky,a.l.lamprecht}@uu.nl
[3] Universidad Austral de Chile, Valdivia, Chile
jguerra@inf.uach.cl

Abstract. Previous research on technology-enhanced learning indicated that exposing students to information related to their peers' performance might positively or negatively affect their behavior and performance. For example, recent research has demonstrated that augmenting traditional open learner models (OLMs) with views of the learner model of peers could boost student engagement and affect navigational patterns. On the other hand, the negative impact of social comparison has been also reported in the literature, which demonstrates that a comparison with much better-performing peers presents a threat to self-integrity. These conflicting findings have not yet been reconciled in the context of technology-enhanced learning. This work attempts to extend research in social comparison in an educational context and on OLMs by examining how the potential negative and positive sides of social comparison could be balanced by enabling students to select their peer comparison group, rather than by being forced to compare themselves with an aggregated class average.

Keywords: Social comparison · User control · Open learner model

1 Introduction

Social comparison theory states that people tend to assess their capabilities and opinions to continuously improve themselves through comparisons to other similar people when no objective comparison measures are available [8]. The reward system in education is based on academic performance, which generates an environment where students compare themselves socially [7]. Due to this nature, social comparison theory has been studied in both regular classroom settings and online learning environments, including MOOCs [2,6,7,13,19].

As shown in Sect. 2, the studies of the impact of social comparison in an educational context revealed both positive and negative cases. Could we design a

social comparison approach for technology-enhanced learning that could minimize the potential negative impact of social comparison while retaining its positive impact? The work presented in this paper attempts to achieve this goal by offering students the opportunity to freely choose their peer comparison group rather than being forced to compare themselves only with the class average. Following a brief review of research on social comparison in education and technology-enhanced learning, we present a design for open learning modeling with user-controllable social comparison. This design has been implemented in an online practice system for learning Python programming and evaluated in a classroom study. The data collected in this study offers interesting insights into the usage of controllable social comparison and its impact on user performance and navigation behavior.

2   Related Work

Social comparison theory explains how comparing oneself to others influences both behavior and self-judgments [8]. What we know about other people's performance, beliefs, and behavior can provide valuable information to self-assess and determine, at least in part, the actions we will pursue. In computer-based learning environments, social comparison has been used to improve navigation support [3] and to design so-called open social learner modeling (OSLM) by augmenting traditional OLMs with views of the learner models of peers [2,12,20]. Previous studies of social navigation support and OSLM have consistently found (although with different strengths) that social comparison features could boost engagement and affect user navigation patterns. In particular, the opportunity to see the models of other learners motivated students to cover more topics in the system and reach higher success rates in problems [12]. It also encouraged weaker students to follow *strong* students in their content navigation [12]. Another OSLM-related study [2] showed that a group of students who enabled social comparison reached higher rates of system usage and higher learning effectiveness than the control group, who had no access to social comparison features. Researchers also attempted a deeper exploration of the engagement effect by considering individual differences. For example, [10] shows how engagement with the system was positively correlated to changes in motivation factors, such as *Performance-Approach* in the group who were exposed to social comparison features. Recent work has also shown how the social comparison features accounted for better completion rates in MOOCs [6]. Similar ideas have been investigated from the broader perspective of learning analytics. For example, Shi and Cristea [20] incorporated visual indicators of different learning-related information, such as learning paths and learner contributions, into a multifaceted OLM with social comparison features.

Although the bulk of research mentioned above has found some positive effects of using social comparison in educational systems, multiple studies show the negative effect of social comparison and competition. Rogers and Feller [19] explored the effects of peer excellence in MOOC settings and found that when

learners are exposed to *perfect* peer performance, their success declined and the drop rate increased. These authors stated that their findings only hold when students compare themselves with excellent peers and that no motivation boost was found when students were compared with *poor performers*. Similarly, in a skill acquisition environment (i.e., dart throwing), students who observed mastered models performed lower than students who observed *coping* models [14].

