Investigating Racial and Ethnic Differences in Learning with a Digital Game and Tutor for Decimal Numbers

Xiaolin Ni[1(✉)], Huy A. Nguyen[1], Nicole Else-Quest[2], Alessandro Pagano[3], and Bruce M. McLaren[1]

[1] Carnegie Mellon University, Pittsburgh, PA 15213, USA
xiaolinn@andrew.cmu.edu
[2] University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA
[3] University of Bari Aldo Moro, Bari, Italy

Abstract. This study investigates the effects on different racial/ethnic groups of middle school students when learning with a digital learning game, *Decimal Point*, and a comparable computer tutor. Using data from three classroom studies with 835 students, we compared learning outcomes and engagement among students from racial/ethnic groups that are *well-represented* in STEM (white and Asian) to those that are *underrepresented* in STEM (Black, Hispanic/Latine, Indigenous, and multiracial). Relative to students from underrepresented groups, students from well-represented groups in STEM scored higher on all tests (pre, post, and delayed, despite similar learning gains from pre-to-post and pre-to-delayed) and showed more engagement and less anxiety. The game also enhanced the experience of mastery only among students from well-represented groups. At the same time, students from underrepresented groups learned from the intervention and matched students from well-represented groups in learning efficiency. In short, we found similar learning gains from the game and tutor interventions among students from well-represented and underrepresented racial/ethnic groups, despite the lower performance and lower engagement among students from underrepresented groups. These insights highlight how students from diverse backgrounds may engage differently with educational technology, guiding future efforts in making *Decimal Point* – as well as digital learning tools in general – more inclusive.

Keywords: Educational games · computer tutor · mathematics · race and ethnicity

1 Introduction

Educational technology, and in particular digital learning games, has demonstrated significant benefits in promoting learning at scale across various grade levels and instructional domains [7, 23]. However, as these tools redefine educational experiences and make learning more engaging and dynamic, their universal application in turn raises questions of equity and inclusion [45]. In particular, the design and function of educational technologies often align with the cultural norms and experiences of students from

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2024
R. Ferreira Mello et al. (Eds.): EC-TEL 2024, LNCS 15159, pp. 330–343, 2024.
https://doi.org/10.1007/978-3-031-72315-5_23

racial/ethnic groups that are well-represented in STEM. The National Science Foundation's report [42] indicates that, relative to their shares of the U.S. population, white and Asian students continue to be well-represented in STEM degree attainment, while Black, Hispanic or Latine, and Indigenous students are underrepresented. Thus, insofar as belonging to a well-represented group confers privilege that supports student persistence and success within that specific context, white and Asian students stand to benefit from some privilege within the STEM educational context, while their Black, Hispanic or Latine, and Indigenous peers do not [10, 25][1].

Notwithstanding important and meaningful differences within and between white and Asian American students in the U.S., this privilege is reflected in and perpetuated by more significant representation within the curriculum and generally more resources and educational opportunities [9, 40]. However, the experiences of students from racial/ethnic groups that are underrepresented in STEM (i.e., Black, Hispanic or Latine, and Indigenous students) are also not well represented in the design of educational technology, leading to a potential mismatch between such tools and the needs of these students [9]. This disconnect can perpetuate and even exacerbate existing disparities in education, particularly within STEM fields.

Digital learning games could contribute to addressing this complex issue, in light of prior research showing that learning games can lower the learning barrier, reduce anxiety and promote engagement, with even superior benefits when compared to conventional tutors [23, 28]. Towards investigating whether these effects could help equalize the performance across racial and ethnic groups, we performed a secondary analysis of the math learning game *Decimal Point* and its equivalent non-game tutor, which have been used by over 1,500 students across many classroom studies [13, 27, 28, 30, 31]. With the data of 857 students from studies during three recent years, we grouped students based not on their racial/ethnic *identities* but rather on their racial/ethnic group *representation* within the STEM education context, as indicated by the National Science Foundation [42]. This provisional grouping calls attention to differences associated with inequitable representation in the STEM education context, consistent with prior work [10] comparing racial/ethnic groups that are well-represented in STEM (i.e., white and Asian) to those that are underrepresented in STEM (i.e., Black, Hispanic or Latine, and Indigenous). This approach informs our investigation of how educational technologies are serving different groups, which we operationalized through two research questions:

RQ1*: How do learning outcomes among students belonging to racial/ethnic groups that are well-represented in STEM differ from those of students belonging to racial/ethnic groups that are underrepresented in STEM?*

RQ2*: Are there racial/ethnic group differences in engagement levels when using a game-based and tutor-based educational technology, and, if there are differences, in what ways are there engagement differences?*

[1] Note that, despite the harmful model minority stereotype [12, 17, 26] and meaningful differences in experience between and within *all* racial/ethnic groups – such as, for instance, between white and Asian students – evidence indicates that these racial/ethnic groups experience more privilege and opportunity with respect to STEM relative to Black/African American, Hispanic / Latine, and Indigenous students [42].

By analyzing students' learning outcomes and engagement, we aim to uncover insights that support the development of genuinely inclusive and equitable educational technologies. In doing so, we respond to the imperative for cultural competence in educational design, as advocated by [24], and contribute to the broader discourse on how to mitigate educational disparities through more equitable technology design, development, and deployment. This analysis is poised to influence future educational technologies, ensuring that these powerful tools serve as bridges, rather than barriers, to educational access and success for all students.

