Exploring Student Identity in Adaptive Learning Systems through Qualitative Data

Clara Belitz[1][0009−0002−1960−0914], HaeJin Lee[1][0009−0000−0260−0462], Nidhi Nasiar[2][0009−0006−7063−5433], Stephen E. Fancsali[3][0000−0002−4016−8145] Frank Stinar[1][0009−0000−3152−6281], Husni Almoubayyed[3][0000−0001−5717−9310], Steve Ritter[3][0000−0003−2807−9390], Ryan S. Baker[2][0000−0002−3051−3232], Jaclyn Ocumpaugh[2][0000−0002−9667−8523], and Nigel Bosch[1][0000−0003−2736−2899]

[1] University of Illinois Urbana–Champaign, Champaign IL 61820, USA
cbelitz2@illinois.edu
[2] University of Pennsylvania, Philadelphia PA 19104 USA
[3] Carnegie Learning, Inc., Pittsburgh PA 15219 USA

Abstract. Adaptive learning systems are increasingly common in U.S. classrooms, but it is not yet clear whether their positive impacts are realized equally across all students. This study explores whether nuanced identity categories from open-ended self-reported data are associated with outcomes in an adaptive learning system for secondary mathematics. As a measure of impact of these social identity data, we correlate student responses for 3 categories: race and ethnicity, gender, and learning identity—a category combining student status and orientation toward learning—and total lessons completed in an adaptive learning system over one academic year. Results show the value of emergent and novel identity categories when measuring student outcomes, as learning identity was positively correlated with mathematics outcomes across two statistical tests.

Keywords: Student Identity · Equity · Adaptive Learning Systems

1 Introduction

AI-driven technologies are increasingly used to improve student learning outcomes and support teachers [22,14], and researchers have aimed to ensure that adaptive learning systems benefit all types of students equitably [28,9]. Choosing which inherent or acquired characteristics to compare when measuring these outcomes, however, is often glossed over in algorithmic fairness methods [1]. This paper explores those choices, examining a new way to define the groups against which equity and algorithmic bias are measured in educational settings.

Many current questions around equity in AI-driven systems utilize census-style categories, often from preexisting sources [18]. Such data offer practical ways to examine differences at scale, but can lack both nuance and breadth, obscuring some variables that may be needed to ensure equity [27]. For example, in classroom contexts, students' intellectual identities could significantly affect how they approach a subject, as has been shown with math identity [8,23]. Both missing data (e.g., learning identities) and underspecified data (e.g., oversimplified race and gender categories) can make it harder to understand how best to support learning for all students.

To capture students' identities and characteristics in a more nuanced and accurate approach, this work asks the following questions:

- **RQ1:** How does a free-response measure of student identity compare to school-provided demographic information in both occurrence and content?
- **RQ2:** Do free-response measures of student identity capture predictive information for learning that is not otherwise found in school-provided demographic data?

This paper presents identity data from students using a validated measure, collecting qualitative (descriptive and free-form) responses from middle and high school students across two academic years (2021–2023). The presence of gender and race responses, commonly used demographic categories, was not correlated with total completed lessons in an AI-driven adaptive math system, while the presence of a novel category that emerged from the data, *learning identity*, had significant positive correlation across two tests.

2 Related Work

Educational research has long recognized that complex sociocultural factors influence whether student opportunities and outcomes are fair and equitable [25]. Education researchers studying artificial intelligence, data science, and learning analytics have examined ways to support marginalized students [24] and whether commonly used algorithms exhibit signs of bias [2]. They have also proposed frameworks for ensuring fairness and equity [15,3]. Similarly, researchers have investigated identities that may not be legal or protected classifications (e.g., math identity) but are nonetheless tied to power dynamics [21,11].

Although the complexity of ensuring equitable learning opportunities in AI-driven learning systems is well known, access to data that would support deep analysis is not always possible, if these data exist at all [16]. Categorization and classification fundamentally require making choices about which information is relevant and important [5]. Making choices is not, in and of itself, problematic, since categorization is crucial to retrieving and using information, but these processes and decisions are often glossed over in artificial intelligence pipelines [19,4]. Categories like race and gender are frequently characterized as inflexible and superficial, though they are socially constructed and nuanced [13,17]. Surveys have tried to develop more flexible categorizations, but may still offer incomplete classifications [13].