These findings suggested some possible negative consequences of *upward comparison* (i.e., comparing yourself with someone much better) in educational contexts. This phenomenon might be explained by the feeling of failure and threats to self-integrity when being inferior to the compared peer [17]. However, the feeling of failure only appears when the gap between the individual and the peer is huge. Rogers and Feller [19] suggested that the negative effect of social comparison might be eliminated by exposing learners to less successful peers along with better performers. Moreover, when the comparison target is selected from *similar people*, it is expected to strengthen the positive effect of social comparison [5].

The definition of who exactly are *similar people* have generated some discussion in the learning sciences research field. The proxy model proposed by [22] stresses the need to find a suitable one to compare with, a *proxy*, who shares relevant similar characteristics with the learner. By comparing to a relevant *proxy*, the learner gains information that allows her to better predict future achievement. In a learning context, a student could provide a better self-assessment (answer the question "Can I do X?") if she knows another learner who is similar (usually in terms of prior knowledge) and for whom she observed their achievement. This model suggests that the overall validity of self-assessment could be increased if we provide an adaptive comparison, in which the learner compares themselves to one who is relevant for the comparison (self-evaluation of knowledge and skills). This approach also explains that self-enhancement through social comparison works whether the one assimilates or contrasts herself to similar, less advantaged, or superior others [21]. Following this, our present work extends the research on social comparison in educational systems and OSLMs by examining the impact of controllable social comparison, a design that enables students to select their comparison group.

3   Python Grids: An Online Practice System

This section presents the design of an online programming practice system, Python Grids (PG), with user-controllable social comparison. This system was used to assess the effect of controllable social comparison in a classroom study that is detailed in the following sections.

3.1   Interface and Social Comparison Features

Python Grids (PG) is an integrated practice system that allows learners to access several types of Python learning content through a unified interface. All learning

content in the system is provided for free practice; i.e., students can decide what they need to work on and how much they need to practice. In this context, the opportunity to track one's progress and the ability to observe the progress of peers could be useful to encourage students to practice more and to guide to the most relevant practice content. In PG, these functions are provided by the Mastery Grids interface [16] to access all practice content (Fig. 1).
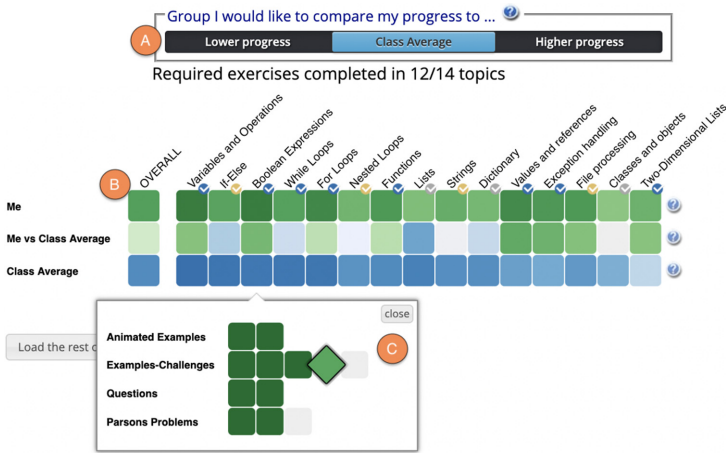

Fig. 1. Python grids (PG) interface with user-controllable social comparison features (A), OSLM grid (B), and content collection for the topic of *Boolean expressions* (C). (Color figure online)

The learning content in Mastery Grids is organized into a set of topics represented by the columns of the OSLM grid (Fig. 1-B). The rows of the grid visualize the topic-by-topic learning progress of the target student and a comparison group while making it easy to compare them to one another. The first row summarizes the topic-level progress of a learner using a green color of different density. The third row shows the aggregated average progress level of students in the selected comparison group (e.g., *Class average*, *Higher progress* or *Lower progress*) using a blue color of different density. In both cases, the darker color indicates a higher level of progress within a topic. The middle *comparison* row presents the progress difference between the learner and the currently selected comparison 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. The density of the color (green or blue) shows the magnitude of the progress difference. By clicking on a specific topic column, a student accesses the learning content available on this topic (see Sect. 3.2 for available learning content). Similar to the topic-level progress visualization, PG also visualizes content-level progress using the green color density (Fig. 1-C). The progress of a topic or content is computed as the proportion of completed activities associated with the topic or content.