2 Background

In the U.S., there is a persistent achievement gap between students from different racial and ethnic backgrounds [16]. The experiences and learning approaches of white children are often privileged as the mainstream standard, as exemplified by the cultural bias in testing [3] and the underrepresentation of minority cultures in learning materials [2]. On the other hand, many sociocultural factors – including inadequately-equipped schools, biased academic tracking, low teacher expectations, and a lack of quality resources – have contributed to lower academic performance among students from racially minoritized communities [11, 35, 39]. In the context of STEM education, such achievement disparities are evident and reflect a range of factors, including inequities within social contexts that determine students' motivation and attainment value [42].

The process of knowledge construction in STEM education research frequently exhibits racial biases, often overlooking power dynamics and privilege. It tends to assess Black, Hispanic/Latine and Indigenous students in comparison to white students, emphasizing deficits rather than recognizing the strengths, skills, and potential of racially minoritized students. This approach in turn favors deficit-based perspectives over asset-based ones when analyzing interventions. These research agendas, policies, and theoretical frameworks often make the erroneous assumption that what benefits all children universally is equally beneficial for children from racially-minoritized communities [22].

This misconception is particularly important to investigate in the area of digital learning games, which aim to employ playful elements to promote immersion, engagement and learning outcomes for all students [50]. Digital games appear to hold strong potential for reducing achievement gaps, given their effects on lowering the learning barriers and market research showing higher gameplay time from Black and Hispanic/Latine youth, when compared to white youth [19]. However, while some research has shown no differences in learning and enjoyment by ethnicity [43], others have reported higher in-game performance among white and Hispanic students than Black/African American students [44]. These mixed results may stem from a range of issues with learning game design, such as a lack of representational diversity among the game developers[2] and the in-game characters [8]. If overlooked, such issues could decrease engagement and cause representational harm by perpetuating problematic images or algorithmic stereotypes [32].

These findings necessitate a critical evaluation of existing digital learning games to ensure that they are not only pedagogically beneficial, but also culturally responsive

[2] https://www.zippia.com/video-game-designer-jobs/demographics/.

and inclusive. To this end, our work examines the digital learning *Decimal Point*, which teaches decimal numbers and operations to middle school students [28]. We consider *Decimal Point* a valuable platform for investigating the impact of learning games on students across different racial/ethnic groups for two reasons. First, it has been used as classroom materials many times in the last decade, by students from a variety of backgrounds [27]. Second, we have found robust evidence for *Decimal Point*'s effectiveness in helping girls catch up to boys in math performance [33]. Thus, we aimed to explore whether *Decimal Point* could be equally effective at bridging the racial/ethnic gap in math performance. In what follows, we describe the game in detail, along with our data collection and analysis procedures.

3 The Learning Game *Decimal Point*

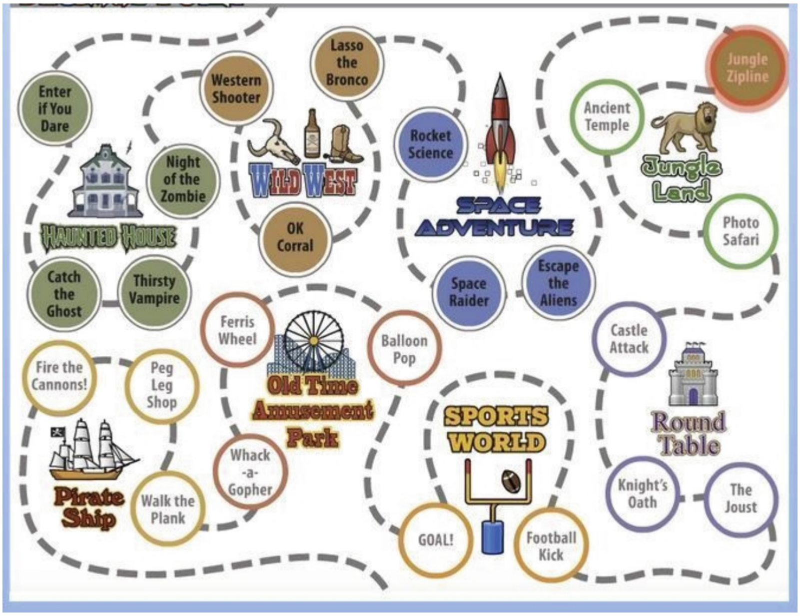

Fig. 1. The main game map in the learning *Decimal Point* which shows the amusement park theme and mini-games.

Decimal Point is a digital learning game that teaches decimal numbers and operations to middle school students [28]. The game features a playful amusement park metaphor, with different theme areas and mini-games (Fig. 1), each covering learning activities that target a specific decimal misconception [15]. In particular, each mini-game (Fig. 2, top) consists of a problem-solving activity (e.g., "Sort the sequence of numbers: 0.5, 0.471, 0.49, 0.365") and a prompted self-explanation activity, expressed as a multiple-choice question (e.g., "Is 0.5 larger or smaller than 0.471? How do you know?"). Students receive immediate corrective feedback after entering their answer and can resubmit any number of times until arriving at the correct answer, which they must do so as to proceed to the next mini-game. During the problem-solving activities, they can also request up to three levels of hints, where the final level provides the solution (i.e., bottom-out hint).

