

Fair Design of Learners Descriptive Cards

Inaya El Alaoui¹, Céline Treuillier¹ and Anne Boyer¹

¹Université de Lorraine, CNRS, LORIA, Vandoeuvre-lès-Nancy, France

Abstract

The analysis and processing of learning traces through Learning Analytics, although essential to help improve learning, requires addressing strong ethical issues. Tools provided to various educational stakeholders must ensure fair learning, and take into account the diversity of students and their behavior while avoiding bias and discrimination. In that context, descriptive cards are an emerging tool allowing for a representative description of the diversity of behaviors adopted within a class. However, their design raises a lot of questions about equity and fairness. To contribute to the need for fair learning, we present a study conducted in collaboration with teachers and propose a new design of descriptive cards, limiting as much as possible the introduction of bias.

Keywords

Descriptive Cards, Trustworthy Learning Analytics, Ethical Learning, Learning Behavior

1. Introduction

Nowadays, it is essential to consider the ethical concerns raised by the processing of the large quantity of data available about learners. In that context, learning analytics tools must be directed towards Trustworthy Learning Analytics (TLA) that respect the criteria of Trustworthy Artificial Intelligence (TAI) defined by the European Commission¹. In that way, LA systems must respect, among others, the diversity criterion by avoiding algorithmic bias and discrimination, as well as integrating stakeholders during their conception [1].


Among the large variety of existing LA systems, we find descriptive cards [2], which allow visualizing learning behaviors recorded in the learning traces. They help to improve explainability by taking the form of fictitious students to whom pedagogical experts can refer during their pedagogical tasks. However, the learners' behavior representation raises questions about bias and equity, leading to our research question **(RQ): How to design learner descriptive cards meeting the need for diverse, non-discriminative, and fair learning analytics?**

To address this question, we conducted a user study with teachers to propose a new way to design descriptive cards, participating in the dissemination of fairer learning analytics.

 inaya.elalaoui@gmail.com (I. E. Alaoui); celina.treuillier@loria.fr (C. Treuillier); anne.boyer@loria.fr (A. Boyer)



© 2022 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

 CEUR Workshop Proceedings (CEUR-WS.org)

¹<https://ec.europa.eu/futurium/en/ai-alliance-consultation.1.html>

2. Related works

2.1. Personas - Descriptive Cards

Digital learning behavior can be reported using descriptive cards which are a tool based on the concept of persona, used for presenting data in a narrative and natural language form [3, 2]. Personas were defined in the context of computer software design, as archetypes of potential application users [4] and used to get as close as possible to users' needs by representing models of user behavior, goals, and motivations compiled in a single individual's fictional description. This concept has been adapted to the education field as "a narrative description of typical learners who can be identified by centroids of machine learning classification processes" [5].

Usually, learners are organized into sub-groups according to their final results: whether they passed or failed, withdrew from the course, or received a distinction. This final result-based classification is common in LA datasets, as OULAD [6]. The descriptive cards will therefore make it possible, and assist teachers, to identify learners in need of specific pedagogical monitoring, whatever their final result.

Nevertheless, the design of personas can introduce several types of biases. This includes human bias, with the introduction of stereotypes, whether cultural, gender, or age-related. As personas represent users' archetypes, individuals who embody personas are often represented and characterized by a photo, an age, and a gender in addition to the narrative description of their behavior. This can indicate ethnicity, socio-economic status, preferences, and sometimes education level which can clearly influence professionals in the design and understanding of the tool, because of the assumptions and stereotypes they may have [7, 8].

Stereotypes are a simplification of reality generalizing personality or behavioral features to a particular social group and are loaded with values, whether positive or negative. It is "a set of shared beliefs about personal characteristics, usually personality traits but also behaviors, of a group of people" [9]. There is extensive literature on the existence of social stereotypes in the educational field, for example concerning racial discrimination [10, 11, 12, 13], or gender discrimination [14, 15]. Obviously, these stereotypes can affect learners' performance [16, 17] because it may cause a feeling of being judged on the basis of a negative stereotype and might also be associated with pressure and anxiety. Considering that personas tend to reinforce cultural stereotypes [18], it is therefore necessary to find a methodology for designing descriptive cards while introducing as little bias as possible. One possible solution to reduce these biases is to build the personas with stakeholders such as the users they embody, or educational experts.

