JOURNAL OF LEARNING ANALYTICS

SOLAR SOCIETY for LEARNING ANALYTICS RESEARCH

Volume 10(1), 71–84. https://doi.org/10.18608/jla.2023.7793

Learning Analytics in Physics Education: Equity-Focused Decision-Making Lacks Guidance!

Adrian Grimm[1], Anneke Steegh[2], Marcus Kubsch[3], Knut Neumann[4]

Abstract

Learning Analytics are an academic field with promising usage scenarios for many educational domains. At the same time, learning analytics come with threats such as the amplification of historically grown inequalities. A range of general guidelines for more equity-focused learning analytics have been proposed but fail to provide sufficiently clear guidance for practitioners. With this paper, we attempt to address this theory–practice gap through domain-specific (physics education) refinement of the general guidelines We propose a process as a starting point for this domain-specific refinement that can be applied to other domains as well. Our point of departure is a domain-specific analysis of historically grown inequalities in order to identify the most relevant diversity categories and evaluation criteria. Through two focal points for normative decision-making, namely equity and bias, we analyze two edge cases and highlight where domain-specific refinement of general guidance is necessary. Our synthesis reveals a necessity to work towards domain-specific standards and regulations for bias analyses and to develop counter-measures against (intersectional) discrimination. Ultimately, this should lead to a stronger equity-focused practice in future.

Notes for Practice

- What is already known: Learning analytics come with potentials like individualized learning as well as threats like explainability challenges. General guidance to navigate these potentials and threats has been developed through the concept of responsibility. However, this guidance has little impact on the day-to-day work of practitioners because of its generality.
- What this study adds: We propose a process for making the general principles domain-specific and thereby actionable. We show how domain-specific analyses of historically grown inequalities can inform the choice of relevant diversity categories and evaluation criteria, suggest focal points for normative decision-making, and make those focal points, diversity categories, and evaluation criteria tangible for the learning analytics community.
- Implications for practice: We should work towards domain-specific, context-sensitive standards and regulations for bias analyses along with developing counter-measures against (intersectional) discrimination.

Keywords

Equity, responsible learning analytics, principles, ethics, bias, critical consciousness, learning progression

Submitted: 30/06/2022 — **Accepted:** 09/01/2023 — **Published:** 18/02/2023

Corresponding author [1] Email: grimm@leibniz-ipn.de Address: IPN-Leibniz Institute for Science and Mathematics Education, Olshausenstrasse 62, D-24118 Kiel, Germany. ORCID ID: https://orcid.org/0000-0003-2701-3349

[2] Email: steegh@leibniz-ipn.de Address: IPN-Leibniz Institute for Science and Mathematics Education, Olshausenstrasse 62, D-24118 Kiel, Germany. ORCID ID: https://orcid.org/0000-0002-0988-5031

[3] Email: kubsch@leibniz-ipn.de Address: IPN-Leibniz Institute for Science and Mathematics Education, Olshausenstrasse 62, D-24118 Kiel, Germany. ORCID ID: https://orcid.org/0000-0001-5497-8336

[4] Email: neumann@leibniz-ipin.de Address: IPN-Leibniz Institute for Science and Mathematics Education, Olshausenstrasse 62, D-24118 Kiel, Germany. ORCID ID: https://orcid.org/0000-0002-4391-7308

1. Introduction and Background

Learning Analytics are an academic field with promising usage scenarios for many educational domains, including the "provision of personalised and timely feedback to students regarding their learning" (SoLAR, 2022). At the same time, learning analytics come with threats such as the amplification of historically grown inequalities (Cheuk, 2021; D'Ignazio & Klein, 2020; Erden, 2020). To address these threats, general guidelines were developed and thoroughly reviewed by the learning analytics community (Cerratto Pargman & McGrath, 2021; Prinsloo & Slade, 2018; Sclater, 2014). However, these general guidelines have been found to have little impact on the day-to-day work of learning analytics practitioners in many domains because their generality makes concrete applications difficult (Kitto & Knight, 2019, pp. 2861–2864). To increase impact and usability, the development of domain-specific guidelines was proposed (Kitto & Knight, 2019, p. 2859). Domain-specific guidelines seem particularly appropriate when addressing the threat of amplifying historically grown inequalities since their nature is domain-dependent. For example, inequalities based on gender (binary gendered in the sources) are present in both secondary school reading as well as science, technology, engineering, and mathematics (STEM). However, female students outperform male students in reading (OECD, 2018), whereas male students are more likely to have STEM career aspirations than female students (OECD, 2016).

As STEM education researchers working on STEM identity development, we wanted to harness the potentials of learning analytics while considering domain-specific threats. We asked ourselves, how do we prevent or counter the threats based on the general guidelines in a concrete physics education project in Germany? Although algorithms making bias visible or even preventing bias have been proposed (Baker & Hawn, 2021; Mitchell et al., 2021; Suresh & Guttag, 2021; Traag & Waltman, 2022), we realized that general guidelines leave many questions unanswered in terms of decision-making when trying to apply them in the concrete context of a domain. In this paper, we therefore describe the steps we took in order to specify where and in what way general guidelines lack practical application to address the threat of amplifying historically grown inequalities in our domain. Our process can serve as a starting point in other domains for translating and further developing the general guidelines into actionable, domain-specific directions. We do not provide new thresholds for particular parameters of learning analytics algorithms applied in physics education in Germany. Instead, we aim to contribute to the field by specifying our process of concretizing the general guidelines into practical directions as a domain-specific example. The proposed process can inform the development of actionable guidelines in other domains, thereby leading to increased impact of the guidelines in addressing the threats.

We started our process by explicitly reflecting our understanding of responsible learning analytics with a focus on equity issues and historically grown inequalities by naming both potentials and threats. With specific equity-related threats in mind, we then reviewed existing general guidelines that fit this issue. In terms of equity issues, these guidelines leave some central questions unanswered: Which diversity categories are most relevant regarding the historically grown inequalities in our domain? Which constructs are most meaningful as evaluation criteria in our domain? To answer these questions, we started by describing the historically grown inequalities in our domain and context: physics education in Germany. Once we identified diversity categories and evaluation criteria, normative decisions on how these categories and criteria are to be considered had to be made. In order to approach the fuzzy decision-making space, we chose *equity* and *bias* as the two focal points for our study. These are inspired by unfolding potentials on the one side and inhibiting threats on the other. Equipped with the general guidelines, the domain-specific diversity categories, the evaluation criteria, and the two focal points in terms of normative decision-making, we applied a method proposed by Kitto and Knight (2019), who suggested reporting on edge cases1 to identify missing domain-specific guidance. The research question that guides our edge case analysis is this: Which tensions and edge cases regarding bias and equity emerge when designing a learning analytics system in physics education using the existing guidance for practice?

Before moving on, we provide the context for the LPA-AFLEK project necessary to identify relevant diversity categories and evaluation criteria and position ourselves as authors aligned with feminist standpoint theory. In our LPA-AFLEK project, we address the finer descriptions of individual secondary school students' learning trajectories on their way through "learning progressions" (Duncan & Rivet, 2018) in understanding the concept of energy. We profile students in competence models through automated labelling of student answers — for example, free text and multiple choice — in digital learning environments. The labels are more specific than simply "correct" or "wrong"; they indicate, for example, whether a particular knowledge element is used in an answer. The automated labelling is done with algorithms trained on student answers previously labelled by researchers. These automated labels are displayed to teachers through a dashboard to support them in timing and individualizing interventions in real-time classroom situations.

