

‘A World Where Many Worlds Fit’

*De-Biasing and Critical Consciousness With A Focus On
STEM Identity Development For All Students In
Physics Education Ecosystems As Pluriverses*

Dissertation

Submitted in fulfilment of the requirements for the degree
Doktor der Naturwissenschaften (Dr. rer. nat.) of the
Faculty of Mathematics and Natural Sciences
of Kiel University

Submitted by
Adrian Grimm

Kiel, January 2025

First Reviewer (Erste*r Gutacher*in):
Second Reviewer (Zweite*r Gutachter*in):
Mentor (Betreuer*in):
Mentor (Betreuer*in):

Prof. Dr. Knut Neumann
Prof. Dr. Marcus Kubsch
Dr. Anneke Steegh
Prof. Dr. Marianela Navarro Camacho

Date of Oral Examination:

April 14th, 2025

Thanks

The time of my dissertation ends as it began: I am living together with Robi and Milena. Well, different place, different time. Same same but different. In between there was our shared time together with Meli, soon Rafa is going to move in. And all of you have been close people in my daily life, carrying me whenever needed. I'd like to express my gratitude to all of you, the other wonder-full human beings who have been with me throughout the last years, in two ways: By sharing one poem that I have written in these last years at the end of each content-section and by inviting you now to a journey through my last years:

I started working at IPN during the pandemics in 2021: Home office. My first day working in the offices of IPN would be months later. Back in that time, I called or video-called all colleagues from my department at least one time for a 1-to-1-session in order to get to know the very people I was going to work with. From that time, for example, I am still in contact with Mirijam and we keep meeting every once in a while, even though Mirijam switched the department and the building at IPN. In these first months I had more frequent contact with Onur, Sebastian, and Isa than with any colleague from IPN – while the three of you would work with me in my project from Frankfurt and Essen! I also still remember my first encounter with Ulrike at May 1st, random in-person at the Exer: "Ah, well, nice to see you – I really did not expect to meet anyone from IPN here but I am really happy to see you!" I still remember the happiness I felt after you had come over to talk a bit with me. You gave me that feeling of having made the right choice to go to IPN. And that was important because I remember that one question that would stay with me for all these years already came up in these first months: "Adrian, where's the physics education?!" Knut would come back and back to that questions, pushing me to connect my ideas stronger to those of the department. And Knut, you took so much time for giving me feedback, more than I had ever expected. I went through all of them and I can say: Even though we did not always agree, I learned something from each of your feedbacks. I feel incredibly grateful for the strong support I have enjoyed from the very beginning, also from my mentors. Marcus who would support me in visiting international conferences from a very early point an. Anneke who was the very reason that I applied at IPN because I had seen: Ah, there are some people publishing on feminist topics!

Even though I had a very strong support network at IPN, the strongest source of energy and support has been outside of IPN for: Robi, Sonni, Mari, and Milena with whom we've built the local chapter of digitalcourage which recently transformed in Digitalmöwen. Morlin and Anna who were with me in my first steps within the Greens. Christina and the yearly trainer work at verdi, offering Bildungsurlaub together – strong unionist positions included. Our swimming sessions, every morning from Monday until Friday between May and September with so many of you, beginning with Mari and Milena. Our anti-racist reading circle which we carried on for years with Leo and Carlotta, until today. In addition to these clearly outside-IPN-networks, I have always tried find people at work with whom I'm first of all friends and only then colleagues – and I'd say we were very successful! In 2022, we founded our Personalrat with Lucas, Matthias, Jan, David, Steffi, Katharina, Kathryn, Kathi, Jannik, und Franzi. Together we wrote the Dienst-Vereinbarung Awareness, established an IPN-public discourse around EntfristungStattBefrustung, made transparent the conflict of interests between as much research as possible on the one hand and good working conditions on the other hand, invited to cultural transformation through narrative auto-actualisation in a fulminant presentation at a conference at IPN, we have risen the awareness for trans people by introducing tampons also men's toilets with an information about the why of that, and so on... Together with Anna, Johannes, Caro, Tobi, Gyde,

Chrissie, Jaika, Martin, and Jan, we have established the IPNosaurus, our IPN-PhD-T-Rex. Together with Mari, we established morning-sauna-sessions at Kieler Förde and have opened the local sauna for more than a year at least once a week now. We have founded a strong dance-karaoke-group who would not only go to the local bar Schaubude biweekly but also bring dance-karaoke to IPN-parties with Amelie, Berrit, Chrissie, Franzi, Gyde, Jannik, Johannes, Anja, Kathi, Lasse, Lea, Leander, Marie, Simon, Stephan, Jasmin, and Caro – and, of course, Milena, Robi, Maria, and many others who do not work at IPN are part of that dance-karaoke-movement as well. Here's where the limits vanish most clearly. And all those groups, events, moments, experiences, memories, they are what carried me through those last for years and where I take huge parts of my daily energy from!