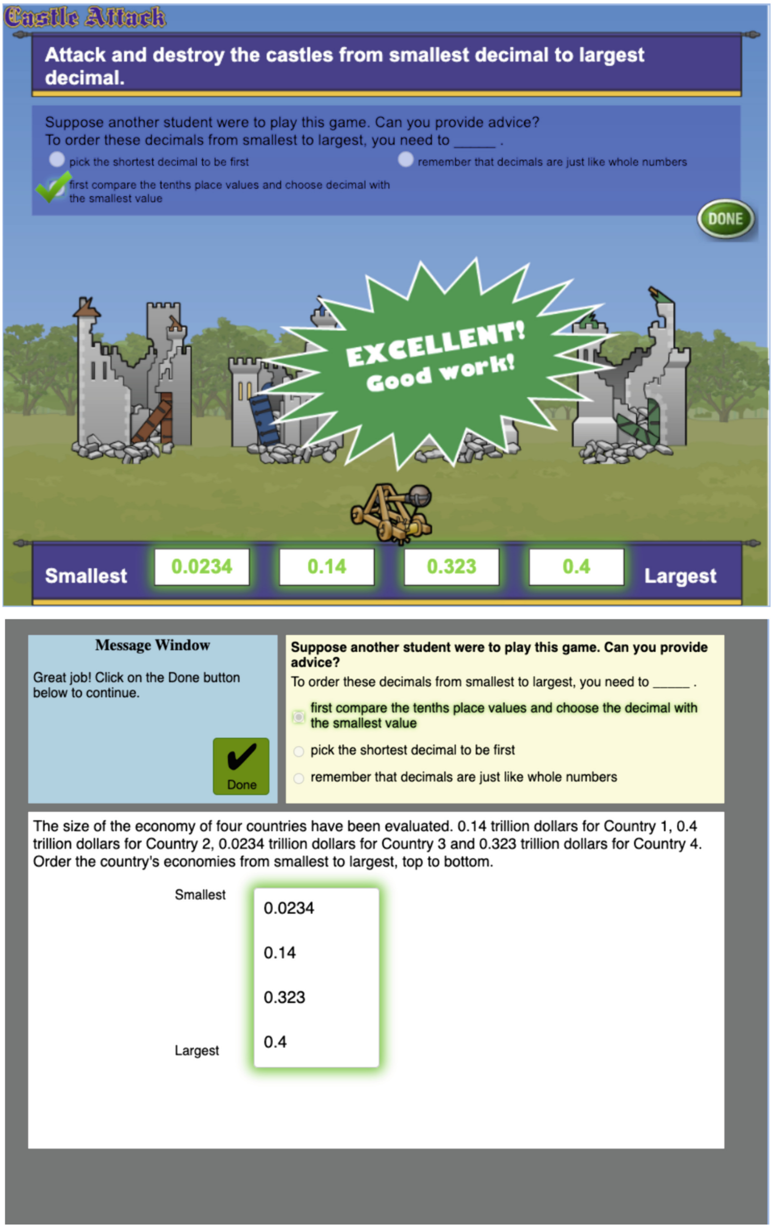

Fig. 2. Screenshot of a sorting mini-game with a self-explanation prompt (top) and the equivalent problem in the decimal tutor (bottom).

During some of our studies we compared *Decimal Point* to a decimal tutor as a control condition (Fig. 2, bottom). The decimal tutor has identical instructional content to *Decimal Point*, but has interactive behavior resembling a traditional intelligent tutoring system [47]. In the analyses reported in this paper we also include students who used the control tutor for learning.

4 Methods

We analyzed data from three classroom studies of the *Decimal Point* game*,* a total of 960 students (see Table 1, column 1), which focused on the three topics of (1) mindfulness induction (n = 237 [5, 34]), (2) self-explanation prompts (n = 480 [41]), and (3) comparing game narratives (n = 243). All three studies, conducted from 2021 to 2023, were conducted in 5th and 6th grade classrooms across public schools in a northeastern U.S city. Each study was conducted with permissions from teachers and parents and followed the procedure described below, which was approved by our university's Institutional Review Board.

For our analyses, we excluded 122 students who did not finish all of the pre- and post-intervention surveys, as well as 3 students whose race/ethnicity data was missing. Thus, our sample then totaled 835 students (see Table 1, column 2) with 551 in the game conditions and 284 in the tutor conditions. These students had an average age of 10.88 years ($SD = 0.65$). In terms of self-reported gender, 414 students identified as girls, 414 as boys, 4 as non-binary, and 3 preferred not to disclose their gender. For analyses of engagement (*RQ2*), we utilized this sample of 835 students. Finally, for analyses of learning (*RQ1*), we considered only the subset of 736 students who finished all of the pretest, posttest and delayed posttest in the study (see Table 1, column 3).

We collected self-reported racial/ethnic identity from a pre-intervention demographic survey. However, because many students were unsure about racial/ethnic categories or their own identities, we opted to use the racial/ethnic data provided by the teachers in our analyses. Of the 835 students who completed the surveys and provided their race/ethnicity, 642 students were white, 101 were Black or African American, 73 were Multiracial or Biracial, 7 were Asian or Pacific Islanders, 11 were Hispanic/Latine, and 1 was Indigenous or Alaskan.