According to Costanza-Chock (2020), "Feminist standpoint theory recognizes that all knowledge is situated in the particular embodied experiences of the knower" (p. 9). We therefore want to position ourselves at the beginning of this research article. We are researchers from Europe — Germany and the Netherlands — conducting our research in northern Germany. All of us have cis-gender identities; three of us identify as men, one as a woman. All of us identify as *white.*[2]

[1] Learning analytics system builders can express tensions they confront in their work through concrete cases (or *edge cases*; Kitto & Knight, 2019). A tension is a conflict between two or more principles that cannot be fully accomplished at the same time, thus requiring a decision in terms of priority.

[2] We set *white* in italics to emphasize it as a privileged position in the structure of racism rather than a skin colour, as was proposed by Black German author Tupoka Ogette (2019) in her book *Exit Racism* (p. 14).

2. Theory

2.1. Responsible Learning Analytics

While learning analytics have great potential for many educational domains, they also bring threats that cannot be ignored. These threats are summarized as *eroding* use: the over-, mis-, and under-use of learning analytics (Floridi et al., 2018, p. 690). While under-use describes the failure to fully utilize the potential of learning analytics, over- and misuse can lead to undesirable outcomes (Kitto & Knight, 2019). The tension between potentials and threats is addressed within the field of responsible learning analytics by a combination of rules and principles in practice with value- and concern-driven approaches (Cerratto Pargman et al., 2021, p. 2). Possible potentials are addressed by a principle called the *obligation to act* — the responsibility and commitment to make use of the potentials of learning analytics. Possible threats are addressed by the principle of *accountability* (Prinsloo & Slade, 2018, p. 3). Responsible learning analytics help to guide practitioners in analyzing potentials and threats by helping them work according to the obligation to act while also accounting for threats and navigating the tensions between the two.

In our understanding, responsible learning analytics are also rooted in *critical theory* that explicitly addresses "power relations" and the "relationships between culture, forms of domination, and society" (Prinsloo & Slade, 2018, p. 4). This aligns well with the *subversive stance* on learning analytics as proposed by Wise et al. (2021) as a way of engaging with issues of power and equity and their interaction with data on learning processes. Being rooted in critical theory and the subversive stance on learning analytics on the one hand, as well as navigating the tensions between the obligation to act and accountability on the other, responsible learning analytics open up a fuzzy, complex space for decision-making in practice. With our paper, we aim at making this fuzzy space clearer and more actionable in practice.

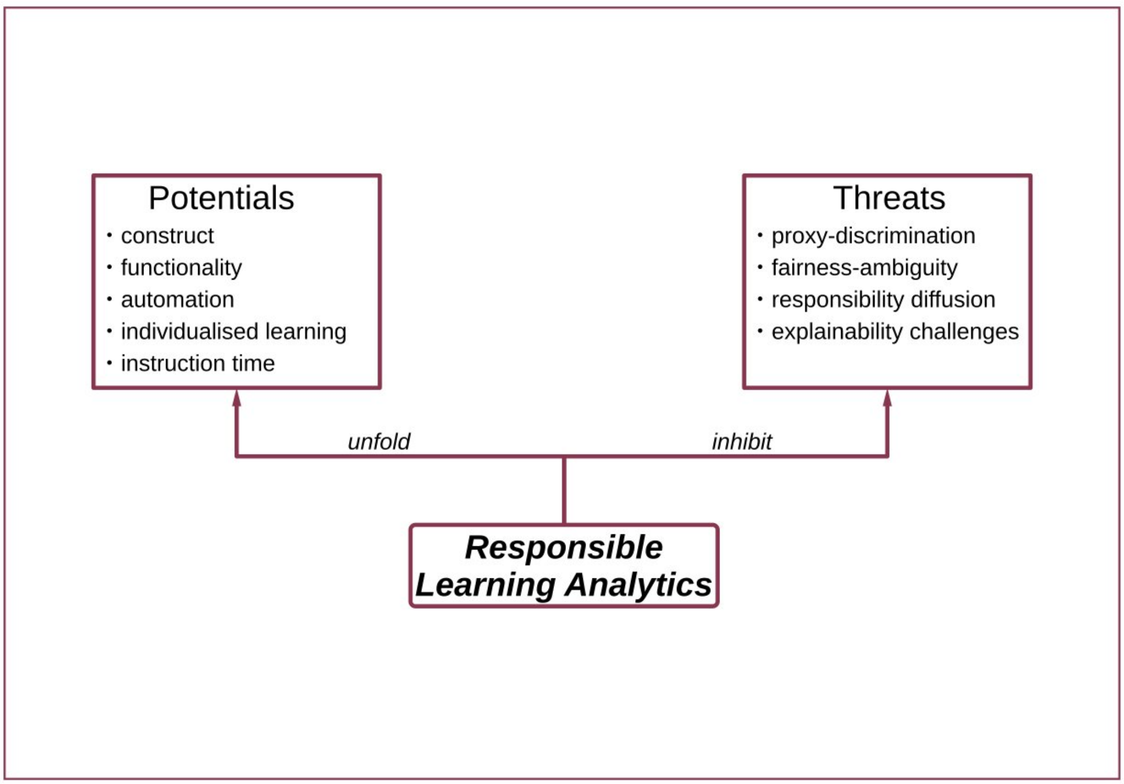

Figure 1. Responsible learning analytics — practice between potentials and threats

The potentials of learning analytics are summarized on the left side of Figure 1. According to Zhai et al. (2019), through learning analytics, we can assess *constructs* with more "complexity, diversity, and structure" (p. 1442). For complex constructs such as understanding energy as a core concept of physics, an assessment with more diverse "cognitive demands" holds great potential (p. 1442). Unpacking a construct in its structure along "three-dimensional learning" and assessing it in all these dimensions can lead to a better understanding of complex constructs (p. 1444). At the same time, increased *functionality* refers to better results in terms of, for example, assessment results enabling more valid representations of students' actual competences whereas *automation* comes with the promise to "save human effort" (p. 1445). These three factors combined hold the potential to increase *individualized learning* through, for example, automated feedback while simultaneously taking over tasks, thus increasing teachers' *instruction time*.

The threats of learning analytics are represented on the right side of Figure 1. According to Erden (2020), data-trained algorithms come with the threat of *proxy-discrimination* (p. 85). This means that even if training is done with a quantifiable goal criterion and no protected category variables, algorithms can still be quite discriminatory if a "proxy" variable is included that correlates with both the quantifiable goal criterion and one or more protected category variables. As an example from the US, Erden gives an algorithm that predicts the duration of staff tenure by travelling distance to their job. Since US postal code is strongly correlated with race, the algorithm turns out to be quite racist through proxy-discrimination. In other words, learning analytics cannot be taken as discrimination-free based simply on the fact that the algorithms do not use discrimination categories.

When and for whom algorithms, such as classifiers, are fair is subject to debate. There are pre-processing, post-hoc procedures, and direct methods to train fair algorithms (Lohaus et al., 2020, p. 6360). Nonetheless, some algorithms can turn out to be surprisingly unfair when applied to different contexts — they are "too relaxed to be fair" as Lohaus et al. put it (2020). This means that an algorithm can be trained to produce unfair results if this leads to a better average accuracy in prediction. Whether this is acceptable or not within a specific context is highly normative. We subsume this phenomenon under *fairness-ambiguity*; an insecurity in the current principles due to a lack of normative specifications for algorithm designers.