On January 23rd in 2023, Anneke and I talked for the first time about my plans of going to Costa Rica. I still remember asking you: I would like to go to Costa Rica. I believe it makes sense but I cannot see any established contact to a university there. What do you think, can I find a way to go there? Within hours you had found a list of people in Costa Rica and Latin America who I could contact – among them Nela. I prepared a mail for Knut, and then a mail to Nela. I had many doubts whether it will work out. And then you, Nela, simply replied not even 24 hours after my mail to you on March 21st: "Estimado Adrián: Gracias por su interés por trabajar conmigo. Con gusto podemos recibirte en la Universidad de Costa Rica" (Dear Adrian, thanks for your interest in working with me. With pleasure we can receive you at the University of Costa Rica). And that's what happened, almost one year later Milena, Mari, and I sat in the plane to Costa Rica, on January 14th, the very day Robi opened his dancing school Hand zu Hand in Kiel. In Costa Rica, I've learned so much: From you, Nela, as a person who would not only take incredibly much time for me but would, for example, also take me to the Communion of your niece. But also from our colleagues Verónica who took me to manifestations, Fran who travelled with me, Alí and Luís who taught Bribri to me, Carlos and Mónica who worked on outlines for scientific papers with us, Dorita, Dora, Fio, Popeye, and Pepe who lived with me, and from the travels I have made with my Mum over there in the mountains of Monteverde. I am so grateful for every single day of these four months in Centro-America. And I am so happy that we managed to keep the contact established and that Berny and Nela came for two weeks to Kiel and Schleswig-Holstein in September 2024 thanks to the great enthusiasm of Knut, Lorian, and everybody involved – with two more months of research stay to come in June and July 2025 to which I am looking forward already!

There are more stories to be told: the support of Bianca in the creation of our PhD-Bollerwagen and the happiness about the glitter at the entrances of the Leibniz-Personalrats-Treffen, Angelika's support in the preparation of my research stay in Costa Rica, Benny and Annina supporting Nela and me in grant application preparation and our Scrumademics project management group with Johannes, Mareike, Paul and Caro. The start of my teaching involvement with engaged discussions with the physics teacher students. The work in the Klima-und-Nachhaltigkeits-Gruppe or the Gesundes-Institut. Our daily lunch breaks with the second floor, our spikeball-tournaments, our boozel-tour. The great support of my family: my mum and dad, Bernd and Regina, Micha, Jesper and Malin, Melvin, Robi, my grandmum Helga and my granddad Heinrich who moved closer to Kiel in about the time when I started my dissertation, and many, many others.

Thanks to all of you. You are the reasons why I have become the way I am today, why I can be who I am today with all my energies, and why I look confidently and positively in our futures. Let's make it more peaceful and just worlds as we say at AFS – 'worlds where many worlds fit'.

Summary

Modern education systems typically shall enable participation and social mobility for all students. However, physics education is characterised by historically grown inequalities along various dimensions, for example more male enter physics-related careers compared to their female or non-binary peers. Even worse, these inequalities tend to reproduce themselves, for example through the mechanism of vulnerability due to underrepresentation towards a lack of recognition of students. In order to break the vicious cycle of reproduction of historically grown inequalities, active structural intervention is needed. In the four presented pieces of scholarship, two central providers of recognition for students are focused: teachers and artificial intelligence algorithms which are used more and more in physics education contexts. The concrete questions asked in the four pieces of scholarship are informed by the greater challenge to provide evidence that informs the design of physics education ecosystems that invite all students.

In the first piece of scholarship, a theoretical framework is developed that describes the impact of artificial intelligence systems on the STEM identity development of students who face historically grown inequalities. The underlying assumption is that the vicious cycle of the reproduction of inequalities in physics education does not manifest in competence development only but instead a broader focus on STEM identity is needed in order to explain the reproduction of inequalities. The developed theoretical framework explains that for students who identify, for example, as female or as students with a low socio-economic status recognition is especially crucial due to their vulnerability. The major contribution of the theoretical framework is to shift the focus of research on algorithmic bias towards the arenas in which the reproduction of inequalities actually takes place: STEM identity development. Relevant suppositions for future research are proposed, including the focus not only on bias in identifying students with low competence levels who are at risk to lose track, but instead also investigate biases in identifying students with high competencies who need recognition in order to successfully develop a STEM identity.

In the second piece of scholarship, the analyses of two concrete cases reveal where existing principles fail to provide guidance for preventing bias in artificial intelligence systems. Research fields such as learning analytics have a long tradition in investigating and discussing algorithmic bias. However, the impact on praxis has been shown to be little, leaving the principles without effect. The piece of scholarship contributes 1) an analysis where exactly guidance is missing and 2) a proposal of a domain-specific process how the missing guidance can be added for the example of physics education. In order to address the specific historically grown inequalities in physics education and having in mind the theoretical framework of the reproduction of these inequalities, concrete diversity dimensions to start with and evaluation criteria that include both the high- and the low-performing students are proposed. Additionally, the role of normativity and two possible normative standpoints is discussed and the implications of both possible normative decisions are highlighted. The conclusion is that 1) domain-specific principles are needed and that 2) counter-measures against intersectional discrimination are needed.

In the third piece of scholarship, two concrete biases and possible approaches to identify and reduce these biases are investigated quantitatively. Well in line with the risk-based approach of regulation of artificial intelligence systems of the European Union, the potential early identification of threats of biases is investigated. The core idea is: If an artificial intelligence system can be trained well to predict whether a student is, for example, male or female, the student answers that were used for the algorithmic training need to contain gendered patterns and hence the threats of bias are bigger. A possible

reduction of bias is investigated by different training dataset configurations. The core idea is: If students are well-represented in the training dataset, the bias should reduce. In a critical discussion from a feminist and de-colonial perspective, it is highlighted that both approaches seem to have promising potentials and relevant limitations as well.