Students participated in the study as part of their regular class activities and could proceed through all of the assignments at their own pace. Each study lasted for 6 days, with the first week covering a demographic survey, a pretest, the learning materials, followed by an evaluation survey and a posttest. A delayed posttest was administered during one day of the second week. The pretest, posttest and delayed posttest consisted of three isomorphic versions of a decimal test that were counterbalanced across students and conditions. Each test covered decimal questions with three question types: near transfer (questions similar to those encountered in the intervention), medium transfer (questions different from the intervention but immediately deducible from the intervention problems), and far transfer (questions related to the same decimal content but conceptually more difficult), for a total of 52 points. From the test scores, we also computed a measure of *learning efficiency* for each student, determined as the z-score of their pre-post or pre-delayed learning gain minus the z-score of the total intervention time [29].

Table 1. Populations of Students in the Three Studies.

Sample	Initial Group, Across Three Studies	Finished Pre and Post Surveys & Provided Race/ethnicity	Finished All Study Materials (pre, post, delayed tests)
Groups Well-Represented in STEM	N/A	649	578
Groups Underrepresented in STEM	N/A	186	158
Total	960	835	736

To complement our learning measures, we collected students' ratings of their experience with the game and tutor in a post-intervention survey covering ten constructs. The constructs and their reliability scores (Cronbach's α) are as follows: *affective engagement* (3 items, $\alpha = .71$); player experience inventory with the *meaning subscale* (3 items, $\alpha = .73$), *mastery subscale* (3 items, $\alpha = .82$) and *challenge subscale* (3 items, $\alpha = .65$); *achievement emotions* (5 items, $\alpha = .90$); *decimal efficacy* (3 items, $\alpha = .83$); *evaluation apprehension* (4 items, $\alpha = .87$); *situational interest* (3 items, $\alpha = .85$); *test self-efficacy* (5 items, $\alpha = .72$), and *test anxiety* (3 items, $\alpha = .72$). Each survey item was rated on

Table 2. The engagement constructs measured in the post-intervention survey. The phrases in brackets were used for students who learned from the tutor instead of the game.

Constructs	Example item
Affective engagement [4]	I felt frustrated or annoyed
Experience of meaning [37]	Playing the game [learning from the tutor] was meaningful to me
Experience of mastery [37]	I felt capable while playing the game [learning from the tutor]
Experience of challenge [37]	The challenges in the game [tutor] were at the right level of difficulty for me
Achievement emotions [1]	Reflecting on my progress in the game [tutor] made me happy
Decimal efficacy [38]	I can do an excellent job on decimal number math assignments
Evaluation apprehension [47]	If I do poorly on this test, people will look down on me
Situational interest [21]	The game [tutor] was exciting
Test self-efficacy [47]	Completing the activity made me doubt my knowledge of math
Test anxiety [6, 48]	During the test, I felt very nervous

a Likert scale from 1 (*strongly disagree*) to 5 (*strongly agree*); see Table 2 for example items.

5 Results

Given the varied representation of the racial/ethnic groups within STEM fields, following [42], we grouped students from white and Asian/Pacific Islander backgrounds as *well-represented* ($n = 649$ for engagement analyses, $n = 578$ for learning analyses). By contrast, Black/African American, Hispanic/Latine, Multiracial/Biracial, and Indigenous students were grouped as *underrepresented* ($n = 186$ for engagement analyses, $n = 158$ for learning analyses). Thus, our research questions focused on comparing measures of learning and engagement between these two broad and diverse racial/ethnic groups. As student data from the two racial/ethnic groups followed a normal distribution and had similar variances, we used analyses of variance (ANOVA) and covariance (ANCOVA) to perform the comparisons.

RQ1: How do learning outcomes among students belonging to racial/ethnic groups that are well-represented in STEM differ from those of students belonging to racial/ethnic groups that are underrepresented in STEM?

Table 3 shows the descriptive statistics of the test scores in each racial/ethnic group. Repeated-measures ANOVA showed significant differences for all students between pretest and posttest $F = 254.40$, $p < .001$, $\eta_p^2 = .257$, as well as between pretest and delayed posttest, $F = 313.89$, $p < .001$, $\eta_p^2 = .299$. In other words, students' performance significantly improved after the intervention. Next, a one-way ANOVA showed that students from well-represented groups performed significantly better than students from underrepresented groups at pretest, $F = 43.27$, $p < .001$, $\eta_p^2 = .056$. With pretest scores as covariates, we then used two-way ANCOVA to assess the effects of racial/ethnic group and learning platform on post-intervention test performance. Compared to students from underrepresented groups, students from well-represented groups had significantly higher posttest scores, $F = 4.09$, $p = .04$, $\eta_p^2 = .006$, and higher delayed posttest scores, $F = 4.85$, $p = .03$, $\eta_p^2 = .007$. However, there was no significant main effect of the learning platform or interaction effect on posttest scores. For learning efficiency, a two-way ANOVA showed no significant difference between the two racial/ethnic groups in pre-post efficiency, $F = 1.76$, $p = .18$, $\eta_p^2 = .002$, and in pre-delayed efficiency, $F = 3.56$, $p = .06$, $\eta_p^2 = .005$. The effects of the learning platform and its interaction with the racial/ethnic groups on learning efficiency were not significant. In sum, the intervention improved learning outcomes among students from both underrepresented and well-represented racial/ethnic groups.