Responsibility in terms of accountability is confronted with new issues when it comes to algorithmic decision-making, especially the "'many-hands'-problem" and the fact that humans interact with computers. Since algorithms also interact with each other, this becomes a huge issue in terms of allocating accountability (Yeung, 2019). Not allocating accountability can also be thought of as *responsibility diffusion* from a victim's perspective. Since there are different responsibility models available, deciding on one of them in a particular case is a normative decision "between our interest, as agents, in freedom of action and our interest, as victims, in rights and interests in security of person and property" (Yeung, 2019, p. 11).

Several major algorithms with, for example, neural-network-based architectures additionally come with *explainability challenges*. An algorithm is explainable if its outputs come with an explanation for the decision, if this explanation is meaningful and accurate, and if the algorithm operates within its knowledge limits (Phillips et al., 2020). Providing an explanation is not possible for all algorithms and existing explanation methods have already been shown to be vulnerable to adversarial attacks, for example (Slack et al., 2020). Additionally, "explanatory power and predictive power do not always point in the same direction" (Bergner, 2017, p. 42). This means that in the testing phases, algorithms that are not explainable can yield better predictive performance. A decision needs to be made about what is more important in a normative and context-sensitive question. Having proxy-discrimination, responsibility diffusion, and the obligation to act in mind, this is not a straightforward decision for algorithm designers.

2.2. Guiding Principles for Practice

A range of sources on addressing equity issues are available, including literature reviews on codes for practice (Cerratto Pargman & McGrath, 2021; Khalil et al., 2022; Sclater, 2014), checklists (Drachsler & Greller, 2016), expert reports (DEK, 2019), laws like the European General Data Protection Regulation, an anti-sexism-law in the German federal state of Schleswig-Holstein ("Gesetz zur Gleichstellung der Frauen") or UN-documents against racism ("International Convention on the Elimination of All Forms of Racial Discrimination"), policies (Slade, 2016) and principles (Floridi et al., 2018; Phillips et al., 2020). These sources range from very general policies outlining a general ethic (for example the UN-documents) to very concrete principles intended to guide practice (for example the checklists provided by Drachsler and Greller). Since the concrete principles provide the most guidance on how to handle equity issues in practice, we specifically focus on three of them in our practical report. The first set of principles are the seven principles of data feminism, which ask us to "examine [… and] challenge power" (D'Ignazio & Klein, 2020). Particularly regarding gender, feminist pedagogy of data science calls to "challenge the gender binary, along with other systems of counting and classification that perpetuate oppression" (D'Ignazio & Klein, 2020).

Second, according to Costanza-Chock (2020), Popular Education principles specify that "Education is never neutral: it either maintains the current system of domination, or it is designed to liberate people" (p. 177). The third and last set of Design Justice Network principles define that "[w]e center the voices of those who are directly impacted by the outcomes of the design process" (pp. 6–7). Additionally, processes need to have assigned accountabilities: "We view change as emergent from an accountable, accessible, and collaborative process, rather than as a point at the end of a process" (pp. 6–7). Costanze-Chock specifies this through the example of so-called universal design: "Universalization erases difference and produces self-reinforcing spirals of exclusion, but personalized and culturally adaptive systems too often are deployed in ways that reinforce surveillance capitalism. Design justice doesn't propose a 'solution' to this paradox. Instead, it urges us to recognize that we constantly make intentional decisions about which users we choose to center and holds us accountable for those choices"

(p. 56). The Design Justice Approach also asks to focus, "Explicitly on the ways that design reproduces and/or challenges the matrix of domination (white supremacy, heteropatriarchy, capitalism, ableism, settler colonialism, and other forms of structural inequality)" (pp. 6–7). Costanza-Chock specifies this with the example of A/B-testing: "A/B testing is widely seen as leading inexorably to 'better UX' and 'better UI.' But a question must be asked: Better for whom? […] we should critique (trouble, queer, de-normalize) the assumption that A/B testing is always geared toward improving UX, for the simple reason that it is actually geared toward increasing the decision-making power of the product designer. […] we might destabilize the underlying assumption that what is best for the majority of users is best for all users" (p. 57).

However, principles so far are rarely applied in many domains as principles and tensions open up a fuzzy space that is neither specified enough nor enforceable in practice (Kitto & Knight, 2019, pp. 2861–2864). For example, for bias analyses, various definitions exist and it is not a straight forward decision which definition to use in a particular context. Traag and Waltman (2022) define bias as "direct causal effect" and thereby exclude correlations from their definition (p. 1). Suresh and Guttag (2021) offer a bias definition focused on historically grown inequalities. Mitchell et al. (2021) even propose an explicitly non-statistical bias definition. Apart from the question of which bias definition to take, the question of for which diversity categories to analyze for remains unanswered. Which categories are most relevant, and the direction of each category, is domain-specific as we already have shown in the example for the diversity category gender with reading versus STEM career aspirations. Defining the most relevant categories for a field as well as choosing, for example, a bias definition are huge tasks, especially from a practitioners' perspective. In the following, we offer an analysis of well described historically grown inequalities in physics education in Germany and two focal points for normative decision-making, equity and bias. The inequalities we describe are not complete and the focal points could be chosen differently. However, we will create explicit descriptions and definitions in order to make both inequalities and focal points visible and thereby debatable. In our results, we will show how the analysis of historically grown inequalities in our specific domain and the two focal points influence decision-making. Thereby, we aim at showing why domain-specific analyses of inequalities as well as explicit focal points as guidance are missing for practitioners.

2.3. Historically Grown Inequalities in Physics Education in Germany

There is a broad range of threats when applying learning analytics in physics education. In Figure 2, we show the already discussed general threats in learning analytics on the left side. On the right side, the threats in physics education are added. We start by describing the historical inequalities in physics education for sexism, racism, classism, and intersectional discrimination. The intersection of threats marks the specific problems of using learning analytics for physics education and the area that lacks guidance.

Multiple prior studies have shown that *sexism* and gender inequality play a major role in physics education (for example Avraamidou, 2019; Steegh et al., 2019). In the European Union in 2018, only 28% of graduates in engineering, manufacturing, and construction were women (EIGE, 2022, p. 21). In Germany in 2020, 21% of physics bachelor and 18% of physics master program graduates were women (Düchs & Ingold, 2018, p. 36). Moreover, girls are less likely to be encouraged by their teachers or to have positive experiences in their physics classes than boys (Mujtaba & Reiss, 2013, p. 1824). From a gender perspective, questions of recognition, STEM identity (Carlone & Johnson, 2007; Godwin, 2016) and career aspirations (Dou et al., 2019) are relevant constructs and evaluation criteria when analyzing differences among students rather than questions of competence and achievement (OECD, 2016).

The Neue deutsche Medienmacher (2021) define *racism* including structural components as follows: "Racism happens when structurally disadvantaged groups or individuals are excluded and depreciated because of actual or supposed physical or cultural attributes (like skin colour, origin, language, religion)."[3] Tupoka Ogette (2019, p. 57) defines institutional and structural racism in her book exit racism by quoting the Federal Agency for Civic Education's historical perspective on racism (Odoi, 2021): "Institutional racism is defined as racism anchored in the structures of public and private institutions. These structures have developed due to states of historical and societal power and violence and have become manifest in the economic as well as cultural and political architecture of a society and its institutions. Invisible in their essence, these structures consciously and unconsciously influence the behaviour, points of view, and ways of thinking of the individuals in the institutions. Conversely, the individuals determine the behaviour of the institutions in which they work."[4]

The Afrozensus reveals that most Black persons in German education systems reported discrimination for racist reasons connected to ethnic origin (88.5%) and skin colour (79.8%). In the context of discrimination, these attributes were much more important than gender (34.5%) and social status or social origin (24.0%; Aikins et al., 2021, p. 170). Moreover, racism has been found to limit the development of science identity (Avraamidou, 2019), algorithmic decision-making (Cheuk, 2021, p. 3),

[3] The quote was translated from German into English by the authors.