The comparison group is set to the class average at the first login, and students can change their comparison group by clicking one of the three options (Fig. 1-A). This selection is recorded and when students access PG again, the comparison group is set to the latest selected group. Based on the selected comparison group, the average of the group progress is updated and students can compare their progress against the newly selected comparison group. The class average group consists of all students who have logged in to PG at least once. The lower and higher progress groups are formed by splitting the students by the median of the average progress. These groups are dynamically updated while the learners are working in the system. In addition to the aggregated progress comparison, students have access to an anonymized ranked list of all individual student models where they can see their current rank in the list. When a student switches the comparison group, the ranked list is updated accordingly and shows only the students in that progress group. The interface features are explained to students in a start-up tutorial that is presented at the first login and is repeated weekly to remind them about the interface features.

3.2   Learning Content

Figure 1-C shows the available interactive practice content for the opened topic of *Boolean Expressions*. In this study, PG provided access to four types of interactive practice content for learning Python programming: *Animated examples*, *Questions*, *Parsons problems*, and *Examples-Challenges*. Animated examples provide an expression-level visualization of the code execution. Questions are parameterized exercises that test student comprehension of program execution by asking for the output of a given program. Parsons problems are code construction exercises in which students do not need to type code, but have to arrange code lines in the correct order. More details about the first three types of Python practice content available in PG can be found in a paper presenting an older version of PG without student-controlled social comparison [4]. An important content-level difference with the older version is the replacement of simple annotated code examples with a new type of content called *Examples-Challenges* [11]. Each example-challenge is a combination of a worked-out example and several "challenge" problems that allows students to practice knowledge learned from the example. Each challenge problem is technically an example that is similar to the original one, but has one or more lines missing. Solving the challenge requires the students to find the correct missing code lines from a set of options.

4   Methodology

4.1   The Study Context

This study was conducted in the context of an introductory programming course at Utrecht University in early 2020. The course is an introduction to computational thinking that is focused on data analysis problems and the implementation

of data analysis programs with Python. It is explicitly intended for students who have zero or little programming experience and who do not have a mandatory programming course in their chosen course of study.

The first half of the course focuses on the basics of programming and of the Python programming language. In a series of eight modules, students learn about outputs and inputs at the command line, string formatting, variables, Boolean expressions, conditional branching, loops, functions, modules, packages, data structures, file I/O, and error handling. The second half of the course then moves on to more advanced, practical topics of programming in Python, such as common data science libraries (e.g., *pandas*, *matplotlib*).

We offered the PG to the students as an optional practice for the basic Python concepts taught in the first half of the course (except for object-orientation). The topics in the PG matched with what the students were expected to master for the midterm exam. In particular, *Questions* content in PG were very similar to the code reading questions in the midterm. To further increase students' motivation to use PG, they were able to collect bonus points for completing the homework requirement. To be admitted to the exam, they needed to submit 80% of the weekly lab/homework exercises. By solving one Questions and one Parsons problem for each of the 14 topics in PG, they could receive up to 2% for this requirement. They could also check their requirement status in PG with the little check-marks on topic columns and a summary text (Fig. 1).

4.2   Data Collection

The PG usage data includes logs of sessions (separate log-ins into the system), problem-solving attempts for Parsons problems, Questions, and Examples-Challenges, viewing animated examples on the level of individual animation steps, and viewing worked-out examples and textual explanations on the level of individual code lines. The logs also contain social comparison group changes and ranked list views. All logs were stored with timestamp information, and each student was given an anonymized login credential by their instructor to access PG. The overall usage statistics are shared in Table 1. In our analysis, we used the logs from students who attempted at least one learning activity. Thus, in our analysis, we only considered logs from 44 students. Given our analysis results in Sect. 5.2, we decided to use the system logs from the whole semester, although the course was taught remotely after the midterm exam due to the Covid-19 pandemic.