RQ2: Are there racial/ethnic group differences in engagement levels when using a game-based and tutor-based educational technology, and, if there are differences, in what ways are there engagement differences?

Table 4 shows the descriptive statistics of the engagement constructs in each racial/ethnic group. As four of those constructs – evaluation apprehension, situational interest, test self-efficacy and test anxiety – were only collected in two out of three studies ($n = 631$, with 491 students from well-represented groups and 140 students from underrepresented groups), we only considered this sub-sample of students when analyzing these constructs.

Table 3. Descriptive statistics of test scores by racial/ethnic group.

Racial/ethnic group	Pretest	Posttest	Delayed Posttest
Underrepresented groups (n = 158)	16.35 (8.66)	19.29 (9.11)	19.62 (9.51)
Well-represented groups (n = 578)	22.52 (10.90)	26.27 (10.87)	27.16 (11.52)

With regards to the main effects of the learning platform, students playing the game reported more affective engagement ($F = 5.29$, $p = .02$, ${\eta_p}^2 = .005$), achievement emotions ($F = 7.46$, $p < .001$, ${\eta_p}^2 = .009$), and situational interest ($F = 13.56$, $p < .001$, ${\eta_p}^2 = .021$), in addition to lower test anxiety ($F = 5.45$, $p = .02$, ${\eta_p}^2 = .009$), than those using the tutor. For the main effects of the racial/ethnic group, students from well-represented groups reported higher affective engagement ($F = 3.90$, $p = .05$, ${\eta_p}^2 = .005$), decimal efficacy ($F = 7.69$, $p < .01$, ${\eta_p}^2 = .009$) and test self-efficacy ($F = 18.23$, $p < .001$, ${\eta_p}^2 = .028$), as well as lower test anxiety ($F = 6.83$, $p < .01$, ${\eta_p}^2 = .011$), compared to those from underrepresented groups. Finally, there was a significant interaction effect between the racial/ethnic group and learning platform on students' experience of mastery, $F = 5.25$, $p = .02$, ${\eta_p}^2 = .006$ (Fig. 3). Pairwise comparison showed that the game ($M = 3.37$, $SD = 1.02$) led to more experience of mastery than the tutor ($M = 3.11$, $SD = 1.07$) for students from well-represented groups, $F = 8.75$, $p < .01$, ${\eta_p}^2 = .013$. However, this effect was not significant among students from underrepresented groups, $F = 0.91$, $p = .34$, ${\eta_p}^2 = .005$.

Table 4. Descriptive statistics for the survey constructs which yielded significant main effects of the learning platform (left) and the racial/ethnic group (right).

Construct	Game	Tutor	Construct	Under-represented	Well-represented
Affective engagement	3.21 (1.15)	3.04 (0.99)	Affective engagement	3.02 (1.11)	3.19 (1.10)
Achievement emotions	3.28 (1.04)	3.01 (0.94)	Decimal efficacy	3.21 (0.97)	3.43 (0.99)
Situational interest	3.19 (1.19)	2.76 (1.01)	Test self-efficacy	3.27 (0.77)	3.62 (0.86)
Test anxiety	2.54 (1.10)	2.80 (1.02)	Test anxiety	2.88 (1.08)	2.59 (1.06)

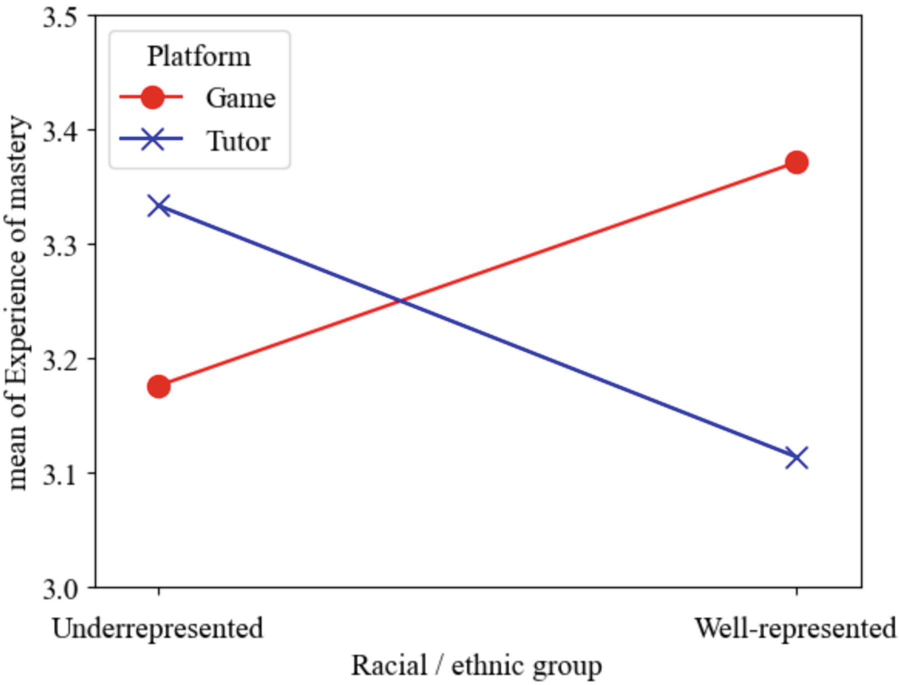

Fig. 3. Interaction effect between the learning platform and racial/ethnic group on experience of mastery.