[4] The quote was translated from German into English by the authors.

sense of belonging (Rainey et al., 2018), and educational pathways in physics (Rosa & Moore Mensah, 2016). Addressing racism in the German education system is a crucial equity issue.

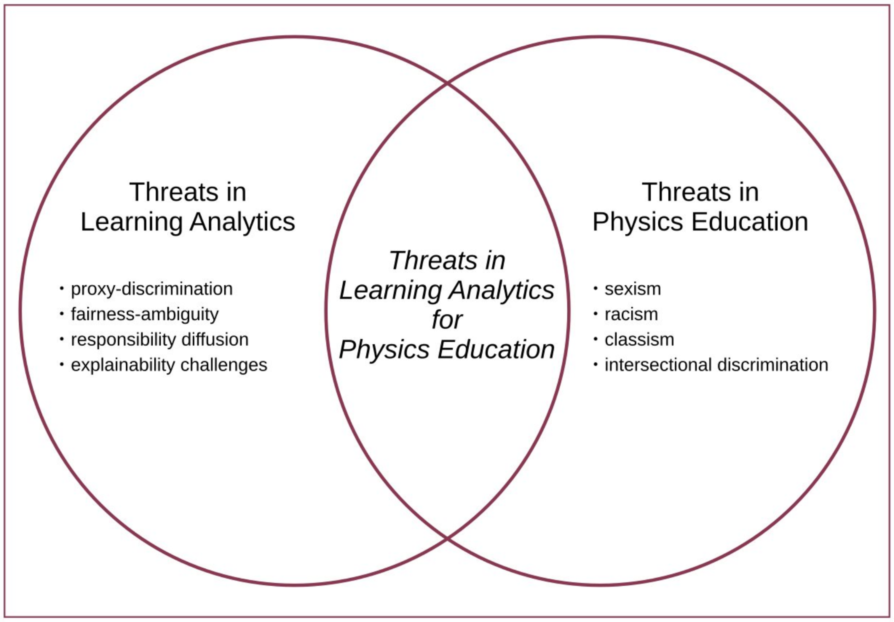

Figure 2. Threats in learning analytics for physics education

Classism is also relevant for physics education as part of STEM as well as the local context of Germany. Students' science aspirations and science, technology, engineering, and mathematics career choices have been shown to be directly related to their socio-economic classification (Avraamidou, 2019, p. 318). In Germany, social class is the strongest predictor for starting an academic career: 79 out of 100 children of academic households start academic studies, compared to only 27 out of 100 children of non-academics and twelve out of 100 children of parents without a professional qualification (Kracke et al., 2018, pp. 5–6, following El-Mafaalani, 2021, pp. 66–67). A view on *intersectional discrimination* reveals distinct, new forms of discrimination that would remain invisible if each category was analyzed only by itself. As Costanza-Chock (2020) puts it, "Black feminist thought fundamentally reconceptualizes race, class, and gender as interlocking systems: they do not only operate on their own, but are often experienced together by individuals who exist at their intersections" (p. 17).

For physics education, we have shown that the diversity categories of gender, race, class, and intersectional discrimination are relevant. As evaluation criteria, competence development and grades are not the most reported constructs; instead, STEM identity development or career aspirations are used to report inequalities. When using learning analytics in a specific domain, addressing the particular historically grown inequalities of that domain is needed. For general guidance to be translated into action, a decision on the most relevant diversity categories in the specific domain needs to be made. This decision can be based on an analysis of historically grown inequalities. However, the decision about which categories to focus on is a normative question and difficult to answer for practitioners. Here, practitioners need guidance in order to be able to translate general guidance into action. In the case of learning analytics for physics education, equity issues exist due to historically grown inequalities and discrimination phenomena in both learning analytics and physics education. Equipped with an analysis of the historically grown inequalities in physics education as well as the threats in learning analytics, we now explicate two focal points for normative decision-making upon which we base our two edge cases in our concrete project context: equity and bias.

2.4. Two Focal Points for Normative Decision Making: Equity and Bias

The aim of the LPA-AFLEK project is to automatically generate labels for student answers through data-driven algorithms. However, reproducing existing inequalities is a highly relevant threat in this context since both physics education and learning analytics face existing inequalities. At the same time, Costanza-Chock (2020) argues that "[for example racial] hierarchies can only be dismantled by actively antiracist systems design, not by pretending they don't exist" (p. 62). The concept of equity allows us to look at how to unfold the potentials of responsible learning analytics in terms of reversing historically grown inequalities instead of simply avoiding or even amplifying them. Since bias is a well-documented threat within the field of

ISSN 1929-7750 (online).

learning analytics, equity connects well to the existing work in the learning analytics community, and both focal points connect well to the potentials and threats within responsible learning analytics.

A system is biased when it has “undesirable […] behaviors or properties” (Cheuk, 2021, p. 2). “Undesirable” is particularly important here. We want to avoid bias, understood as a preference for an already privileged group, such as an algorithmic preference for men. A negatively biased treatment of men may therefore serve as a justifiable anti-discrimination feature for gender equity according to our definition. In the case of LPA-AFLEK, a bias in the selected algorithms would lead to incorrect labels for student competence. These wrong labels would then be displayed to teacher dashboards and guide interventions in biased directions. This could, for example, result in unwarranted negative feedback from teachers to non-male students, possibly leading to a lack of recognition and a weaker science identity in individual students, and to the reproduction of existing sexist structures in physics education as a whole.

According to Costanza-Chock (2020), not reproducing historically grown inequalities is not enough to achieve equity (p. 62). For equity, it is necessary to also focus on the “differences [the students] brought with them due to the effects of past discrimination or even discrimination in other venues” (p. 62). This “requires redistributive action” and “means that the algorithm designers must discuss, debate, and decide upon what they believe to be a just distribution of outcomes” (pp. 63–64). When existing inequalities and discrimination are not actively addressed by counter-measures, they will persist or be amplified by learning analytics.

3. Methods

We aimed at specifying the missing guidance of existing guiding principles for practical application by asking which tensions and edge cases regarding equity and bias emerge when designing a learning analytics system in physics education. In order to meet this aim, we analyzed project-specific tensions and edge cases in accordance with the structure developed by Kitto and Knight (2019, p. 2867): 1) description of the problem, 2) relevant principles, 3) how we handled the tension in LPA-AFLEK, 4) difficulties, and 5) which guidance is missing? We found the structure developed by Kitto and Knight particularly suitable since it stems from the learning analytics community as well as the authors’ demand for domain-specific edge cases — which is our application. To choose the principles we referred to in our edge cases, we focused on sexism, racism, classism, and their intersections in physics education.