3 Method

This project adapts the Twenty Statements Test [20,12], which uses a constructivist approach to developing categories, rather than choosing features a priori to represent the classes of interest. In this paper, a subset of representative categories is described in depth, while the entire coding schema is available online.

3.1 Student Demographics and Learning Context

Our study examines data from middle and high school students in a small city in the northeastern United States. Students used MATHia as part of their regular instruction in 6th to 8th grade mathematics, geometry, and algebra. As an AI-driven adaptive system, MATHia provides personalized hints and just-in-time feedback as learners progress [22]. Given this capability, it is important to understand whether these lessons, hints, and feedback are working equally effectively for all students. A typical school year encompasses 90–120 lessons, depending on standards, individual teacher and administrator preferences, and other factors impacting customization. Lesson count is a more effective measure of student outcomes than total time spent with MATHia [10]. The mean lessons completed per student in our study was 104.7 (SD = 44.5). Lesson count completion per student had a non–normal distribution. The school district, which collects demographic data as part of its standard practice, shared these data under the guidance of the University of Illinois Institutional Review Board #IRB24-1190.

3.2 Twenty Statements Test

The Twenty Statements Test (TST) uses a free-form answer format to elicit self-concept by having individuals answer the question "Who am I?" up to 20 different ways in a short amount of time [20]. Responses are then coded by researchers both into well-established categories, such as the schema developed by [12], and into categories that arise inductively via thematic analysis [6]. The survey design permits students to respond with multiple answers that fall into the same category (e.g., multiple nationalities). The survey was delivered online, where each response had its own text box. The survey was available in the four languages most commonly spoken at home by students in the school district. The survey was piloted with data from 53 students, with 19.7 responses on average per student (SD=1.4) in the 2020–2021 academic year. Full instructions regarding the development of the schema, interrater reliability, and inclusion criteria are available online. The rest of this paper presents results from students in the 2021–2023 academic years (n=118). Note that the responses reported in this paper are not verbatim, to respect student privacy, but are emblematic of the type of responses seen.

4 Results

Overall, student-provided demographic categories aligned with those provided by the school. A few students provided responses that actively disagreed with their demographic data, including students with ethnic identities from countries outside of the U.S. that did not map neatly to American racial boxes ($n = 5$).

Table 1. Student responses (rows) vs. school-provided categories (columns). The category of ≥ 1 *Race* + ≥ 1 *Ethnicity* refers to students who provided at least one racial identity as well as at least one ethnic identity. The categories of *Multiple Races* and *Multiple Ethnicities* refer to those who provided at least two racial identities but no ethnic identities, and at least two ethnic identities but no racial identities, respectively.

	African-American	Asian	Hispanic	White	Multi-Race, Non-Hispanic
One Race	6	1	2	1	0
Multiple Races	0	0	0	0	1
≥ 1 Race + ≥ 1 Ethnicity	9	0	0	1	1
One Ethnicity	9	0	1	4	1
Multiple Ethnicities	1	1	2	1	0
Ethnicity + American	1	1	0	0	0
Person of Color + Ethnicity	1	0	0	0	0
No Ethnic or Racial Response	43	0	7	22	2
Total (2021–2023)	70	3	12	29	5

However, most students did not relay this information at all, or–when they did–provided a more nuanced or specific category than the one provided by the school. Importantly, more students provided responses related to learning identity than related to gender or racial categories, which are commonly used demographic features. As such, learning identity was chosen for further examination of the importance of novel, emergent categories. Specifically, we compare student performances through lesson completion (a measure that indicates how far through the MATHia cuirriculum a student has progressed) across these categories.

RQ1: How does a free-response measure of student identity compare to school-provided demographic information in both occurrence and content?