2.2. Trustworthy Guidelines

Ethical integrity concerns must be considered when processing data about learners in a LA context [19]. Some researchers recently promote the development of Trustworthy Learning Analytics (TLA) [1]. In that way, Ladjal *et al.* explain that "a crucial aspect of developing TLA is centered around the algorithms and methods that would allow us to measure and mitigate potential risks", and also argue that notions of privacy are closely related to the trust in algorithms and systems using learners data [20]. Altogether, these findings highlight the need for further research answering critical questions about fairness and student well-being [21].

3. User Study: How to avoid representation bias?

3.1. Methodology

We evaluated three different types of cards (See Fig. 1). The first one is the **avatar description card** (See Figure 1a) inspired by the one in Treuillier and Boyer [2]. To address our research question about stereotype bias, we represented learners from different ethnic backgrounds: North African, South American/Hispanic, Asian, African/Black Americans, and Caucasian, for both women and men. The second type of card is the **icons descriptive cards** (See Figure 1b), designed to counter the potential bias due to avatars. An icon reflecting the learner's result replaces the avatar, and a non-gendered first name is used. The third type of card is the **group descriptive cards** (See Figure 1c) which displays a group of avatars representing learners with the same learning behavior. All cards were colored with a neutral color, except for the icon cards with gender-neutral names, where colors could vary between pink and blue, to test whether it could influence the perception of gender.

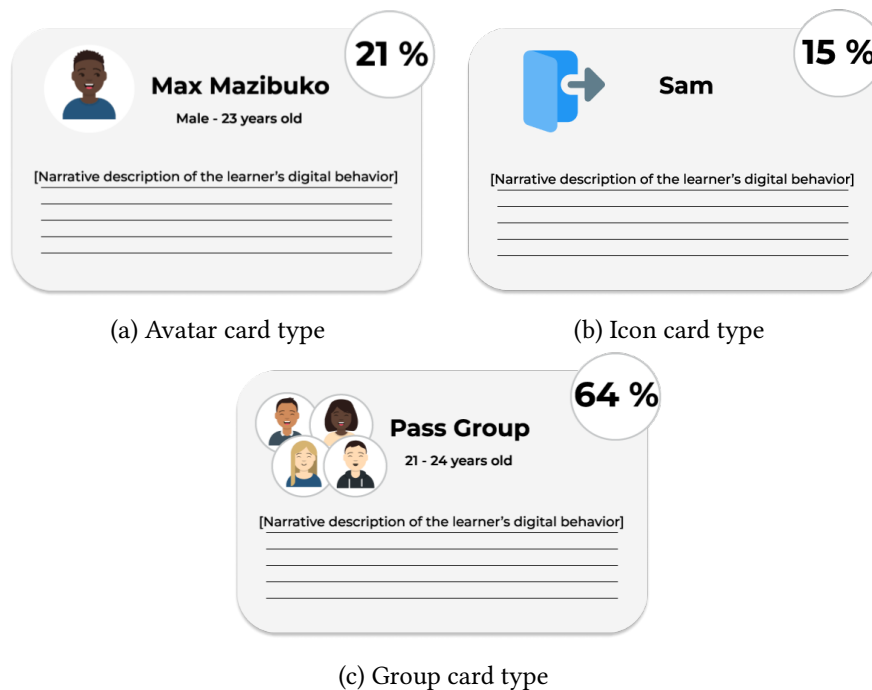


Figure 1: Evaluated Descriptive cards

Following our information campaign about our studies, 30 volunteer teachers were recruited. Precisely, we included twenty-four French higher education teachers and six French middle or high school teachers. In total there were 19 male participants and 11 females. As the recruitment of our participants mostly took place in a computer science department, 50% of the teachers in our sample taught computer sciences or related fields, 30% taught sciences-related fields such as Neuroscience, Physics, Chemistry, or Biology and we also recruited Languages and Art teachers which represent the last 20% of our sample.