In the fourth piece of scholarship, the focus is shifted away from artificial intelligence systems and moved towards teachers and their role in making physics education ecosystems places that invite all students. Drawing heavily from theory established in the Americas and by scholars from the Global South, the Critical Consciousness of physics teachers in Germany was explored qualitatively. 14 interviews were conducted with seven teachers and afterwards coded with a coding manual developed by researchers from Germany and Costa Rica. The coded interviews were then grouped into different types which represented different resources and obstacles for the development of Critical Consciousness of teachers. The typology can serve as a guide for professional developments of physics teachers in Northern European contexts.

The four studies provide a solid ground that future research can be built upon. The implications of specific normative standpoints as well as the need for research based on different normative standpoints were discussed and highlighted. From a pluriversal standpoint which was the point of departure for the four studies, various future research directions emerged from the studies on Critical Consciousness and De-Biasing. For Critical Consciousness, qualitative explorations of the concrete mechanisms and effects on a micro-level of teachers' Critical Consciousness on students' STEM identity development are needed to increase understanding as a resource for professional developments. Additionally, criteria for programmes with long-lasting effects that aim at increasing teachers' Critical Consciousness in the context of physics education in Northern Europe need to be investigated and culture-sensitive quantitative instruments are needed for up-scaling. For De-Biasing, the provided evidences can serve as indications for directions of future research but do not yet lay a solid ground for evidence-informed political decision-making. More evidences are needed to 1) investigate whether the findings on effectiveness remain stable along other datasets, and 2) compare the effects of the concrete approaches in order to find the most efficient approach. Additionally, the findings indicate that De-Biasing has limited potentials and in order to reach a pluriverse additional counter-measures are needed. Counter-measures can include increasing teachers' Critical Consciousness and/or using artificial intelligence systems not only for reducing bias but, for example, also for actively supporting teachers in making materials more inviting especially for currently under-served students. Ultimately, research on both De-Biasing and Critical Consciousness with a focus on STEM identity development can inform the design of physics education ecosystems as pluriverses that successfully invite all students – creating a “world where many worlds fit” (Escobar, 2017; Mignolo, 2007).

Zusammenfassung

Moderne Bildungssysteme sollen typischerweise Teilhabe und soziale Mobilität für alle Schüler*innen ermöglichen. Bildung in der Physik ist allerdings charakterisiert durch historisch gewachsene Ungleichheiten entlang mehrerer Dimensionen, zum Beispiel beginnen mehr männliche als weibliche oder nicht-binäre Schüler*innen eine Physik-bezogene Karriere. Schlimmer noch, die bestehenden Ungleichheiten tendieren dazu reproduziert zu werden, beispielsweise durch den Mechanismus der Vulnerabilität einiger Schüler*innen gegenüber fehlender Anerkennung aufgrund von Unter-Repräsentation. Um diesen Kreislauf der sich reproduzierenden Ungleichheiten zu durchbrechen, braucht es strukturelle Intervention. In den vier präsentierten Arbeiten sind zwei zentrale Geber*innen von Anerkennung im Fokus: Lehrer*innen und Algorithmen auf Basis Künstlicher Intelligenz, die im Physik-Unterricht mehr und mehr genutzt werden. Die wissenschaftlichen Fragestellungen der einzelnen Arbeiten haben ihre Basis in der übergeordneten Herausforderung, Evidenz für ein Design von Physik-Bildungs-Ökosystemen zu sammeln, in die alle Schüler*innen eingeladen sind.

In der ersten Arbeit wird ein theoretischer Rahmen entwickelt, der den Einfluss von Systemen Künstlicher Intelligenz auf die MINT-Identitäts-Entwicklung von Schüler*innen beschreibt, die Diskriminierung aufgrund historisch gewachsener Ungleichheiten erfahren. Die zugrunde liegende Annahme ist, dass der Kreislauf der Reproduktion von Ungleichheiten in der Physik-Bildung sich nicht nur in der Kompetenz-Entwicklung manifestiert, sondern dass stattdessen eine breitere Perspektive auf MINT-Identität notwendig ist, um die Reproduktion der Ungleichheiten erklären zu können. Der entwickelte theoretische Rahmen erklärt, dass für weibliche Schüler*innen oder Schüler*innen mit niedrigem sozio-ökonomischem Status Anerkennung besonders wichtig ist aufgrund ihrer spezifischen Vulnerabilität. Der größte Beitrag des theoretischen Rahmens liegt in der Fokus-Verschiebung innerhalb der Forschung zu algorithmischem Bias hin zu dem Ort, an dem die Reproduktion der Ungleichheiten in der Physik-Bildung tatsächlich stattfindet: MINT-Identitäts-Entwicklung. Relevante Suppositionen für zukünftige Forschungs-Arbeiten werden vorgeschlagen, inklusive des Fokus nicht nur auf Bias bei der Identifikation von Schüler*innen mit niedrigen Kompetenz-Niveaus sondern auch auf Bias bei der Identifikation von Schüler*innen mit hohen Kompetenz-Niveaus, die Anerkennung ihrer Leistungen brauchen, um eine MINT-Identität entwickeln zu können.