In total, 118 students provided valid responses across the 2021–2022 ($n = 62, \mu = 19.8, SD = 0.81$), and 2022–2023 ($n = 56, \mu = 19.4, SD = 2.6$) academic years. Race and ethnicity data are combined here to reflect the school-provided demographic data, even as we recognize that they are distinct dimensions. Overall, 45 out of 118 students (38.1%) indicated a racial and/or ethnic identity. The students who did provide racial and ethnic responses were classified using an emergent schema of Middle Eastern or North African, Multiracial, East Asian, South Asian, Pacific Islander, Latino/a, White, and Other or Unclear. Though the school uses the group "African-American," the term "African" only appeared in this set of responses to explicitly identify a student as from a specific African country or to differentiate a specific relationship to place (e.g., "from North Africa"). Students instead broadly referred to themselves as Black or from specific locations. Students in other racial groups represented in this dataset exhibited a similar pattern, indicating specific ethnic origins like Portuguese, Italian,

Table 2. Spearman's *rho* and Mann–Whitney *U* test comparing identity (presence/absence) to lessons completed. **Bold** indicates significance.

Category	*rho*	*p*	*U*	*p*
Gender	-.029	.753	1748.5	.753
Race/Ethnicity	-.034	.716	1708.5	.717
Learning Identity	**.227**	**.013**	**970.5**	**.014**

or Irish for White students and Japanese, Filipino, or Chinese for Asian students (see Table 1, which uses school-provided demographic language).

Just over half of respondents (65 of 118; 55.1%) provided gender responses. Nearly all of these (n=64) aligned with the school demographics (40 girls, 24 boys), but 1 nonbinary student did not. Gendered responses ranged from implicit (e.g., familial roles like son; pronouns like she/her) to explicit statements (e.g., "I'm a boy").

Learning identity emerged from two categories (i.e., intellectual concerns and student role), which have both appeared in prior work that investigates the effects of (a) performance beliefs and (b) interest in mathematics identity development [7]. Intellectual concerns emerged at roughly the same rate as gendered responses (62 of 118; 52.3%), while student role appeared at higher rates (79 of 118; 66.9%). Intellectual role was most likely to be positive (75.8% positive; 6.1% negative; 18.2% neutral), while student role (which often emerged as "I am a student") was most likely to be neutral (53.6% neutral; 41.9% positive; 4.5% negative). Notably, negatively valenced responses for student role and intellectual concerns were always accompanied by at least one positive valence response. Such responses for these two categories of student role and intellectual concerns frequently overlapped, with negligible qualitative difference. For example, one student reported both "I am an honors student," coded as student role for the explicit inclusion of school, as well as "I am good at math," coded as intellectual concerns for the relationship to self-conception. Therefore these two forms of response are treated as a single category in our analysis.

RQ2: Do free-response measures of student identity capture predictive information for learning that is not otherwise found in school-provided demographic data?

Because **RQ2** explores whether the presence of an identity is significantly related to MATHia outcomes, specific identity responses were binarized as *present* or *not present*. We performed two statistical tests: the correlation between total lessons completed in MATHia and the presence of each of the categories of interest as well as the difference in completion count across presence and absence of those categories (Table 2). The presence of gender and race responses had no significant correlation with total lessons completed in MATHia across the school year. In contrast, learning identity did ($rho = .227$, $p = .013$; $U = 970.5$, $p = .014$).

5 Discussion and Conclusion

Our results demonstrate the importance of emergent categories, like learning identity, when exploring the impact of adaptive learning systems. Similarly, we show (1) how students' TST responses differ from the demographic data their schools use to describe them and (2) how students' relationship to the schools emerges in the data. Our results show more students mentioned their learning identity than did their gender. Notably, many students grounded their learning identities in a highly specific fashion (e.g., specific subjects in which they excel or struggle). Such nuances are not captured by typical school demographic categories, but are likely meaningful to interest in learning and subsequent educational outcomes. As such, finding ways to expand beyond traditional demographics, perhaps by working to elicit a broader range of learning identity categories, may help researchers to better understand whether student success in adaptive learning systems is occurring in an equitable fashion.

Learning identities also emerged more often than racial or ethnic ones in both school years, alluding to the significance of this designation to students. Although the context of completing the TST while at school may have contributed to the higher rates for these categories, it is also likely that these are important to students' everyday lived experiences with adaptive learning systems. These results may raise questions about the relevance of the social identities by which (in)equity is measured for individuals in this specific context. We hypothesize that, as with math identity, learning identity may be an important factor for student performance [26,23]. Given that our results demonstrate the positive impact of learning identity on outcomes, future designs of AI-driven educational systems might consider how they can build up students' learning identities and math confidence. While this work has not yet explored all the possible emergent categories from the data, nor the intersections of multiple categories, this reflects the rich nature of qualitative work and will be pursued in future research. Though this type of qualitative work is time-intensive, the varied data may help to avoid missing differentiations that matter for detecting nuanced algorithmic bias in adaptive learning systems.