6 Discussion and Conclusion

This work examined the effects of the digital learning game *Decimal Point* and an equivalent decimal tutor on the learning and engagement of students from different racial/ethnic backgrounds. Our research is motivated by prior results on the gender effects of *Decimal Point* which showed that girls started with lower pretest scores but caught up to boys by posttest [33]. While our results showed that performance gaps remained between students from well-represented and underrepresented groups after game play and tutor use, we have also identified racial/ethnic group differences across learning and enjoyment measures in the game and the tutor. The implications of these findings are discussed below.

First, we found that students from racial/ethnic groups well-represented in STEM scored higher at pretest, posttest and delayed posttest relative to those from racial/ethnic groups underrepresented in STEM. This performance gap is consistent with the systemic bias in K-12 learning outcomes [14] and in educational technologies [40]. This bias may arise from educational content favoring the cultural norms and learning styles of majority groups, potentially leading to disparities in test scores and long-term knowledge retention [12, 25]. In the case of the *Decimal Point* game and the decimal tutor, the goal is to provide additional practice opportunities to students with basic knowledge of decimal numbers, but not to offer any remedial resources to students who may need them. Therefore, incorporating more worked examples and review materials could benefit students with lower prior knowledge [29]. On the other hand, we found that students from underrepresented groups had lower pretest performance but similar learning efficiency as students from well-represented groups. This finding is consistent with the results reported in [18] and suggests that, with more practice opportunities, the learning trajectory of students from underrepresented groups can improve such that it matches or surpasses that of students from well-represented groups.

When examining racial/ethnic group differences and learning platform effects on engagement, we identified several notable trends. First, compared to the tutor, *Decimal Point* led to more affective engagement, achievement emotion and situational interest, as well as lower test anxiety. These results are in line with the established benefits of learning games, which can promote both learning outcomes and motivation [23]. On the other hand, students from underrepresented groups had lower affective engagement, decimal efficacy and test efficacy, as well as higher test anxiety, than students from well-represented groups, regardless of the learning platform. As discussed earlier, the lack of remedial resources could hinder students from underrepresented groups in absorbing the learning content, resulting in worse self-reported measures of engagement, efficacy and anxiety. Another possibility lies in the game and the tutor's emphasis on textual information, including the question prompts, hint content and feedback messages (Fig. 2). Given the sharp decline in reading skills of middle school students, particularly students from racially minoritized groups and other underserved communities, during and after the COVID-19 pandemic [20], the relatively large amount of textual information in these two technologies may cause difficulties for those students in learning the decimal content, a phenomenon that we informally observed during the classroom studies. Thus, identifying ways to improve the learning materials' accessibility (e.g., using non-textual hints [36]) would likely promote students' learning and engagement more equitably.

Finally, we identified an interaction effect between racial/ethnic group and learning platform, where the game led to more experience of mastery than the tutor among students from well-represented groups, but not among those from underrepresented groups. While the two learning platforms featured identical learning content, we had hypothesized that the game's more prominent and playful feedback would enable a greater sense of mastery. However, our finding suggests this feedback did not resonate as well with students from underrepresented groups. Future research should therefore investigate how students react to each component of the game, through field observations or automated affect detectors [49], to evaluate its effectiveness on different student populations.

In conclusion, our analyses have revealed differences in several aspects of learning and engagement between racial/ethnic groups, in the context of learning decimal numbers through a game and a tutor. We found that students whose racial/ethnic backgrounds are well-represented in STEM were more engaged and less anxious than those from underrepresented groups. While the game has been shown to help narrow the gender gap in math performance [33], our findings indicate that there remains important future research to help improve its inclusiveness to students from diverse racial/ethnic groups. Ultimately, educational technology design should consider both pedagogical objectives and students' diverse backgrounds and interests to foster genuinely inclusive and engaging learning experiences.

Acknowledgements. This work was supported by the National Science Foundation (NSF) Award #DRL-2201796. The opinions expressed are those of the authors and do not represent the views of NSF.