5   The Use of Social Comparison

5.1   Overall Social Comparison Usage

We started our analysis by checking the overall usage of social comparison features in PG. As the data shows, the opportunity to change the basis of social comparison has been used quite extensively. Out of 44 students who performed

at least one learning activity, 40 (91%) changed their social comparison group at least once, and 15 of them (34%) performed more than 10 group changes. Also, 26 (65%) of them viewed the anonymous ranked list of individual models at least once. As shown in Table 1, on average, students made 17.65 comparison group changes and viewed the ranked list 3.48 times. Out of 706 total comparison group changes, 32% of them performed to the higher progress group (i.e. the comparison group was changed to the higher group), 21% of them performed to the lower group, and 47% of them performed to the average group. We observed very different preferences among students. For example, most of the students generally stayed in the average comparison group, while 3 students preferred the lower progress group and 7 students preferred the higher progress group as their dominant comparison group. Finally, we observed a significant correlation between the number of ranking views and total comparison group changes $(r = .70, p < .001)$.

Table 1. Overall usage statistics and social comparison events.

Overall (N = 44)	Median	Mean	SD
Total activity access	99.5	112.73	77.98
Number of sessions	9.0	12.41	11.50
Unique questions attempted	27.0	24.07	14.10
Unique parsons attempted	19.5	17.64	11.84
Unique challenges attempted	23.0	30.23	26.54
Unique animated examples viewed	11.5	16.82	13.26
Unique examples viewed	19.0	23.98	18.57
Social comparison events (N = 40)			
Total comparison group changes	13.5	17.65	19.09
Change to lower group	3.0	3.67	3.75
Change to average group	6.0	8.28	9.12
Change to higher group	4.0	5.70	6.57
Ranked list views	1.0	3.48	7.74

5.2   Time Distribution of Social Comparison Activity

When we checked the distribution of comparison events over the course duration, we saw a clear change in the number of comparison group changes after the midterm exam date. The majority of the comparison changes (92%) happened in the first half of the course. This behavior change was likely caused by the decreased usage of the PG in the second half of the course: 79% of all sessions occurred in the first half. This was natural given that PG content mostly

covers topics from the first part of the course before the midterm exam. To support this explanation, we observed that the ratio of social comparison events (group changes and ranked list views) to the total number of all actions in PG stayed reasonably stable throughout the semester (between 3% and 8%), with the lowest ratio reached at the midterm week due to the increase in practice events. Furthermore, the preference for each comparison group option had also been stable throughout the semester (Fig. 2). Each week, almost half of the group changes made to the average group. This is followed by changes to the higher progress group (roughly 30%). Finally, the lower progress group preferred the least (roughly 20%). Moreover, throughout the semester, the proportion of students who performed at least one comparison change per session fluctuated between 40% to 80%. We observed the highest proportion right before the midterm exam and the lowest proportion right after the midterm exam. On average, each week, half of the students (53%) performed a social comparison event. This shows that a considerable number of students used the social comparison features of PG, even after the midterm exam. This leads us to consider the whole semester data for a detailed analysis, which is presented in the following sections.

5.3   Stability of Comparison Group Selection

The *stability* of a comparison group selection examines how long students stay with a specific selection of a social comparison group. The longer a student works with a specific group selected for comparison, the more this selection might affect her behavior and performance. If a student performs a comparison change immediately following another group change (i.e. before performing any learning activity), we labeled this change as *exploratory*. This labeling might show us the intention of the student when changing the comparison group. For example, students might change the comparison group just to *explore* to see their current progress state relative to different groups, or to *stick* to the comparison group
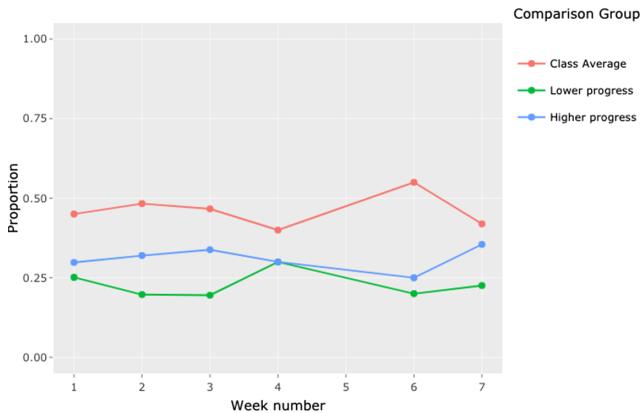

Fig. 2. Proportions of comparison group changes per week during the semester.

and set a possible target to reach. The data revealed that 67% of comparison changes were *exploratory*.