Though the TST takes more time and resources than using extant demographics, these ground-up approaches provide greater detail for understanding the contextual realities of algorithmic systems and their sociotechnical impacts on classrooms. As demonstrated by our results, a broader range of more nuanced categories is an important part of understanding how algorithmic bias emerges in AI systems.

Acknowledgments. This research was supported by NSF grant no. 2000638. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

References

1. Abdu, A.A., Pasquetto, I.V., Jacobs, A.Z.: An empirical analysis of racial categories in the algorithmic fairness literature. In: 2023 ACM Conference on Fairness, Accountability, and Transparency. pp. 1324–1333. FAccT '23, Association for Computing Machinery, Chicago, IL (2023). https://doi.org/10.1145/3593013.3594083
2. Ashwin, T.S., Biswas, G.: Identifying and mitigating algorithmic bias in student emotional analysis. In: Artificial Intelligence in Education: 25th International Conference, AIED 2024. pp. 89–103. Springer Nature Switzerland, Cham (2024), https://link.springer.com/10.1007/978-3-031-64299-9_7
3. Belitz, C., Lee, H., Nasiar, N., Fancsali, S.E., Ritter, S., Almoubayyed, H., Baker, R.S., Ocumpaugh, J., Bosch, N.: Hierarchical dependencies in classroom settings influence algorithmic bias metrics. In: Proceedings of the 14th Learning Analytics & Knowledge Conference. pp. 210–218. ACM, Kyoto Japan (2024). https://doi.org/10.1145/3636555.3636869
4. Belitz, C., Ocumpaugh, J., Ritter, S., Baker, R.S., Fancsali, S.E., Bosch, N.: Constructing categories: Moving beyond protected classes in algorithmic fairness. Journal of the Association for Information Science and Technology **74**(6), 663–668 (2022). https://doi.org/10.1002/asi.24643
5. Bowker, G.C., Star, S.L.: Sorting Things Out: Classification and Its Consequences. MIT Press, Cambridge, Mass (1999)
6. Braun, V., Clarke, V.: Using thematic analysis in psychology. Qualitative Research in Psychology **3**(2), 77–101 (2006). https://doi.org/10.1191/1478088706qp063oa
7. Cribbs, J.D., Hazari, Z., Sonnert, G., Sadler, P.M.: Establishing an explanatory model for mathematics identity. Child Development **86**(4), 1048–1062 (2015). https://doi.org/https://doi.org/10.1111/cdev.12363
8. Crossley, S., Liu, R., McNamara, D.: Predicting math performance using natural language processing tools. In: Proceedings of the Seventh International Learning Analytics & Knowledge Conference. pp. 339–347. LAK '17, Association for Computing Machinery, New York, NY (Mar 2017). https://doi.org/10.1145/3027385.3027399
9. Doroudi, S., Brunskill, E.: Fairer but not fair enough: On the equitability of knowledge tracing. In: Proceedings of the 9th International Conference on Learning Analytics & Knowledge. pp. 335–339. ACM, Tempe, AZ (2019). https://doi.org/10.1145/3303772.3303838
10. Fancsali, S.E., Ritter, S., Yudelson, M., Sandbothe, M., Berman, S.R.: Implementation factors and outcomes for intelligent tutoring systems: A case study of time and efficiency with Cognitive Tutor Algebra. Proceedings of the Twenty-Ninth International Florida Artificial Intelligence Research Society Conference pp. 473–478 (2016), https://cdn.aaai.org/ocs/12965/12965-57684-1-PB.pdf
11. Flores, A.: Examining disparities in mathematics education: Achievement gap or opportunity gap? The High School Journal **91**(1), 29–42 (2007). https://doi.org/10.1353/hsj.2007.0022
12. Gordon, C., Gergen, K.J.: Self-conceptions: Configurations of content. In: The Self in Social Interaction. John Wiley & Sons, Inc., New York, NY (1968)
13. Guyan, K.: Queer Data: Using gender, sex and sexuality data for action. Bloomsbury Academic, London (2022)
14. Heffernan, N.T., Heffernan, C.L.: The ASSISTments ecosystem: Building a platform that brings scientists and teachers together for minimally invasive research on human learning and teaching. International Journal of Artificial Intelligence in Education **24**(4), 470–497 (2014). https://doi.org/10.1007/s40593-014-0024-x