To account for both equity and bias, and thereby potentials as well as threats, we analyzed two edge cases that helped to define the field of responsible learning analytics more precisely. In the edge case on equity, we focused on the part of responsible learning analytics that calls for cultural change, augmented by diversity as a core value, and obtainable through building critical consciousness. The edge case on bias further developed the equity issues that can be addressed by, for example, standards and checklists. For the analyses of the edge cases, no code book nor structured interviews were used. Instead, all authors discussed where exactly missing guidance identified in the theory sections can be explicated in our concrete project context. In both edge cases, we showed where existing principles fell short in providing clear guidance.

4. Edge Cases

4.1. Edge Case 1: How much effort is enough? Need for intersectional bias analyses versus obligation to act

4.1.1. Description of the problem

In using learning analytics in physics education, there is a particular threat for (intersectional) biases in our algorithms, for example through proxy discrimination. Bias analyses need to address explainability challenges as well as fairness-ambiguity. Particularly in physics education, existing inequalities due to sexism, racism, classism, and intersectional discrimination are a good starting point for bias analyses. In Figure 3, an exemplary bias analysis for three algorithms that score student artefacts in LPA-AFLEK is shown. The questions we faced in our project are these: Which analyses do we perform? What does “fair” mean in the context of learning in physics education? Do we need to perform context-specific analyses ourselves or are references to existing analyses and choices for algorithms based on them sufficient? On the one hand, we saw a need for bias analyses due to the threats we faced, while the existing principles did not provide us with a clear checklist of how to screen our algorithms. On the other hand, finding a strategy for bias analyses could conflict with making use of the potentials of learning analytics: How much of our efforts are necessary in order to not have biased algorithms, and when should we focus instead on the potentials and our obligation to act to avoid under-use?

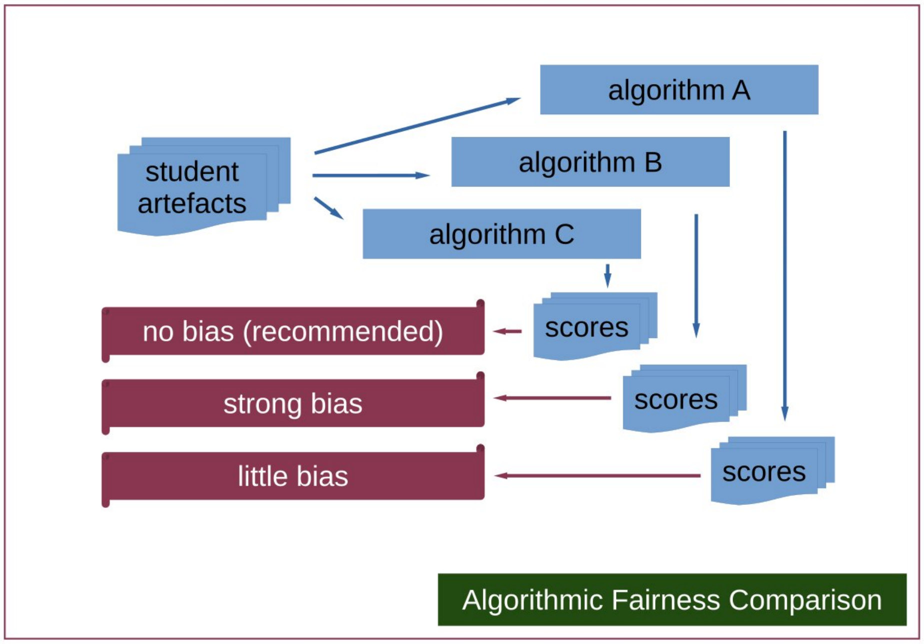

Figure 3. Bias analyses

4.1.2. Relevant principles

Both data feminism and Popular Education challenge existing power structures, such as explicitly naming and critically analyzing sexism, racism, and classism in physics education. Moreover, data feminism specifically asks us to challenge binary gender models. Design Justice requires us to centre those directly impacted, in our case students facing discrimination, and make these decisions explicit. If A/B-testing, for example, is used to judge the performance of our digital learning environment, we should ask for whom we perform the performance test and whose perspectives are overlooked or outweighed by naturally occurring bigger groups in our testing samples.

4.1.3. How we handled the tension in LPA-AFLEK

We implemented this project with a team of three professors, three postdocs, five doctoral students, and some student assistants. In German school settings, unlike with privacy guidelines, no explicitly allocated legal requirements exist for bias analyses in algorithm usage. In previous projects on which LPA-AFLEK was based, information on gender, race, and class was not assessed because of data minimization. This made sense when no bias analyses were performed, but made it more challenging to implement future analyses, since less data was available for algorithm development. Within LPA-AFLEK, we planned to perform a bias analysis regarding sexism and classism with the project's reduced data set.

4.1.4. Difficulties

Difficulties arose when we tried to concretely define a bias analysis in LPA-AFLEK with the principles for practice in mind. We explicitly analyzed power structures. It was challenging to choose which power structures to consider — and how. In our case, we focused on the power structures manifest in historically grown inequalities in physics education. Tangible assessment and operationalization were challenges since no standardized assessment survey for physics education was available. Regarding gender, we discussed our survey with anti-discrimination experts, and for class, we had an internal project group discussion. We decided to not assess race and to focus on class and gender instead. We decided to do so since we identified a lack of assessment opportunities for our context and no necessity to assess all the power structures relevant in physics education in order to make our argument.[5]

We placed students with the following identity markers at the centre of our analyses: 1) none of the legal guardians holds an academic degree and 2) a non-German language is the most spoken at home. We chose to do so since academic background

[5] In Germany, ethnicity is not commonly surveyed. Race would usually be self-declared, as done by Aikins et al. (2021). We generally considered that students had not self-reflected enough for a meaningful self-declaration, making the assessment questionable in terms of validity. For the sake of our argument, we aimed at investigating biases particularly relevant in the domain under investigation (i.e., physics education). Furthermore, we aimed to point out the difficulties in choosing the categories to investigate in first place, as well as the difficulties in translating into practice the commonly surveyed constructs.

of legal guardians could have an impact on the academic language usage of the students, which was what we based our scores on. Non-German languages spoken at home could impact the words used by students, which could lead to truncated scoring of the student artefacts due to missed lingual clues. Judging the results in terms of equity was difficult. We could, for example, quantify precision for our algorithms on the basis of gender-specific groups. We could quantify how results changed after we trained the algorithm with students from one gender group only or after applying slicing analyses as proposed by Gardner et al. (2019).

Judging the fairness was even more difficult: Does our algorithm need to score students from all gender groups equally well? What does "equally well" mean in our context? Would it be enough if our algorithms scored students from all gender groups at least with greater precision rather than a certain threshold (for example, a precision of 90%)? Is it enough to show that, on average, students of all gender groups learn the energy concept better via our algorithms, even if male students profit and female students are more often misclassified? The answers to these ethical questions cannot be defined scientifically but should be defined politically. In order to perform meaningful scientific analyses with practical relevance on a bigger scale, guidelines are needed. We decided to describe our results from different, explicit normative viewpoints and not to judge the fairness from an authors' perspective in a concluding statement. In practice, this means that even though we performed bias analyses we could not tag our algorithms as "fair" or "bias-free." This is problematic for a design practice that needs to justify the effort needed for bias analyses within a resource context with an obligation to act.