References

1. Abeele, V.V., Spiel, K., Nacke, L., Johnson, D., Gerling, K.: Development and validation of the player experience inventory: a scale to measure player experiences at the level of functional and psychosocial consequences. Int. J. Hum.-Comput. Stud. **135**, 102370 (2020)
2. Armstrong, A.L.: The representation of social groups in US educational materials and why it matters: a research overview. New Am. (2021)
3. Banks, K.: A comprehensive framework for evaluating hypotheses about cultural bias in educational testing. Appl. Meas. Educ. **19**(2), 115–132 (2006). https://doi.org/10.1207/s15324818ame1902_3
4. Ben-Eliyahu, A., Moore, D., Dorph, R., Schunn, C.D.: Investigating the multidimensionality of engagement: affective, behavioral, and cognitive engagement across science activities and contexts. Contemp. Educ. Psychol. **53**, 87–105 (2018)
5. Bereczki, E.O., Takacs, Z.K., Richey, J.E., Nguyen, H.A., Mogessie, M., McLaren, B.M.: Mindfulness in a digital math learning game: insights from two randomized controlled trials. J. Comput. Assist. Learn. jcal.12971 (2024). https://doi.org/10.1111/jcal.12971
6. Chung, B.G., Ehrhart, M.G., Holcombe Ehrhart, K., Hattrup, K., Solamon, J.: Stereotype threat, state anxiety, and specific self-efficacy as predictors of promotion exam performance. Group Organ. Manag. **35**(1), 77–107 (2010)
7. Clark, D.B., Tanner-Smith, E.E., Killingsworth, S.S.: Digital games, design, and learning: a systematic review and meta-analysis. Rev. Educ. Res. **86**(1), 79–122 (2016)
8. Cooke, L.: Metatuning: a pedagogical framework for a generative STEM education in game design-based learning. In: 2016 IEEE Integrated STEM Education Conference (ISEC), pp. 207–214. IEEE (2016)
9. Finkelstein, S., Yarzebinski, E., Vaughn, C., Ogan, A., Cassell, J.: The effects of culturally congruent educational technologies on student achievement. In: Lane, H.C., Yacef, K., Mostow, J., Pavlik, P. (eds.) AIED 2013. LNCS, vol. 7926, pp. 493–502. Springer, Heidelberg (2013). https://doi.org/10.1007/978-3-642-39112-5_50
10. French, A.M., et al.: An intersectional application of expectancy-value theory in an undergraduate chemistry course. Psychol. Women Q. **47**(3), 299–319 (2023). https://doi.org/10.1177/03616843231153390
11. Gregory, A., Skiba, R.J., Noguera, P.A.: The achievement gap and the discipline gap: two sides of the same coin? Educ. Res. **39**(1), 59–68 (2010). https://doi.org/10.3102/0013189X09357621
12. Grindstaff, K., Mascarenhas, M.: "No one wants to believe It": manifestations of white privilege in a stem-focused college. Multicult. Perspect. **21**(2), 102–111 (2019). https://doi.org/10.1080/15210960.2019.1572487
13. Hou, X., Nguyen, H.A., Richey, J.E., McLaren, B.M.: Exploring how gender and enjoyment impact learning in a digital learning game. In: Bittencourt, I., Cukurova, M., Muldner, K., Luckin, R., Millán, E. (eds.) AIED 2020. LNCS, vol. 12163, pp. 255–268. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-52237-7_21
14. Howard, T.C.: Why Race and Culture Matter in Schools: Closing the Achievement Gap in America's Classrooms. Teachers College Press (2019)

15. Isotani, S., McLaren, B.M., Altman, M.: Towards intelligent tutoring with erroneous examples: a taxonomy of decimal misconceptions. In: Aleven, V., Kay, J., Mostow, J. (eds.) ITS 2010. LNCS, vol. 6095, pp. 346–348. Springer, Heidelberg (2010). https://doi.org/10.1007/978-3-642-13437-1_66
16. Kena, G., et al.: The condition of education 2015. Natl. Cent. Educ. Stat. **144** (2015)
17. King, G.P., Russo-Tait, T., Andrews, T.C.: Evading race: STEM faculty struggle to acknowledge racialized classroom events. CBE—Life Sci. Educ. **22**(1), ar14 (2023). https://doi.org/10.1187/cbe.22-06-0104
18. Koedinger, K.R., Carvalho, P.F., Liu, R., McLaughlin, E.A.: An astonishing regularity in student learning rate. Proc. Natl. Acad. Sci. **120**(13), e2221311120 (2023). https://doi.org/10.1073/pnas.2221311120
19. Kolko, J., Strohm, C.Q., Lonian, A.: Hispanics and Blacks game more. Forrester Consum. Technographics N. Am. Boston MA (2003)
20. Kuhfeld, M., Lewis, K., Peltier, T.: Reading achievement declines during the COVID-19 pandemic: evidence from 5 million US students in grades 3–8. Read. Writ., 1–17 (2022)
21. Linnenbrink-Garcia, L., et al.: Measuring situational interest in academic domains. Educ. Psychol. Meas. **70**(4), 647–671 (2010)
22. Loewenberg, D.: Mathematical Proficiency for All Students: Toward a Strategic Research and Development Program in Mathematics Education. Rand Corporation (2003)
23. Mayer, R.E.: Computer games in education. Annu. Rev. Psychol. **70**, 531–549 (2019)
24. Mayfield, V.: Cultural Competence Now: 56 Exercises to Help Educators Understand and Challenge Bias, Racism, and Privilege. ASCD (2020)
25. McGee, E.O.: Interrogating structural racism in stem higher education. Educ. Res. **49**(9), 633–644 (2020). https://doi.org/10.3102/0013189X20972718
26. McGee, E.O., Thakore, B.K., LaBlance, S.S.: The burden of being "model": racialized experiences of Asian STEM college students. J. Divers. High. Educ. **10**(3), 253 (2017)
27. McLaren, B.M.: Decimal point: a decade of learning science findings with a digital learning game. In: Ilic, P., Wegerif, R., Casebourne, I. (eds.) Artificial Intelligence in Education: The Intersection of Technology and Pedagogy. Springer Verlag Germany. Part of the "Artificial Intelligence-Enhanced Software and Systems Engineering" (AI-SSE) Series (2024)
28. McLaren, B.M., Adams, D.M., Mayer, R.E., Forlizzi, J.: A computer-based game that promotes mathematics learning more than a conventional approach. Int. J. Game-Based Learn. IJGBL **7**(1), 36–56 (2017)
29. McLaren, B.M., Lim, S.-J., Koedinger, K.R.: When and how often should worked examples be given to students? New results and a summary of the current state of research. In: Proceedings of the 30th Annual Conference of the Cognitive Science Society, pp. 2176–2181 (2008)
30. McLaren, B.M., Richey, J.E., Nguyen, H., Hou, X.: How instructional context can impact learning with educational technology: lessons from a study with a digital learning game. Comput. Educ. **178**, 104366 (2022). https://doi.org/10.1016/j.compedu.2021.104366
31. McLaren, B.M., Richey, J.E., Nguyen, H.A., Mogessie, M.: Focused self-explanations lead to the best learning outcomes in a digital learning game. In: Proceedings of the 17th International Conference of the Learning Sciences, pp. 1229–1232. ISLS (2022)
32. Misra, R., Eyombo, L., Phillips, F.T.: Digital games for minority student engagement: emerging research and opportunities: emerging research and opportunities (2019)
33. Nguyen, H.A., Hou, X., Richey, J.E., McLaren, B.M.: The impact of gender in learning with games: a consistent effect in a math learning game. Int. J. Game-Based Learn. IJGBL. **12**(1), 1–29 (2022)
34. Nguyen, H.A., Takacs, Z.K., Bereczki, E.O., Richey, J.E., Mogessie, M., McLaren, B.M.: Investigating the effects of mindfulness meditation on a digital learning game for mathematics. In: Rodrigo, M.M., Matsuda, N., Cristea, A.I., Dimitrova, V. (eds.) AIED 2022. LNCS, vol. 13355, pp. 762–767. Springer, Cham (2022). https://doi.org/10.1007/978-3-031-11644-5_80