The first objective is to investigate whether teachers are actually influenced by stereotypes, and which descriptive cards representation are least likely to induce them. To do that, participants were asked to autonomously carry out a computer-assisted categorization task inspired by the Implicit Association Test (IAT) [22], which is intended for measuring the stereotyping biases linked to the representation of learners on the descriptive cards, in an implicit way. Participants had to associate as quickly as possible attribute concepts that had either positive or negative connotations and that were located on either side of the screen with target concepts being descriptive cards displayed at the center of the screen. Participants had to respond, using a left or right key press when choosing an attribute concept between one of the indicators and their negative (e.g. Regular/ Irregular) (See Fig 2).

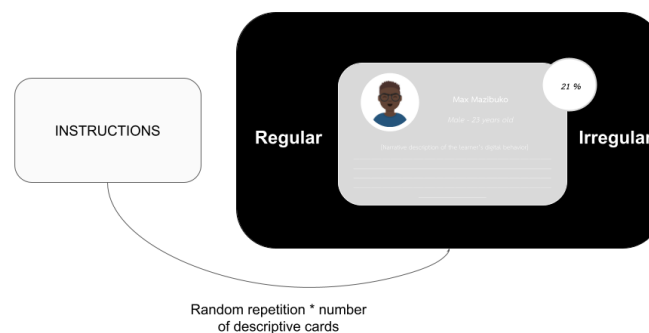


Figure 2: Implicit Association Test example

Next, the second objective is to assess the participants' perception about descriptive cards. Once the part on stereotyping bias was completed, participants were thus invited to fill out a questionnaire in order to evaluate how teachers receive and perceive the descriptive cards in a pedagogical context. This questionnaire was adapted from the Persona Perception Scale [23], and we evaluated five relevant factors: completeness, coherence, clarity, empathy and usefulness, and intention or willingness to use. Participants had to fill out this questionnaire regarding each type of descriptive card that was presented in front of them by indicating their level of agreement with the statements on a 7-point Likert scale ranging from "strongly disagree" (1) to "strongly agree" (7). Before ending the session, participants had to specify which type of card they preferred and why.

3.2. Results

3.2.1. Representation bias

Results related to the categorization task show that, when facing avatar descriptive cards (See Figure 1a), participants associate positive terms with all learners, regardless of gender or ethnicity. However, this association is more important for Caucasian and Asian avatars, while students from the North African, Black (especially the Black male student), and Hispanic minorities were more often linked to negative terms than their counterparts mentioned above

(only 75% to 90% of participants associate positive terms to those ethnicities). Concerning the icon descriptive cards (See Figure 1b) , results are coherent with most of the icons displayed on the descriptive cards. However, few participants expressed their reluctance to use several icons, which in their words *"does not seem clear"*, because they are not associated with the right concept. Finally, for the group descriptive cards (See Figure 1c), the names of the groups seem to be fairly explicit. Besides, about gender bias, which can be introduced by the names on the cards or by their colors, results indicate that some of the non-gendered names are very strongly associated with the male gender, regardless of the color of the card (for about 70% of participants). Regarding what guided their association choice, only 39% felt influenced by the cards' colors. Among this sample, half associated the color with a gender whereas the other half thought it was related to the learner's success when it was blue or failure when it was pink. Also, 10 participants said that they automatically associated failure with the male gender and success with the female gender. Finally, a large majority of the participants stated that they based their choice on their interpretation of the icons and the group names. They all added that this information was the most explicit, making the descriptive cards more understandable.