In der zweiten Arbeit wird anhand der Analyse von zwei konkreten Fallbeispielen aufgezeigt, wo existierende Leitlinien zum Umgang mit Bias nicht ausreichend Orientierung anbieten, um Bias in Systemen Künstlicher Intelligenz zu verhindern. Viele Forschungsfelder wie das der Learning Analytics haben eine lange Tradition in der Erforschung und Diskussion von algorithmischem Bias. Dennoch: Studien zeigen, dass der Effekt der existierenden ethischen Leitlinien auf die Design-Praxis der Algorithmen gering ist, die Leitlinien also keine Wirkung entfalten. Die vorliegende Arbeit sucht diese Lücke zu überbrücken, indem 1) eine Analyse vorlegt wird, wo genau Orientierung für die Praxis fehlt, und 2) am Beispiel der Physik-Bildung ein Vorschlag für einen Domänen-spezifischen Prozess ausgearbeitet wird, wie diese fehlende Orientierung hinzugefügt werden kann. Um die Physik-spezifischen historisch gewachsenen Ungleichheiten dem in der ersten Arbeit entwickelten theoretischen Rahmen folgend adressieren zu können, werden konkrete Diversitäts-Dimensionen vorgeschlagen und Evaluations-Kriterien abgeleitet, die sowohl niedrig- als auch hoch-leistende Schüler*innen berücksichtigen. Zusätzlich werden die Rolle von Normativität und zwei mögliche normative Standpunkte diskutiert sowie die Implikationen beider Standpunkte beleuchtet. Die wichtigsten

Schlussfolgerungen lauten, dass 1) Domänen-spezifische Leitlinien notwendig sind für Wirksamkeit in der Design-Praxis, und dass 2) zusätzliche Maßnahmen neben der Weiter-Entwicklung der Leitlinien gegen intersektionale Diskriminierung notwendig sind.

In der dritten Arbeit werden zwei konkrete Biases und mögliche Herangehensweise quantitativ erforscht, um diese Biases zu identifizieren und zu reduzieren. Sich mit dem Risiko-basierten Ansatz der Regulierung von Systemen Künstlicher Intelligenz auf Ebene der Europäischen Union gut deckend wird das Potenzial erforscht, Risiken für Bias möglichst früh zu erkennen. Die Kern-Idee lautet: Wenn ein System Künstlicher Intelligenz darauf trainiert werden kann, die Geschlechts-Identität von Schüler*innen mit den Schüler*innen-Antworten zuverlässig vorherzusagen, dann enthalten die Schüler*innen-Antworten Geschlechts-spezifische Muster, das Risiko für Bias ist also größer. Eine mögliche Reduktion von Bias wird über verschiedene Konfigurationen der Trainings-Datensätze erforscht. Die Kern-Idee lautet: Wenn Gruppen von Schüler*innen in den Trainings-Daten gut repräsentiert sind, sollte der Bias reduziert werden. In einer Diskussion aus feministischer und de-kolonialer Perspektive wird herausgearbeitet, dass beide Ansätze vielversprechende Potenziale aber auch relevante Limitationen enthalten.

In der vierten Arbeit wechselt der Fokus weg von Systemen Künstlicher Intelligenz hin zu Lehrer*innen und ihrer Rolle dabei, Physik-Bildungs-Ökosysteme zu Orten zu machen, die alle Schüler*innen einladen. Aufgebaut wird dabei auf in den Americas durch Forscher*innen des Globalen Südens etablierten Theorien. Die Arbeit umfasst eine qualitative Exploration des Kritischen Bewusstseins von Physik-Lehrer*innen in Deutschland. Es wurden 14 Interviews mit insgesamt Sieben Lehrer*innen durchgeführt und anschließend codiert mit einem Codier-Manual, das von Forscher*innen aus Deutschland und Costa Rica entwickelt wurde. Die Interviews wurden genutzt für die Entwicklung einer Typologie, auf deren Basis die Interviews entsprechend verschiedener Ressourcen und Hindernisse von Lehrer*innen bei der Entwicklung von Kritischem Bewusstsein eingeteilt werden konnten. Die so entwickelte Typologie kann als Orientierung dienen für Aus- und Fortbildungen von Physik-Lehrer*innen in nordeuropäischen Kontexten.

Die vier Arbeiten sind ein solider Grund für zukünftige Forschung. Die Implikationen von spezifischen normativen Standpunkten als auch die Notwendigkeit von Forschung aus unterschiedlichen normativen Standpunkten heraus wurden diskutiert und hervorgehoben. Aus einer normativen Perspektive des Pluriversums heraus kamen mehrere mögliche Richtungen für zukünftige Forschung an Kritischem Bewusstsein und De-Biasing auf. Für Kritisches Bewusstsein wurde die Notwendigkeit der qualitativen Exploration von konkreten Mechanismen und Effekten von Kritischem Bewusstsein von Lehrer*innen auf die MINT-Identitäts-Entwicklung von Schüler*innen auf der Mikro-Ebene als möglicher nächster Schritt identifiziert, um dadurch weitere Ressourcen für die Aus- und Fortbildung von Lehrer*innen zu erschließen. Zusätzlich braucht es die Erforschung von Erfolgs-Faktoren von Aus- und Fortbildungs-Programmen zur Entwicklung von Kritischem Bewusstsein mit lang-anhaltenden Effekten im Kontext von Physik-Bildung in Nordeuropa. Für die Skalierung und damit Wirkung in der Breite sind ebenfalls Kultur-sensible, quantitative Erhebungs-Instrumente für Kritisches Bewusstsein notwendig. Für De-Biasing konnten Indikationen für Richtungen zukünftiger Forschung erarbeitet werden, die allerdings noch keinen soliden Grund für Evidenz-informierte Politik darstellen. Mehr Evidenz ist notwendig, um 1) die Indikationen über Evidenzen aus anderen Datensätzen abzusichern, und 2) die effizienteste Herangehensweise zu identifizieren. Auch die Limitationen von De-Biasing zur Erreichung eines Pluriversums und die Notwendigkeit von weiteren Gegen-Maßnahmen wurde herausgearbeitet. Zusammen genommen können

De-Biasing und Kritisches Bewusstsein mit einem Fokus auf MINT-Identitäts-Entwicklung
die Gestaltung von Physik-Bildungs-Ökosystemen als Pluriversen informieren, die
erfolgreiche alle Schüler*innen einlädt – „eine Welt, in die viele Welten passen“ (Escobar,
2017; Mignolo, 2007).