To further understand the stability of the comparison group selection, we calculated the proportion of practice time while a comparison group is selected. Repeated-measures ANOVA analysis revealed a statistically significant effect of the comparison group on practice time proportion ($F(1.61, 64.3) = 27.34, p < .001, \eta_p^2 = 0.41$) after a Greenhouse-Geisser correction was used. Bonferroni-adjusted pairwise comparisons showed that there was a significant practice time difference between the average group ($M = 0.56$) and the higher group ($M = 0.20$), $t(40) = 4.58, p < .001$ and between the average group and the lower group ($M = 0.07$), $t(40) = 7.27, p < .001$. However, there was no significant difference between higher and lower comparison groups, $t(40) = -1.91, p = .064$, even if the average practice time per student with the higher group selection ($M = 75.4$ min) was more than 3 times larger than that of the lower group selection($M = 22.13$ min). The practice time before the first comparison switch (17%) is not considered in this analysis. In summary, students mainly practiced while the average group was set as the comparison group.

We also considered the final comparison group selection of the students at the end of the term as another stability measure. The average group was picked by the majority of students (63%, n = 25) as their final comparison group, followed by the higher group (25%, n = 10) and then the lower group (12%, n = 5).

The stability results highlighted that the average comparison group was the most preferred comparison group during the semester. As stated earlier, the average group contains all students in the class and is a combination of both the lower and the higher progress group. We believe that this is one of the main reasons why students engaged with many comparison group changes (i.e., *exploratory* switches), but stuck to the average group most of the time when using PG.

6   The Impact of Social Comparison

6.1   Social Comparison and System Usage

In this section, we have used the session-based log data. A session starts when a student logs in to PG and finishes when there is no activity for a period of 30 min. For each session, we have calculated the number of learning activities (e.g. example viewing, line clicks, problem-solving attempts), and the number of comparison events performed (e.g. comparison group change, ranked list view). We also labeled each session with the initial progress state and the progress group that a student belongs to at the start of a session (e.g. lower or higher progress group).

To understand the relationship between social comparison events and other usage metrics, in some of the analysis below, we conducted linear mixed-effect analysis using the *lme4* and *lmerTest* packages in *R* [1,15]. We selected this approach since it allows us to investigate students' session-based data, where each student had a different number of sessions (i.e., allowing missing data)

[18]. More importantly, we should account for the individual differences in social comparison [9] in our analysis models. Thus, we included learner ids as a random effect in all our linear mixed models, which also resolves the non-independence issue of our session-based data. Since we predicted counts data (e.g., number of learning activities), we used Poisson regression.

Following the prior research findings, we hypothesized that sessions where students performed a social comparison event would boost overall levels of student engagement with PG. Thus, there would be a difference in the total number of learning activities performed in sessions either with or without a comparison event. To test this hypothesis, we performed a paired Wilcoxon signed-rank test. The test result revealed that students performed significantly more learning activities when there was a comparison event ($M_{comparison} = 31.06, M_{wocomparison} = 19.77$), $V = 596.5, p = .01$. These results indicate that we found a significant correlation between engagement and social comparison events.

We further checked if the number of learning activities performed in PG was associated with session time and the number of social comparison events (SCE). To achieve this, we divided the sessions into two main session periods: early and later sessions. We fitted a linear mixed model to predict the number of learning activities within a session, where session period in the student timeline and SCE were added as the fixed effects. The results indicated that there was no significant effect of session period on the number of learning activities performed in a session ($B = -0.009, z = -1.09, p = .27$). However, we found a significant effect of SCE on the number of learning activities ($B = 0.051, z = 19.04, p < .001$). Thus, in sessions where students engaged with more SCE, students also engaged with more learning activities, regardless of the session period.