3.2.2. Persona Perception Scale

Concerning the persona perception scale results, we first normalized participants' scores for comparison. The higher the total score on the survey is, the more the user seems to like the tool. A mean of the scores obtained for each modality investigated was calculated, as well as the mean of the total scores obtained for the questionnaire, for each type of descriptive card. Scores obtained for each factor (completeness, coherence, clarity, empathy, usefulness) were high for all types of descriptive cards. However, the group card type (See Fig.1c) is the one with better scores of coherence, empathy, and intention to use, while the icon card type (See Fig. 1b) has better results for clarity and completeness scores. Finally, concerning the users' preferences for a type of card and the reasons for their choice, it appears that 10% of the 30 participants stated that they preferred the avatar cards (See Fig. 1a), against 66% for the group cards (See Fig.1c) and 24% for the icon cards (See Fig. 1b). We can therefore see that a majority of participants expressed a preference for group cards. Among this majority, 25% of them expressed the fact that the visual aspect of the icon cards was interesting to combine with the formulation and presentation of the group cards. In their opinion, this would help to identify learners needing special attention more quickly and easily.

4. Conclusion

The implementation of TLA requires joint actions of all the actors of LA. The protocol put in place made it possible to highlight key elements to decrease representation biases.

With this user study, we wanted to determine how descriptive cards could enhance pedagogical monitoring while respecting Trustworthy guidelines, especially by minimizing representation bias. Accordingly, we have highlighted stereotype bias that could emerge while using avatars, names, or colors and determine the teachers' preferred cards' features. In order to endorse the diversity and non-discrimination criterion of TAI as well as the transparency one,

we characterized how to design learner descriptive cards avoiding data representation bias, and thus allowing a fairer learners' behaviors characterization.

Taking into account participants' feedback, we then propose an updated version of descriptive cards, minimizing bias as much as possible. All presented results have been taken into account to propose this new version of descriptive cards. This user study carried out with the end-users of the studied LA tool enabled us to highlight some representation bias and stereotypes, and thus to propose a solution resulting from a joint reflection with educational stakeholders. This collaborative approach is essential to foster the development of fairer learning analytics tools.

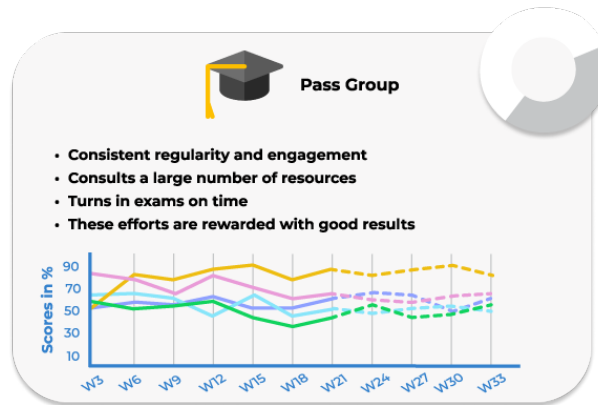


Figure 3: Enhanced version of descriptive card

References

- [1] H. Veljanova, C. Barreiros, N. Gosch, E. Staudegger, M. Ebner, S. Lindstaedt, Towards trustworthy learning analytics applications: An interdisciplinary approach using the example of learning diaries, in: HCI International 2022 Posters: 24th International Conference on Human-Computer Interaction, HCII 2022, Virtual Event, June 26–July 1, 2022, Proceedings, Part III, Springer, 2022, pp. 138–145.
- [2] C. Treuillier, A. Boyer, A new way to characterize learning datasets., in: CSEDU (2), 2022, pp. 35–44.
- [3] C. Treuillier, A. Boyer, Identification of class-representative learner personas, in: LA4SLE 2021-Learning Analytics for Smart Learning Environments, volume 3024, 2021, pp. 38–45.
- [4] A. Cooper, The inmates are running the asylum, Springer, 1999.
- [5] C. Brooks, J. Greer, Explaining predictive models to learning specialists using personas, in: Proceedings of the fourth international conference on learning analytics and knowledge, 2014, pp. 26–30.
- [6] J. Kuzilek, M. Hlosta, Z. Zdrahal, Open university learning analytics dataset, Scientific data 4 (2017) 1–8.
- [7] D. Spiliotopoulos, D. Margaritis, C. Vassilakis, Data-assisted persona construction using social media data, Big Data and Cognitive Computing 4 (2020) 21.