Table of Contents

Thanks.....	III
Summary	V
Zusammenfassung	VII
Table of Contents.....	X
List of Figures	XIV
List of Tables	XVI
Declaration of Ethical Conduct	XVII
1 Introduction.....	1
1.1 STEM Education for all	1
1.2 Project Design	2
1.3 Overview of the Four Pieces of Scholarship.....	4
2 Theory	10
2.1 STEM Education for all	10
2.1.1 STEM Education for all and Historically Grown Inequalities in Physics Education.....	10
2.1.2 The Need for a more Concrete Justice Theory.....	11
2.1.3 Justice Theory and Focus: The Pluriverse and STEM Identity Development	12
2.2 De-Biasing and Critical Consciousness.....	14
2.3 Overarching Research Questions	16
3 Responsible Learning Analytics and STEM Identities	20
3.1 Introduction.....	21
3.2 STEM identities of under-served students.....	22
3.2.1 STEM identities.....	22
3.2.2 Under-served students.....	23
3.2.3 Intersectionality and individual needs of students	24
3.2.4 Summary: STEM identities of under-served students.....	24
3.3 Responsible learning analytics.....	25
3.3.1 Learning analytics.....	25
3.3.2 Learning analytics and responsibility.....	26
3.3.3 Summary: Responsible learning analytics.....	27
3.4 Responsible learning analytics in the context of STEM identities of under-served students.....	27
3.4.1 Proposal of a theoretical framework.....	27
3.4.2 Issues with normativity: Bias and equity.....	28

3.4.3	Suppositions	29
3.5	Discussion.....	32
3.5.1	Implications for responsible learning analytics researchers	33
3.5.2	Connections to critical consciousness of teachers.....	33
3.5.3	Conclusions for STEM identity development of under-served students	34
	References of the Piece of Scholarship	36
4	Equity-Focused Decision-Making Lacks Guidance!	42
4.1	Introduction and Background	43
4.2	Theory.....	44
4.2.1	Responsible Learning Analytics	44
4.2.2	Guiding Principles for Practice	46
4.2.3	Historically Grown Inequalities in Physics Education in Germany.....	48
4.2.4	Two Focal Points for Normative Decision Making: Equity and Bias	50
4.3	Methods	50
4.4	Edge Cases	51
4.4.1	Edge Case 1: How much effort is enough? Need for intersectional bias analyses versus obligation to act.....	51
4.4.2	Edge Case 2: Bias-free is not enough! Need for counter-measures versus obligation to act	54
4.5	Synthesis	56
4.5.1	Working towards domain-specific standards and regulations for bias analyses56	
4.5.2	Working towards domain-specific counter-measures against intersectional discrimination	57
	References of the Piece of Scholarship	58
5	De-Biasing	64
5.1	Introduction	65
5.1.1	Relevance and Contribution	65
5.1.2	Authors' positions based on feminist standpoint theory.....	66
5.2	Theoretical Background	66
5.2.1	Identity Development	66
5.2.2	Learning Analytics.....	68
5.2.3	Bias in Learning Analytics	69
5.2.4	De-Biasing Learning Analytics	70
5.2.5	Research Questions.....	72
5.3	Methodology	73
5.3.1	Research Design	73
5.3.2	Data Base	74

5.3.3	Data Analysis Procedure	76
5.3.4	Algorithmic Architecture	79
5.4	Results.....	79
5.4.1	Reducing Bias: Slicing Analysis	80
5.4.2	Identifying Threats of Bias: Training Dataset Analysis.....	84
5.5	Discussion of Implications.....	86
5.5.1	Implications for Education Researchers.....	86
5.5.2	Implications for Political Decision Makers	86
5.5.3	Implications for Learning Analytics Researchers.....	87
5.6	Conclusions	88
5.6.1	De-biasing through regulation of training datasets – a promising starting point	88
5.6.2	6Needs Addressing Researchers and Politicians.....	88
	Author Contributions.....	88
	Acknowledgments	89
	References of the Piece of Scholarship.....	90
6	Critical Consciousness.....	98
6.1	Introduction.....	99
6.2	Theory	100
6.2.1	Navigating Meritocracy and Social Justice: Inequalities in Physics Education	100
6.2.2	Critical Consciousness as a Means of Preparing Teachers.....	101
6.2.3	Professional Development Needs for Critical Consciousness in Northern Europe	102
6.3	Methodology and Data.....	103
6.3.1	Setting the Context: Semi-Structured Interviews within the Project	103
6.3.2	Type-Building Deductive Qualitative Analysis	103
6.3.3	Reflections on the Methodology.....	106
6.4	Results: Professional Development Needs	107
6.4.1	Critical Consciousness in the Context of Northern Europe	107
6.4.2	Typology	108
6.4.3	Resources and Obstacles for Professional Developments.....	110
6.5	Discussion	114
6.5.1	Critical Consciousness as Structure Against Inequalities in Identity Development.....	114
6.5.2	Implications for Research Communities.....	114
6.5.3	Limitations	115
6.6	Synthesis	116