4.1.5. Which guidance is missing?

From our perspective, as long as no standards for bias analyses exist, they will not become part of practice in designing learning analytics algorithms in physics education since the effort and work do not result in rewards and recognition. To make sure that these analyses are performed — such as existing privacy guidelines in Germany — minimum requirements for algorithms can be implemented. Such standards are strongly normative and thus should be defined politically by parliaments, funding agencies, or the learning analytics community. Defining domain-specific standards makes sense in order to address only the domain-relevant threats since imposing too much effort on practitioners leads to the under-use of learning analytics. Specific standards can also increase comparability. Without standards, different operationalizations of identity markers for gender, race, and class can create difficulties in terms of comparing results. Until standards exist, researchers can analyze biases from different possible normative points of view and thus provide decision makers with edge cases, examples, and descriptions of practical consequences of particular normative decisions.

4.2. Edge Case 2: Bias-free is not enough! Need for counter-measures versus obligation to act

4.2.1. Description of the problem

In physics education, sexism, racism, classism, and intersectional discrimination and the related historically grown inequalities are well known. We asked how we could go beyond fair (in terms of bias-free) algorithms and have a praxis of equitable learning analytics with counter-measures, as shown in Figure 4. The questions we faced in our project were these: How do we design counter-measures for those most impacted? Where do students and teachers stand in terms of their critical consciousness[6] about existing inequalities? On the one hand, we saw a need for counter-measures due to the threats and inequalities, but the existing principles did not provide us with clear guidance. For example, how can we create cultural change for understanding that diversity is valuable and how can we develop critical consciousness? On the other hand, defining concrete counter-measures created tensions with the potentials of learning analytics: How much effort is necessary in order to create equitable learning analytics? and When should we focus instead on the potentials and our obligation to act to avoid under-use?

4.2.2. Relevant principles

Cerratto Pargman and McGrath (2021) suggest that equity-focused research and further investigation of "enabling interventions triggered by analytics" are needed. Data feminism explicitly challenges power. Design Justice holds us accountable for our "intentional decisions about which users we choose to center" and asks us to reflect explicitly on how our designs do or do not counter or reproduce discrimination (Costanza-Chock, 2020, pp. 56, 6–7).

[6] Critical consciousness is the "reflection and action upon the world in order to transform it" (Freire, 1970, p. 51). For teachers, it is "about students' identities that constitute 'diversity' and how they are situated within systems of oppression and privilege" (Baggett, 2020, p. 34).

ISSN 1929-7750 (online).

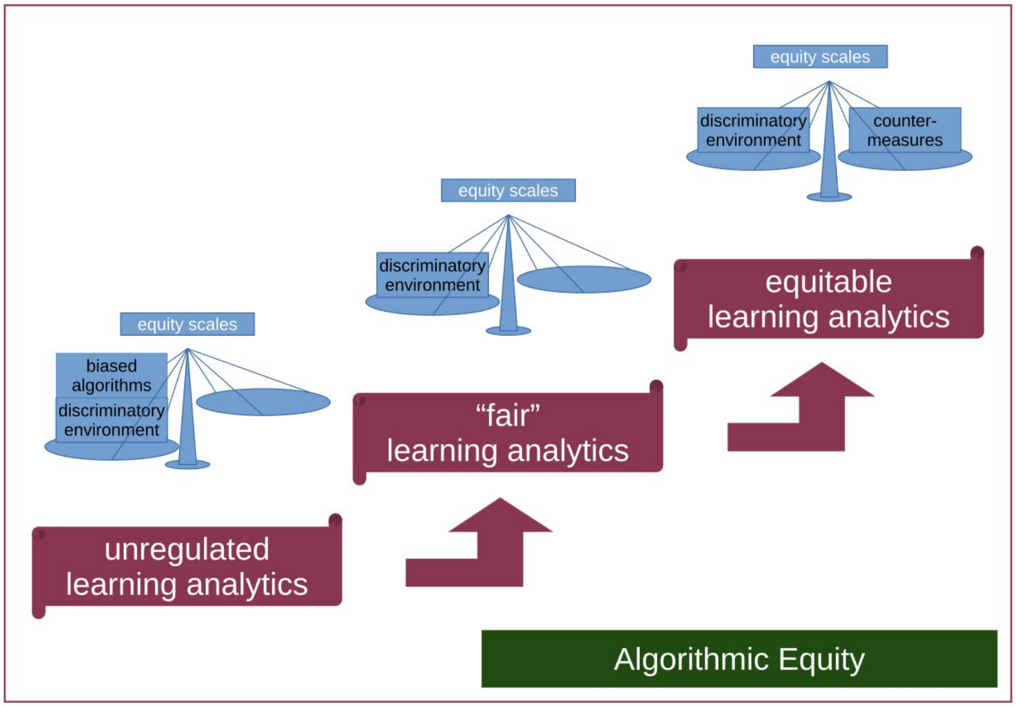

Figure 4. Measures to counter discrimination

4.2.3. How we handled the tension in LPA-AFLEK

In LPA-AFLEK, there was no plan to implement counter-measures, but the future development of counter-measures was prepared. Our theoretical approach built critical consciousness as a modern equity-focused approach with a long tradition — and it was already used in learning analytics (Broughan & Prinsloo, 2020; Francis et al., 2020). In order to have impact, our measures to build critical consciousness needed to connect to the context of physics and teacher beliefs. This also aligns with the theoretical approach of cultural relevant/responsive and sustaining pedagogy (Smith et al., 2022). We planned to contribute to the development of counter-measures aiming at critical consciousness by asking, What level of critical consciousness do physics teachers in Schleswig-Holstein have regarding discrimination phenomena in physics education and learning analytics?

4.2.4. Difficulties

Before designing counter-measures, we needed a strong normative grounding to provide legitimation. This legitimization makes counter-measures understood as working against structural discrimination instead of designed biases. This is especially important within a "hard" sciences environment such as physics, where checklists and algorithms are wider spread than debates about cultural change and discrimination. Normative regulation could look like systematic structures, such as specific funding for addressing equity, including counter-measures. We did, however, also see a normative foundation within the responsible learning analytics community, supported by normative formulations about challenging power and the accountability for our decisions on who to centre. This led us to ask, would preparing future design processes for counter measures be enough to tackle the existing threats and inequalities? Having our obligation to act with respect to the potentials of learning analytics in mind, solving this tension within the given normative foundation was not straight forward. The principle to include those most impacted in the design process would have needed a lot of resources that would have made it increasingly difficult to justify relative to the obligation to act. As these were neither principles of the entire learning analytics or physics education community nor legally binding, we found the existing normative foundation not strong enough to justify more engagement within our project. At the same time, threats and inequalities remained under-addressed.

4.2.5. Which guidance is missing?

In order to be able to address these difficulties, a solid normative ground is needed as legitimization and accountability for learning analytics designers. This could, for example, take the form of a regulatory environment or badge system in the review process of scientific journals. Once such systems are put in place, a tool box of counter-measures can reduce the resources needed in concrete projects and design cases. Counter-measures must be context-sensitive and domain-specific. This includes a sensitivity to the existing critical consciousness regarding sexism, racism, classism, and intersectional discrimination. To

ISSN 1929-7750 (online).

inform the development of regulations and principles, and the design of counter-measures, more context-sensitive practical reports are needed. Prototype designs for context-sensitive counter-measures can be helpful in making the different options for decision-making tangible and understandable for decision makers.

5. Synthesis

In this research, we asked which tensions and edge cases regarding bias and equity emerge when designing a learning analytics system in a physics education context using the existing principles for practice. In doing so, we aimed at contributing to a more equity-focused practice by showing where existing principles fail to provide clear guidance. We propose concrete steps on how to make the general guidance domain-specific and actionable. For physics education, we provided an analysis of historically grown inequalities as well as two focal points for normative decision-making. However, there are more historically grown inequalities in physics education that could be analyzed and the focal points we chose could be chosen differently. The two edge cases that we analyzed from an equity and bias perspective led us to the following conclusions and implications for practice.