- [8] P. Turner, S. Turner, Is stereotyping inevitable when designing with personas?, *Design studies* 32 (2011) 30–44.
- [9] J.-P. Leyens, V. Yzerbyt, G. Schadron, *Stereotypes and social cognition.*, Sage Publications, Inc, 1994.
- [10] D. Katz, K. Braly, Racial stereotypes of one hundred college students., *The Journal of Abnormal and Social Psychology* 28 (1933) 280.
- [11] G. M. Gilbert, Stereotype persistence and change among college students., *The Journal of Abnormal and Social Psychology* 46 (1951) 245.
- [12] M. Karlins, T. L. Coffman, G. Walters, On the fading of social stereotypes: studies in three generations of college students., *Journal of personality and social psychology* 13 (1969) 1.
- [13] C. Chateignier, M. Dutrévis, A. Nugier, P. Chekroun, French-Arab students and verbal intellectual performance: Do they really suffer from a negative intellectual stereotype?, *European Journal of Psychology of Education* 24 (2009) 219. URL: <https://doi.org/10.1007/BF03173013>. doi:10.1007/BF03173013.
- [14] B. Dardenne, M. Dumont, T. Bollier, Insidious dangers of benevolent sexism: consequences for women's performance., *Journal of personality and social psychology* 93 (2007) 764.
- [15] A. Chatard, S. Guimond, L. Selimbegovic, "how good are you in math?" the effect of gender stereotypes on students' recollection of their school marks, *Journal of Experimental Social Psychology* 43 (2007) 1017–1024.
- [16] É. Sales-Wuillemin, *La catégorisation et les stéréotypes en psychologie sociale*, Dunod, 2006.
- [17] C. M. Steele, A threat in the air: How stereotypes shape intellectual identity and performance., *American psychologist* 52 (1997) 613.
- [18] D. G. Cabrero, H. Winschiers-Theophilus, J. Abdelnour-Nocera, A critique of personas as representations of "the other" in cross-cultural technology design, in: *Proceedings of the First African Conference on Human Computer Interaction*, 2016, pp. 149–154.
- [19] W. Holmes, K. Porayska-Pomsta, K. Holstein, E. Sutherland, T. Baker, S. B. Shum, O. C. Santos, M. T. Rodrigo, M. Cukurova, I. I. Bittencourt, et al., Ethics of ai in education: Towards a community-wide framework, *International Journal of Artificial Intelligence in Education* (2021) 1–23.
- [20] D. Ladjal, S. Joksimović, T. Rakotoarivelo, C. Zhan, *Technological frameworks on ethical and trustworthy learning analytics*, 2022.
- [21] E. Hakami, D. Hernández Leo, How are learning analytics considering the societal values of fairness, accountability, transparency and human well-being?: A literature review, Martínez-Monés A, Álvarez A, Caeiro-Rodríguez M, Dimitriadis Y, editors. *LASI-SPAIN 2020: Learning Analytics Summer Institute Spain 2020: Learning Analytics. Time for Adoption?*; 2020 Jun 15-16; Valladolid, Spain. Aachen: CEUR; 2020. p. 121-41 (2020).
- [22] A. G. Greenwald, D. E. McGhee, J. L. Schwartz, Measuring individual differences in implicit cognition: the implicit association test., *Journal of personality and social psychology* 74 (1998) 1464.
- [23] J. Salminen, H. Kwak, J. M. Santos, S.-G. Jung, J. An, B. J. Jansen, Persona perception scale: developing and validating an instrument for human-like representations of data, in: *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems*, 2018, pp. 1–6.