Acknowledgements.....	116
Supplemental Material	116
Author Contributions	116
Supplemental Material: Coding Manual Critical Consciousness.....	118
References of the Piece of Scholarship	131
7 General Discussion	136
7.1 Construct-Specific Discussions over Findings from all Pieces of Scholarship	136
7.1.1 De-Biasing: Addressing Bias – Mitigation Possible, Elimination Out of Sight	136
7.1.2 Critical Consciousness: A Promising Piece for a Social Justice Architecture	138
7.1.3 STEM Identity Development in a Pluriverse: Action Needed	141
7.2 Implications	143
7.2.1 Research	143
7.2.2 Practice	144
7.2.3 Policy	145
8 References of the Dissertation Frame	150

List of Figures

Figure 1-1 – Project: Overview.....	3
Figure 1-2 - Project: Scoring of Student Answers.....	3
Figure 1-3 - Project - Teacher Facing Dashboard	4
Figure 2-1 - Pluriverse.....	14
Figure 3-1 - STEM identities of under-served students.....	25
Figure 3-2 - Responsible learning analytics.....	27
Figure 3-3 - Responsible learning analytics in the context of STEM identities of under-served students	29
Figure 4-1 - Responsible learning analytics — practice between potentials and threats	45
Figure 4-2 - Threats in learning analytics for physics education	49
Figure 4-3 - Bias analyses	52
Figure 4-4 - Measures to counter discrimination.....	55
Figure 5-1 - STEM Identities of Under-Served Students and Responsible Learning Analytics (Grimm et al., 2023).....	69
Figure 5-2 - Bias Entry Points in Learning Analytics and Focus on This Paper (inspired and informed by (Baker & Hawn, 2021, p. 9))	71
Figure 5-3 - Example item, answer, label, and score – the item “How should solarcells be installed on a roof in order to transform as much energy as possible?” with the answer “in a way that as much energy can be transformed as possible” with the label “transformation process” scored positively as transformation process identified.....	74
Figure 5-4 - Identifying Threats of Bias: Training Dataset Analysis.....	77
Figure 5-5 - Reducing Bias: Slicing Analysis	79
Figure 5-6 - Gender and Biases for all Label; MEE – Manifestation Electric Energy, MEv – Manifestation Electric variable, MRE – Manifestation Radiant Energy, MRv – Manifestation Radiant variable, MTE – Manifestation Thermal Energy, MTv – Manifestation Thermal variable, TP – Transformation Process, M_lp – Manifestation Learning Performance, T_lp – Transformation Learning Performance	81
Figure 5-7 - Language and Biases for all Label; MEE – Manifestation Electric Energy, MEv – Manifestation Electric variable, MRE – Manifestation Radiant Energy, MRv – Manifestation Radiant variable, MTE – Manifestation Thermal Energy, MTv – Manifestation Thermal variable, TP – Transformation Process, M_lp – Manifestation Learning Performance, T_lp – Transformation Learning Performance	82
Figure 5-8 -Educational Background and Biases for all Label; MEE – Manifestation Electric Energy, MEv – Manifestation Electric variable, MRE – Manifestation Radiant Energy, MRv – Manifestation Radiant variable, MTE – Manifestation Thermal Energy, MTv – Manifestation Thermal variable, TP – Transformation Process, M_lp – Manifestation Learning Performance, T_lp – Transformation Learning Performance	83
Figure 6-1 - Typology built by analyses for Critical Action: 1) Initial Action (n=7) and 2) First Critical Action (n=7)	108
Figure 6-2 - Results for Initial Action on Critical Consciousness with its Main Categories	110
Figure 6-3 - Results for First Critical Action on Critical Consciousness with its Main Categories	112

Figure 7-1 - Results from all pieces of scholarship with relevance to De-Biasing – I refer to the fields by indicating the field, for example ‘DB-1-a’ refers to the results on De-Biasing from this figure, piece of scholarship 1 the theoretical model, and the field a) special focus needed: STEM identity and under-served students	136
Figure 7-2 - Results from all pieces of scholarship with relevance to Critical Consciousness – I refer to the fields by indicating the field, for example ‘CC-1-a’ refers to the results on Critical Consciousness from this figure, piece of scholarship 1 the theoretical model, and the field a) rather stable historically grown inequalities	139

List of Tables

Table 3-1 - Suppositions	31
Table 5-1 - Numbers of Students per Diversity Dimension	74
Table 5-2 - Numbers of Student Answers per Diversity Dimension	75
Table 5-3 - Numbers of Student Answers per Diversity Dimension Well-Served/Under-Served per Label Positive/Negative; Abbreviations: lan – Language, gen – Gender, edu – Educational Background; ws – Well-Served, us – Under-Served; pos – Positive, neg – Negative; MEE – Manifestation Electric Energy, MEv – Manifestation Electric variable, MRE – Manifestation Radiant Energy, MRv – Manifestation Radiant variable, MTE – Manifestation Thermal Energy, MTv – Manifestation Thermal variable, TP – Transformation Process, M_lp – Manifestation Learning Performance, T_lp – Transformation Learning Performance, colour scheme groups relevant numbers for one model training (27 models in total)	76
Table 5-4 - Results for Gender in Training Dataset Analysis	84
Table 5-5 - Results for Language in Training Dataset Analysis	84
Table 5-6 - Results for Educational Background in Training Dataset Analysis.....	85
Table 6-1 - Coding Manual on Critical Consciousness	104
Table 6-2 - Comparing Initial Action (IA) and First Critical Action (FCA).....	109