5.1. Working towards domain-specific standards and regulations for bias analyses

In the existing literature on bias analyses, many studies focus on reporting percentages and thresholds for concrete applications and the impacts of parameter changes (for example, Gardner et al., 2019; Lohaus et al., 2020). In our first edge case, we took a different approach: We chose an existing method for case analyses and reported on our steps in order to identify missing guidance. We first reflected our understanding of learning analytics and described existing general guidance. In addition to this general perspective, we analyzed our specific domain regarding historically grown inequalities. From there, we were able to identify the most relevant diversity categories as well as evaluation criteria for our specific domain. In order to address potentials and threats as responsible learning analytics do, we formulated two focal points for the analysis. With these focal points, we were able to point at missing guidance and political questions that are not straight forward to answer for practitioners. Instead, the historically grown inequalities in a specific domain need to be analyzed and the focal points for normative ethical decision-making need to be made: How much effort needs to be put into bias analyses for concrete diversity categories and evaluation criteria in order to find a balance between accountability and not strengthening existing inequalities? How pressing is the obligation to act and to harness the potentials of learning analytics in our domain?

For physics education in Germany, we found that existing principles were not concrete and mandatory enough to address the threats of biased algorithms when using learning analytics. Since threats in other domains differ, we recommend developing domain-specific standards and regulations for bias analyses to prevent the under-use of learning analytics due to too strong regulations. Defining these standards and regulations is a normative political task. Researchers can contribute by providing domain-specific, concrete cases, scenarios, and examples.

5.2. Working towards domain-specific counter-measures against intersectional discrimination

We found counter-measures particularly relevant when considering intersectional discrimination. When analyzing more than one diversity category in terms of bias analyses, analyzing for all possible intersections is relevant but often requires much effort. In addition, developing counter-measures becomes an effort itself, especially when those most impacted are to be included. Tension also exists between these efforts and the obligation to act. Always analyzing all biases neither seems feasible nor doable. Instead, we propose working with counter-measures against (intersectional) discrimination. These counter-measures can provide the analysis that discrimination exists but that huge efforts would be needed to avoid it. Additionally, counter-measures can be used to dismantle historically grown inequalities. Whether addressing threats or implementing counter-measures is more feasible is strongly context dependent. Both can be viable tools to dismantle historically grown inequalities.

For physics education in Germany, we found the current normative foundation to build counter-measures not yet solid enough. Even if counter-measures were implemented, a tool box of context-sensitive counter-measures does not yet exist for physics education in northern Germany and would need to be developed. Although defining normative guidance is a political task that includes allocating accountability, research communities can, for example, contribute through their own regulations or badge systems in their journals.

Declaration of Conflicting Interest

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The publication of this article received financial support from the Federal Ministry of Education and Research (BMBF), grant number 01JD2008.

Author Contributions

Conceptualization: A.G., A.S., M.K., K.N.; Methodology: A.G., A.S., M.K., K.N.; Original Draft Preparation: A.G.; Writing-Review and Editing: A.G., A.S., M.K., K.N.; Mentoring: A.S., M.K.; Funding Acquisition: K.N., M.K.; Project Management: M.K., A.G., K.N.; All authors have read and agreed to the published version of the manuscript.

Acknowledgments

Our work is built on the shoulders of various great thinkers who do not yet receive the visibility they deserve. We want to highlight especially the work of Sasha Costanza-Chock, nonbinary trans* femme author, on design justice (2020), and Paulo Freire, from the Global South, on critical consciousness (1970).

References

Aikins, M. A., Bremberger, T., Aikins, J. K., Gyamerah, D., & Yıldırım-Caliman, D. (2021). Afrozensus 2020: Perspektiven, Anti-Schwarze Rassismuserfahrungen und Engagement Schwarzer, afrikanischer und afrodiasporischer Menschen in Deutschland. https://afrozensus.de

Artificial intelligence, platform work and gender equality. (2022). European Institute for Gender Equality. https://doi.org/10.2839/372863

Avraamidou, L. (2019). "I am a young immigrant woman doing physics and on top of that I am Muslim": Identities, intersections, and negotiations. Journal of Research in Science Teaching, 57, 311–341. https://doi.org/10.1002/tea.21593

Baggett, H. C. (2020). Relevance, Representation, and Responsibility: Exploring World Language Teachers' Critical Consciousness and Pedagogies. L2 Journal, 12(2), 34–54. https://doi.org/10.5070/L212246037

Baker, R., & Hawn, A. (2021). Algorithmic Bias in Education. https://doi.org/10.1007/s40593-021-00285-9

Bergner, Y. (2017). Chapter 3: Measurement and its Uses in Learning Analytics. In Handbook of Learning Analytics (1st ed., pp. 35–48). SoLAR. https://doi.org/10.18608/hla17.003

Broughan, C., & Prinsloo, P. (2020). (Re)centring students in learning analytics: In conversation with Paulo Freire. Assessment & Evaluation in Higher Education, 45(4), 617–628. https://doi.org/10.1080/02602938.2019.1679716

Carlone, H. B., & Johnson, A. (2007). Understanding the Science Experiences of Succesful Women of Color: Science Identity as an Analytic Lens. Journal of Research in Science Teaching, 44(8), 1187–1218. https://doi.org/10.1002/tea.20237

Cerratto Pargman, T., & McGrath, C. (2021). Mapping the Ethics of Learning Analytics in Higher Education: A Systematic Literature Review of Empirical Research. Journal of Learning Analytics, 8(2), 123–139. https://doi.org/10.18608/jla.2021.1

Cerratto Pargman, T., McGrath, C., Viberg, O., Kitto, K., Knight, S., & Ferguson, R. (2021). Responsible Learning Analytics: Creating just, ethical, and caring LA systems. Companion Proceedings. LAK21. https://www.solaresearch.org/wp-content/uploads/2021/04/LAK21_CompanionProceedings.pdf

Cheuk, T. (2021). Can AI be racist? Color-evasiveness in the application of machine learning to science assessments. Science Education, 1–12. https://doi.org/10.1002/sce.21671

Costanza-Chock, S. (2020). Design justice: Community-led practices to build the worlds we need. The MIT Press.

D'Ignazio, C., & Klein, L. (2020). Introduction: Why Data Science Needs Feminism. In Data Feminism. https://data-feminism.mitpress.mit.edu/pub/frfa9szd/release/6

Dou, R., Hazari, Z., Dabney, K., Sonnert, G., & Sadler, P. (2019). Early informal STEM experiences and STEMidentity: The importance of talking science. Science Education, 103, 623–637. https://doi.org/10.1002/sce.21499

Drachsler, H., & Greller, W. (2016). Privacy and Analytics – it's a DELICATE Issue. LAK 16, Edinburgh. http://dx.doi.org/10.1145/2883851.2883893

Düchs, G., & Ingold, G.-L. (2018). Frauenanteil bleibt stabil. Physik Journal, 17(8/9), 32–37.

Duncan, R. G., & Rivet, A. E. (2018). Learning progressions. In F. Fischer, C. E. Hmelo-Silver, S. R. Goldman, & P. Reimann (Eds.), International Handbook of the Learning Sciences (pp. 422–432). Routledge.