Notwithstanding the above, performing more learning actions does not always mean *better* practice. Students who performed more SCE might have engaged with easier content just to increase their progress within PG if they were concerned with their progress position in the class. To investigate this behavior, we divided learning activities into easy/hard content based on the levels of complexity. We labeled animated examples and worked-out examples as less complex activities (i.e., easy content) and Parsons problems and Questions as more complex activities (i.e., hard content). Then, we calculated a complexity ratio per session by dividing the number of hard content activity by the total learning activities. We performed a linear mixed model analysis to predict the complexity ratio with SCE and session period as fixed effects. These results indicated that students worked significantly less on complex learning activities in their earlier sessions ($M = 0.43$) compared to later sessions ($M = 0.57$) ($B = -0.069, t = -3.615, p < .001$). However, there was no significant effect of SCE on the complexity ratio ($B = -0.005, t = -0.755, p = .45$). Given these findings, we could not conclude that social comparison directed students to easier or harder content, but rather that it increased the engagement with both easier and harder learning activities.

Following our stability analysis in Sect. 5.3, we checked if *exploratory* social comparison behavior had any effect on the number of learning activities. For this aspect, we calculated the *exploratory* ratio by dividing the number of *exploratory* comparison changes to the total number of comparison group changes. We fitted a mixed model to predict the number of learning activities in a session with *exploratory* ratio as the fixed effect. We found that students who performed more exploration with social comparison features in a session performed significantly less learning activity, as compared to a session with fewer exploration of these features ($B = -.48, z = -12.48, p < .001$). Exploratory switches might let students see a glimpse of their progress state compared to other groups while limiting the exposure to the comparison result to a longer period of time; excessive switching back and forth might reduce their concentration on practicing with the learning content.

6.2   Does Direction of Social Comparison Matter?

In previous sections, we explored the relationship between social comparison and system usage without considering the direction of the social comparison. In this section, we investigated the impact of the progress state (i.e., belonging to the higher/lower progress group) and the direction of the comparison group selection on learning activities. To specify the direction of the social comparison, we checked students' initial progress states at the beginning of a session and their comparison group selections. For example, if a student started the session in the higher progress group and then selected the *higher* progress group as the comparison group, this means that the student performed a *matched* social comparison event. This means their progress state was matched with the selected comparison group. If the student performed at least one matched comparison in a session, we labeled the session as a comparison-matched session and used as a dummy variable (comparison-matched) in the following linear mixed model analysis.

To check if performing matched social comparison group changes has any relationship to the number of learning activities performed in a session, we fitted a linear mixed model to predict the number of learning activities (content). We used the initial progress state (lower/higher), the comparison-matched dummy variable (true/false), and the interaction term as the fixed effects. We added the interaction term into the model to check if the effect of comparing with similar peers (i.e., performing a matched comparison) was similar at different progress states. We found a significant interaction effect ($B = -.709, z = -16.88, p < .001$) along with a significant effect on the initial progress state ($B = .516, z = 16.62, p < .001$) and a significant effect of comparison-matched sessions ($B = .450, z = 19.92, p < .001$) on the number of learning activities. As shown in the interactions effect plot (Fig. 3), results revealed that students who started the session in the higher progress state performed a significantly higher number of learning activities if they also performed a matched social comparison event in that session (i.e. selecting the higher progress group as the comparison

group) ($M = 25.78$), as compared to sessions without any matched social comparison events ($M = 16.43$). In contrast, students who started the session in a lower progress state performed significantly fewer learning activities if the session had a matched social comparison event (i.e., selecting the lower progress group as the comparison group) ($M = 21.24$), as compared to the sessions without a matched comparison ($M = 27.52$). This analysis revealed that engagement with the learning activities was correlated with both the direction of the social comparison and the progress state. This finding might be explained by the positive outcome of the upward social comparison.
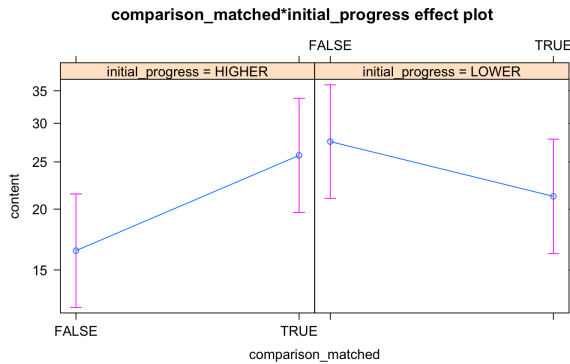

Fig. 3. The interaction effect plot of the matched comparison events and the initial progress state on the number of learning activities(content) in a session.