35. Oakes, J.: Keeping Track: How Schools Structure Inequality. Yale University Press (2005)
36. Ostrow, K., Heffernan, N.: Testing the multimedia principle in the real world: a comparison of video vs. text feedback in authentic middle school math assignments. In: Educational Data Mining 2014 (2014)
37. Pekrun, R.: Progress and open problems in educational emotion research. Learn. Instr. **15**(5), 497–506 (2005)
38. Pintrich, P.R.: A manual for the use of the Motivated Strategies for Learning Questionnaire (MSLQ) (1991)
39. Reardon, S.F.: The widening academic achievement gap between the rich and the poor. In: Social Stratification, pp. 536–550. Routledge (2018)
40. Reyes, L.H., Stanic, G.M.: Race, sex, socioeconomic status, and mathematics. J. Res. Math. Educ. **19**(1), 26–43 (1988)
41. Richey, J.E., et al.: Understanding gender effects in game-based learning: the role of self-explanation. In: Proceedings of the International Conference on Artificial Intelligence in Education. Springer (in press)
42. National Center for Science and Engineering Statistics (NCSES): Diversity and STEM: Women, minorities, and persons with disabilities 2023. Spec. Rep. NSF 23–315 (2023)
43. Shute, V., et al.: Maximizing learning without sacrificing the fun: stealth assessment, adaptivity and learning supports in educational games. J. Comput. Assist. Learn. **37**(1), 127–141 (2021)
44. Smith, G., Fulwider, C., Liu, Z., Lu, X., Shute, V.J., Li, J.: Examining students' perceived competence, gender, and ethnicity in a digital stem learning game. Int. J. Game-Based Learn. IJGBL **12**(1), 1–17 (2022)
45. Smolansky, A., Nguyen, H.A., Kizilcec, R.F., McLaren, B.M.: Equity, diversity, and inclusion in educational technology research and development. In: Wang, N., Rebolledo-Mendez, G., Dimitrova, V., Matsuda, N., Santos, O.C. (eds.) AIED 2023. CCIS, vol. 1831, pp. 57–62. Springer, Cham (2023). https://doi.org/10.1007/978-3-031-36336-8_8
46. Spencer, S.J., Steele, C.M., Quinn, D.M.: Stereotype threat and women's math performance. J. Exp. Soc. Psychol. **35**(1), 4–28 (1999)
47. VanLehn, K.: The behavior of tutoring systems. Int. J. Artif. Intell. Educ. **16**(3), 227–265 (2006)
48. Veit, C.T., Ware, J.E.: Mental health inventory. Psychol. Assess. (1983)
49. Yang, T.-Y., Baker, R.S., Studer, C., Heffernan, N., Lan, A.S.: Active learning for student affect detection. In: Proceedings of the 12th International Conference on Educational Data Mining, EDM 2019, Montréal, Canada, 2–5 July 2019. International Educational Data Mining Society (IEDMS) 2019, pp. 208–217 Université du Québec; Polytechnique Montréal (2019)
50. Zeng, J., Parks, S., Shang, J.: To learn scientifically, effectively, and enjoyably: a review of educational games. Hum. Behav. Emerg. Technol. **2**(2), 186–195 (2020). https://doi.org/10.1002/hbe2.188