15. Holstein, K., Doroudi, S.: Fairness and equity in learning analytics systems (FairLAK). In: Companion Proceedings of the Ninth International Learning Analytics & Knowledge Conference. p. 2. Society for Learning Analytics Research, Arizona (2019)
16. Jo, E.S., Gebru, T.: Lessons from archives: Strategies for collecting sociocultural data in machine learning. In: Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency. pp. 306–316. FAT* '20, Association for Computing Machinery, New York, NY (2020). https://doi.org/10.1145/3351095.3372829
17. Kayyali, R.: US Census classifications and Arab Americans: Contestations and definitions of identity markers. Journal of Ethnic and Migration Studies **39**(8), 1299–1318 (2013). https://doi.org/10.1080/1369183X.2013.778150
18. Le Quy, T., Roy, A., Iosifidis, V., Zhang, W., Ntoutsi, E.: A survey on datasets for fairness-aware machine learning. WIREs Data Mining and Knowledge Discovery **12**(3), e1452 (2022). https://doi.org/10.1002/widm.1452
19. Lu, C., Kay, J., McKee, K.R.: Subverting machines, fluctuating identities: Relearning human categorization. In: 2022 ACM Conference on Fairness, Accountability, and Transparency. pp. 1005–1015. FAccT '22, ACM, Seoul, Republic of Korea (2022). https://doi.org/10.1145/3531146.3533161
20. M. Kuhn, McPartland, T.: An empirical investigation of self-attitudes. American Sociological Review **19**, 68–76 (1954)
21. Osborne, J.W.: Testing stereotype threat: Does anxiety explain race and sex differences in achievement? Contemporary Educational Psychology **26**(3), 291–310 (2001). https://doi.org/10.1006/ceps.2000.1052
22. Ritter, S., Fancsali, S.E.: MATHia X: The next generation Cognitive Tutor. In: Proceedings of the EDM 2016 Workshops and Tutorials. pp. 624–625. International Educational Data Mining Society, Raleigh, North Carolina (2016)
23. Slater, S., Ocumpaugh, J., Baker, R., Li, J., Labrum, M.: Identifying changes in math identity through adaptive learning systems use. In: Proceedings of the 26th International Conference on Computers in Education. p. 7. Asia-Pacific Society for Computers in Education, Philippines (2018)
24. Sloan-Lynch, J., Morse, R.: Equity-forward learning analytics: Designing a dashboard to support marginalized student success. In: Proceedings of the 14th Learning Analytics & Knowledge Conference. pp. 1–11. LAK '24, Association for Computing Machinery, Kyoto, Japan (2024). https://doi.org/10.1145/3636555.3636844
25. Stobart, G.: Fairness in multicultural assessment systems. Assessment in Education: Principles, Policy & Practice **12**(3), 275–287 (2005). https://doi.org/10.1080/09695940500337249
26. Tywoniw, R., Crossley, S.A., Ocumpaugh, J., Karumbaiah, S., Baker, R.: Relationships between math performance and human judgments of motivational constructs in an online math tutoring system. In: Proceedings of Artificial Intelligence in Education: 21st International Conference, AIED 2020. pp. 329–333. Springer-Verlag, Ifrane, Morocco (2020). https://doi.org/10.1007/978-3-030-52240-7_60
27. Westbrook, L., Saperstein, A.: New categories are not enough: Rethinking the measurement of sex and gender in social surveys. Gender & Society **29**(4), 534–560 (Aug 2015). https://doi.org/10.1177/0891243215584758
28. Wise, A.F.: Designing pedagogical interventions to support student use of learning analytics. In: Proceedings of the Fourth International Conference on Learning Analytics & Knowledge. pp. 203–211. LAK '14, Association for Computing Machinery, Indianapolis, Indiana (Mar 2014). https://doi.org/10.1145/2567574.2567588