6.3   Social Comparison as Navigational Support

As stated earlier, PG visualized both individual and comparison group progress per topic. If students pay attention to what OSLM presents to them, we expect them to work on topics where their average progress is behind the comparison group. For this aspect, controlling the comparison group might help them to *navigate* better through PG by visualizing a better target group for them and highlighting the topics where they need to put in more effort.

In this analysis, we used a different type of log data that is related to the topic selections in PG. At each moment of a topic selection, we have recorded the selected comparison group (average/lower/higher), the progress state of the learner (i.e., belonging to the higher/lower progress group), and the progress state of the selected topic (i.e., whether the student's progress is behind/ahead of the comparison group). Using this data, we calculated the ratio of topic selections that happened at each progress state (both student and topic) and the comparison group.

General statistical results revealed that students mainly selected topics where they were behind the selected comparison group average (66% of topic

selections), which shows that students followed the general navigational support provided by OSLM. Following this result, we concentrated on only topic-selections where students selected topics they were behind the comparison group. We wanted to see if having a matched comparison (similar to the idea in Sect. 6.2) had any positive effect on these types of topic selections. We conducted a repeated-measures ANOVA analysis, which revealed a statistically significant effect of matched comparison on the number of *better* topic selections ($F(1, 22) = 5.358, p = .03, \eta_p^2 = 0.2$). This means that when students selected a comparison group that matched their progress state, OSLM guided their attention to the topics where their progress was behind the selected comparison group.

7 Conclusions and Future Work

In this paper, we presented a design of an online practice system, Python Grids (PG), with user-controllable social comparison features that were integrated into an OSLM. We assessed the effect of the controllable features in a semester-long classroom study. We introduced controllability to the social comparison features (i.e., controlling the peer group to compare) as a step to diminish any possible negative effects of social comparison on students and improve the positive motivation, as shown in the related literature.

The study demonstrated that students extensively used the controllability features presented in the system to select and explore different comparison groups. At the same time, the majority of comparison changes (67%) were classified by us as *exploratory*; i.e., where students briefly examined their standing in the context of a different peer group and returned to a stable preferred setting. For most of the students, this stable setting was the average comparison group, which was displayed while the majority of practice activities were selected and performed.

When we checked the impact of social comparison on system usage, we found out that students who performed social comparison events engaged with significantly more learning content activity. Moreover, the social comparison was not associated with the complexity of the learning activities, which suggests that social comparison generally motivated students to practice more. Further, the results showed that performing more *exploratory* social comparison was associated with less practice, as opposed to performing a stable group switch.

The study also revealed some interesting results regarding the importance of social comparison direction. We discovered that in learning sessions where students selected the comparison group that was appropriate to their progress state (i.e. being in the higher/lower progress group and selecting the higher/lower progress group to compare), they performed a higher number of learning activities. This result indicated a positive outcome of controllable comparison group selection. Similarly, we found out that students benefited from the navigational support presented by PG and mostly selected the topics where they were behind the comparison group. Besides, switching to an appropriate comparison group further helped students to select *better* topics, which indicate improved navigation support as a result of the controllability features.

It is important to mention that the findings in this paper only refer to correlations, which do not indicate any causality. Due to a lack of data collection, we could not investigate the effect of prior knowledge on system usage. Similarly, we did not examine how individual differences (e.g., personality traits, achievement goals, and social comparison orientation) affected the system usage. However, to account for individual differences, we conducted linear mixed model analysis, and we believed that given the limited information that we had about students, this approach helped us to discover some important findings. In the future, we plan to conduct a randomized controlled study with students using PG both with and without control features. We expect to see the positive impact of these features on motivation and engagement in these studies and also re-evaluate our findings with a higher number of learners.