El-Mafaalani, A. (2021). Mythos Bildung (2nd ed.). Kiepenheuer & Witsch (KiWi).

Erden, D. (2020). KI und Beschäftigung: Das Ende menschlicher Vorurteile oder der Beginn von Diskriminierung 2.0? In Wenn KI, dann feministisch (pp. 77–90). netzforma* eV. https://netzforma.org/publikation-wenn-ki-dann-feministisch-impulse-aus-wissenschaft-und-aktivismus

Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., Rossi, F., Schafer, B., Valcke, P., & Vayena, E. (2018). AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations. Minds & Machines, 28(4), 689–707. https://doi.org/10.1007/s11023-018-9482-5

Francis, P., Broughan, C., Foster, C., & Wilson, C. (2020). Thinking critically about learning analytics, student outcomes, and equity of attainment. Assessment & Evaluation in Higher Education, 45(6), 811–821. https://doi.org/10.1080/02602938.2019.1691975

Freire, P. (1970). Pedagogy of the Oppressed. Penguin Random House UK.

Gardner, J., Brooks, C., & Baker, R. (2019). Evaluating the Fairness of Predictive Student Models Through Slicing Analysis. LAK19: Proceedings of the 9th International Conference on Learning Analytics & Knowledge, 225–234. https://doi.org/10.1145/3303772.3303791

Godwin, A. (2016). The Development of a Measure of Engineering Identity. 2016 ASEE Annual Conference & Exposition, New Orleans, Louisiana. https://doi.org/10.18260/p.26122

Gutachten der Datenethikkommission der Bundesregierung (pp. 1–32). (2019). [Kurzfassung]. DEK. https://www.bmi.bund.de/SharedDocs/downloads/DE/publikationen/themen/it-digitalpolitik/gutachten-datenethikkommission-kurzfassung.pdf?__blob=publicationFile&v=4

Khalil, M., Prinsloo, P., & Slade, S. (2022). A Comparison of Learning Analytics Frameworks: A Systematic Review. LAK22: LAK22: 12th International Learning Analytics and Knowledge Conference, 152–163. https://doi.org/10.1145/3506860.3506878

Kitto, K., & Knight, S. (2019). Practical ethics for building learning analytics. British Journal of Educational Technolog, 50(6), 2855–2870. https://doi.org/10.1111/bjet.12868

Kracke, N., Buck, D., & Middendorff, E. (2018). Beteiligung an Hochschulbildung, Chancen(un)gleichheit in Deutschland. DZHW Brief, 3. https://doi.org/10.34878/2018.03.dzhw_brief

Lohaus, M., Perrot, M., & von Luxburg, U. (2020). Too Relaxed to Be Fair. Proceedings of Machine Learning Research, 119, 6360--6369. https://proceedings.mlr.press/v119/lohaus20a.html

Mitchell, S., Potash, E., D´Amour, A., & Lum, K. (2021). Algorithmic Fairness: Choices, Assumptions, and Definitions. Annual Review of Statistics and Its Application, 8, 141–163. https://doi.org/10.1146/annurev-statistics-042720-125902

Mujtaba, T., & Reiss, M. J. (2013). Inequality in Experiences of Physics Education: Secondary School Girls' and Boys' Perceptions of their Physics Education and Intentions to Continue with Physics After the Age of 16. International Journal of Science Education, 35(11), 1824–1845. https://doi.org/10.1080/09500693.2012.762699

Neue deutsche Medienmacher (NdM)—Glossar. (2021). https://glossar.neuemedienmacher.de/

Odoi, N. (2021). Die Farbe der Gerechtigkeit ist weiß | bpb. bpb.de. https://www.bpb.de/gesellschaft/migration/afrikanische-diaspora/59470/rassismus-im-strafrechtssystem

OECD. (2016). Excellence and equity in education (Volume I; PISA 2015 Results). OECD.

OECD. (2018). Where All Students Can Succeed (Volume II; PISA 2018 Results). OECD.

Ogette, T. (2019). Exit racism (5th ed.). unrast-Verlag.

Phillips, P. J., Hahn, C. A., Fontana, P. C., Broniatowski, D. A., & Przybocki, M. A. (2020). Four Principles of Explainable Artificial Intelligence. National Institute of Standards and Technology. https://doi.org/10.6028/NIST.IR.8312-draft

Prinsloo, P., & Slade, S. (2018). Mapping responsible learning analytics: A critical proposal. In Responsible Analytics & Data Mining in Education: Global Perspectives on Quality, Support, and Decision-Making. Routledge.

Rainey, K., Dancy, M., Mickelson, R., Stearns, E., & Moller, S. (2018). Race and gender differences in how senseof belonging influences decisions to majorin STEM. International Journal of STEM Education, 5. https://doi.org/10.1186/s40594-018-0115-6

Rosa, K., & Moore Mensah, F. (2016). Educational pathways of Black women physicists: Stories of experiencing and overcoming obstacles in life. Physical Review Physics Education Research, 12(2), Article 2. https://doi.org/10.1103/PhysRevPhysEducRes.12.020113

Sclater, N. (2014). Code of practice for learning analytics (pp. 1–64) [Literature review]. Jisc. https://repository.jisc.ac.uk/5661/1/Learning_Analytics_A-_Literature_Review.pdf

Slack, D., Hilgard, H., Jia, E., Singh, S., & Lakkaraju, H. (2020). Fooling LIME and SHAP: Adversarial Attacks on Post hoc Explanation Methods. Proceedings of the 2020 AAAI/ACM Conference on AI, Ethics, and Society (AIES '20), 180–186. https://doi.org/10.1145/3375627.3375830

Slade, S. (2016). The Open University Ethical use of Student Data for Learning Analytics Policy. The Open University. https://doi.org/10.13140/RG.2.1.1317.4164

Smith, T., Avraamidou, L., & Adams, J. D. (2022). Culturally relevant/responsive and sustaining pedagogies in science education: Theoretical perspectives and curriculum implications. https://doi.org/10.1007/s11422-021-10082-4

SoLAR. (2022). What is Learning Analytics? Society for Learning Analytics Research (SoLAR). https://www.solaresearch.org/about/what-is-learning-analytics/

Steegh, A. M., Höffler, T. N., Keller, M. M., & Parchmann, I. (2019). Gender differences in mathematics and science competitions: A systematic review. Journal of Research in Science Teaching, 56(10), Article 10. https://doi.org/10.1002/tea.21580

Suresh, H., & Guttag, J. (2021). A Framework for Understanding Sources of Harm throughout the Machine Learning Life Cycle. EAAMO '21: Equity and Access in Algorithms, Mechanisms, and Optimization, 1–9. https://doi.org/10.1145/3465416.3483305

Traag, V. A., & Waltman, L. (2022). Causal foundations of bias, disparity and fairness. ArXiv. https://doi.org/10.48550/arXiv.2207.13665

Wise, A. F., Sarmiento, J. P., & Boothe Jr., M. (2021). Subversive Learning Analytics. 639–645. https://doi.org/10.1145/3448139.3448210

Yeung, K. (2019). Responsibility and AI (DGI(2019)05; Issue DGI(2019)05). Council of Europe. https://rm.coe.int/responsability-and-ai-en/168097d9c5

Zhai, X., Haudek, K. C., Shi, L., Nehm, R. H., & Urban-Lurain, M. (2019). From substitution to redefinition: A framework of machine learning-based science assessment. Journal of Research in Science Teaching, 57, 1430–1459. https://doi.org/10.1002/tea.21658