Declaration of Ethical Conduct

I hereby declare that the work presented in this dissertation – apart from the advice given by my supervisors – is my own in both format and content. Two research articles (sections 3 and 4) presented in this dissertation have been published in peer-reviewed scientific journals. Two manuscripts (sections 5 and 6) have been submitted for publication in peer-reviewed scientific journals. This is my first dissertation and the contents have not been used in any prior attempts to obtain a PhD. The dissertation complies with the standards for good scientific practice as proposed by the German Research Foundation. I have not been deprived of an academic degree.

In the process of refining the manuscript from section 6, language editing was conducted with the assistance DeepL, a language model that helps to enhance the clarity, coherence, and style of the text, ensuring that the integrity of the original content was maintained.

Kiel, Januar 3rd 2025

Adrian Grimm

1 Introduction

1.1 STEM Education for all

Modern education systems commonly aim to enable all students to participate in society as well as to make social mobility possible (*EU Charter*, 2012; KMK, 2020; MNC, 2021; Muñoz Izquierdo, 2012). In many countries, including Germany, social justice in education along dimensions of diversity¹ is a means to reach this aim. Dimensions such as gender or social class should not play a role when it comes to access to education. Women should not face stronger barriers in physics education than men and students with parents who studied at the university should not have an easier access to careers in physics than students whose parents did not study. However, historically grown inequalities in terms of gender or socio-economic status persist in physics education. For example, only 27.3 % of all students in engineering, manufacturing and construction within the European Union were women in 2022 (Eurostat, 2022). Reducing the existing inequalities is a crucial task for societies in order to maintain social cohesion by fulfilling the promise of participation and social mobility for all. The role of science is to provide evidence for how to allow for participation and social mobility for all students in an effective and efficient way.

One growing approach to enable participation for all students in education are artificial intelligence systems. Artificial intelligence systems come with the promise to offer individualised support to students who need it where teachers currently cannot provide such individual support due to the sheer number of students in one classroom. For example, artificial intelligence systems can provide immediate and individualised feedback to students (Dennis et al., 2016; Pardo et al., 2019), and can be used to automate tasks (Zhai et al., 2019, p. 1145) which frees time for teachers who can use that time to provide students with the individual support they need. However, it has been shown that the use of artificial intelligence systems does not come without threats. Artificial intelligence systems can be racist (Dressel & Farid, 2018), biased against students with lower socio-economic status (Fletcher et al., 2021), and reproduce gender stereotypes (Bolukbasi et al., 2016). For example, if artificial intelligence systems systematically tend to over-estimate the abilities of *white*² students, that can lead to a so-called 'benefit of doubt' for *white* students, providing these students with an increased level of recognition (Jeong et al., 2021). Recognition is a central dimension of STEM identity development (Carlone & Johnson, 2007) which means that, staying with the example, Black³ students can be less likely to develop a STEM identity. Such a behaviour of artificial intelligence systems is concerning as it further disadvantages groups of students who are the most vulnerable, such as historically under-represented groups of students in physics education. Hence, artificial intelligence systems in physics education may not only fail to live up to the promise of increased participation for all students but instead comes with substantial threats to make education less equitable (Uttamchandani & Quick, 2022).

¹ Dimensions of diversity in this dissertation refers to the seven dimensions gender, race, socio-economic status, religion and beliefs, ability and chronic illnesses, sexual identity, and age.

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).

³ We write Black instead of black in order to make a reference to the strategy of empowerment – the resistance against racism (M. M. Eggers et al., 2023, p. 13; Gunda-Werner-Institut & Center for Intersectional Justice, 2019, p. 6).

1 Introduction

In order to provide STEM education for all students, the potentials of artificial intelligence systems for participation need to be harvested while the threats of decreasing equity need to be prevented. STEM education needs to provide learning opportunities for students of all gender and of any socio-economic status. Students bring different histories, strengths, needs, and vulnerabilities to physics classes. One approach to analyse inequalities along dimensions of diversity, such as gender and socio-economic status, is to design STEM education as 'a world where many worlds fit', to design for a 'pluriverse' (Escobar, 2017; Kayumova & Dou, 2022). Next to harvesting the potentials of artificial intelligence systems, designing for a pluriverse allows to address the complexity of impacts and threats of artificial intelligence systems, addressing potentials and threats at the same time.

1.2 Project Design

The project "Learning Progression Analytics - Analyzing and Fostering Learning for the Development of Competence" (LPA-AFLEK) aimed at individualising students' learning for better addressing students' needs by real-time analysis of student answers through artificial intelligence systems. Physics teachers in Northern Germany enacted different instructional units which included a digital learning environment and workbook with their students from 7th or 8th grade in their regular classes on the topic of energy over the period of five lessons of 90 minutes each. Teachers participated in programmes of professional development of about three hours and an individual coaching of about one hour. For the learning unit, the teachers could choose between the enactment of two different units with respectively two different topics, laptops or solar cells. Each of the instructional units was designed following a project based learning approach (Fischer, 2022; Krajcik & Blumenfeld, 2005), having on guiding question that aimed at motivating the students. The guiding question was introduced in the first lesson. Then, the guiding question was worked on through three experiments guided by one sub-question each. In the fifth and final lesson, the guiding question was answered by the students.