References

1. Bates, D., Mächler, M., Bolker, B., Walker, S.: Fitting linear mixed-effects models using lme4. J. Stat. Softw. **67**(1), 1–48 (2015)
2. Brusilovsky, P., Somyürek, S., Guerra, J., Hosseini, R., Zadorozhny, V., Durlach, P.J.: Open social student modeling for personalized learning. IEEE Trans. Emerg. Top. Comput. **4**(3), 450–461 (2016)
3. Brusilovsky, P., Chavan, G., Farzan, R.: Social adaptive navigation support for open corpus electronic textbooks. In: De Bra, P.M.E., Nejdl, W. (eds.) AH 2004. LNCS, vol. 3137, pp. 24–33. Springer, Heidelberg (2004). https://doi.org/10.1007/978-3-540-27780-4_6
4. Brusilovsky, P., et al.: An integrated practice system for learning programming in Python: design and evaluation. Res. Pract. Technol. Enhanced Learn. **13**(1), 1–40 (2018). https://doi.org/10.1186/s41039-018-0085-9
5. 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. Pers. Soc. Psychol. Bull. **25**(10), 1242–1253 (1999)
6. 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 and Knowledge Conference, pp. 454–463 (2017)
7. Dijkstra, P., Kuyper, H., van der Werf, G., Buunk, A.P., van der Zee, Y.G.: Social comparison in the classroom: a review. Rev. Educ. Res. **78**(4), 828–879 (2008)
8. Festinger, L.: A theory of social comparison processes. Hum. Relat. **7**(2), 117–140 (1954)
9. Gibbons, F.X., Buunk, B.P.: Individual differences in social comparison: development of a scale of social comparison orientation. J. Pers. Soc. Psychol. **76**(1), 129 (1999)
10. Guerra, J., Hosseini, R., Somyürek, S., Brusilovsky, P.: An intelligent interface for learning content: combining an open learner model and social comparison to support self-regulated learning and engagement. In: Proceedings of the 21st International Conference on Intelligent User Interfaces, pp. 152–163. ACM (2016)
11. Hosseini, R., et al.: Improving engagement in program construction examples for learning Python programming. Int. J. Artif. Intell. Educ. **30**(2), 299–336 (2020). https://doi.org/10.1007/s40593-020-00197-0

12. Hsiao, I.H., Bakalov, F., Brusilovsky, P., König-Ries, B.: Progressor: social navigation support through open social student modeling. New Rev. Hypermedia Multimedia **19**(2), 112–131 (2013)
13. 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. Eur. J. Soc. Psychol. **31**(5), 557–578 (2001)
14. Kitsantas, A., Zimmerman, B.J., Cleary, T.: The role of observation and emulation in the development of athletic self-regulation. J. Educ. Psychol. **92**(4), 811 (2000)
15. Kuznetsova, A., Brockhoff, P., Christensen, R.: lmerTest package: tests in linear mixed effects models. J. Stat. Softw. **82**(13), 1–26 (2017)
16. Loboda, T.D., Guerra, J., Hosseini, R., Brusilovsky, P.: Mastery grids: an open source social educational progress visualization. In: Rensing, C., de Freitas, S., Ley, T., Muñoz-Merino, P.J. (eds.) EC-TEL 2014. LNCS, vol. 8719, pp. 235–248. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-11200-8_18
17. Muller, D., Fayant, M.P.: On being exposed to superior others: consequences of self-threatening upward social comparisons. Soc. Pers. Psychol. Compass **4**(8), 621–634 (2010)
18. Nguyen, Q., Huptych, M., Rienties, B.: Linking students' timing of engagement to learning design and academic performance. In: Proceedings of the 8th International Conference on Learning Analytics and Knowledge, pp. 141–150 (2018)
19. Rogers, T., Feller, A.: Discouraged by peer excellence: exposure to exemplary peer performance causes quitting. Psychol. Sci. **27**(3), 365–374 (2016)
20. Shi, L., Cristea, A.I.: Learners thrive using multifaceted open social learner modeling. IEEE Multimedia **23**(1), 36–47 (2016)
21. Suls, J., Martin, R., Wheeler, L.: Social comparison: why, with whom, and with what effect? Curr. Dir. Psychol. Sci. **11**(5), 159–163 (2002)
22. Wheeler, L., Martin, R., Suls, J.: The proxy model of social comparison for self-assessment of ability. Pers. Soc. Psychol. Rev. **1**(1), 54–61 (1997)