Two larger phases of data collection were carried out, one in 2022 and one in 2023. In the first phase of 2022, the students worked in a digital learning environment and the teachers did not yet have an automatic evaluation or feedback option. In the second phase in 2023, the digital learning environment was extended by an artificial intelligence system (Gombert et al., 2022) which provided teachers with real-time evaluations of the student answers through a dashboard (Karademir et al., 2024) and an option to provide individualised and direct feedbacks to the students. In order to evaluate the teachers' perspectives on various aspects of their work with the digital learning environment and the dashboard, we conducted three interviews of 90 minutes each, one after each lesson with a set of epistemic activities which included an experiment. The overview of the project can be seen in Figure 1-1.

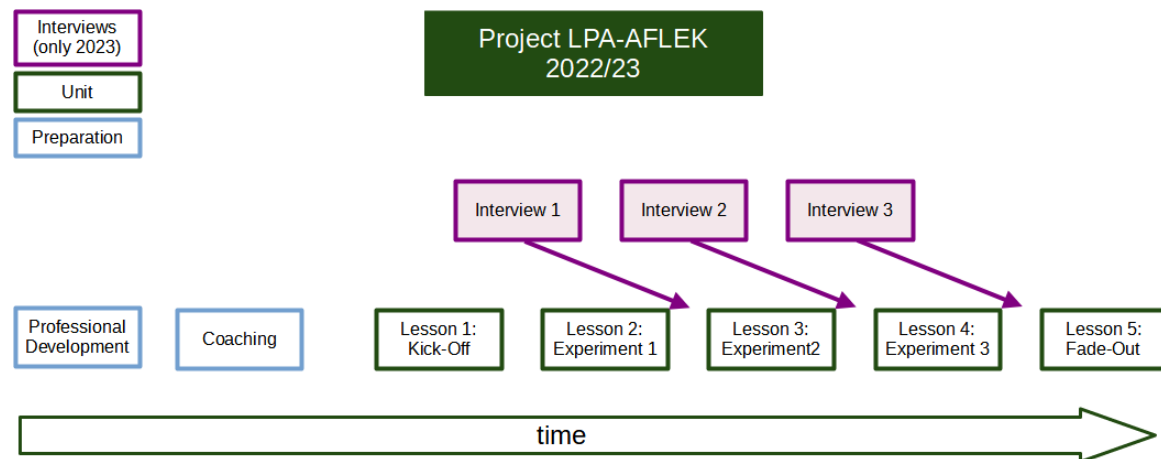


Figure 1-1 – Project: Overview

In order to be able to analyse student answers by means of an artificial intelligence system, evidence-centred design was used. More specifically, we built upon a student competence model to formulate evidence statements for each task. These statements delineated what would serve as evidence that students had demonstrated the knowledge, skill, and/or learning performance that they were expected to demonstrate in an answer to the task in question. Most of the tasks out of the 31 (Laptops) and 36 (Solar Cells) tasks were free text answers. In one task, there could be one or more labels of competence that we scored. Based on the evidentiary statements, all student answers were scored by human raters with respect to competence, namely the knowledge, skills and learning performances demonstrated. With the students' answers and the scores from the first phase, we trained the artificial intelligence system for the second phase. You can see an exemplary item with a student answer and a score in Figure 1-2.⁴

Figure 1-2 - Project: Scoring of Student Answers

⁴ The evaluative scores are only visible in the figure for better understandability in order to see with which information the artificial intelligence systems were trained– the students could not see these evaluations.

1 Introduction

The teacher-facing dashboard displayed the competence evaluation with the opportunity to provide feedback to individual students or groups of students. There were visualisations on a task level, on the level of a specific knowledge element, but also as an overview for all students over all tasks and competence elements. The example of a class overview with four possible competence levels (red, orange, yellow, and green) is shown in Figure 1-3.



Figure 1-3 - Project - Teacher Facing Dashboard

1.3 Overview of the Four Pieces of Scholarship

In our first piece of scholarship, we⁵ addressed the lack of a theoretical framework that connects the research on bias in artificial intelligence systems with STEM identity development. We already knew that artificial intelligence systems in education come with the threat of being biased (Baker & Hawn, 2021). That bias is relevant to addressing inequalities because existing inequalities are known to operate at the level of identity development (Carlone & Johnson, 2007; Çolakoğlu et al., 2023; Kayumova & Dou, 2022) with receiving recognition being one central pillar for successful STEM identity development. Precisely that recognition can (or not) be provided by artificial intelligence systems with the risk of bias, under-serving the same students with recognition who already face historically grown inequalities. In our piece of scholarship, we proposed a theoretical model that connects STEM identity development of under-served students with bias through the mechanisms of vulnerability and iterability (Butler, 2005). Building on that theoretical framework, we explored two normative issues and deduced six suppositions. These suppositions can inform political decision-making as well as future research in the context of artificial intelligence systems in physics education that aims at addressing issues of social justice.

In our second piece of scholarship, we addressed the gap between existing principles to prevent bias (Cerratto Pargman & McGrath, 2021; D'Ignazio & Klein, 2020) and the lack of impact of these principles on practice (Kitto & Knight, 2019). Kitto and Knight found that one

⁵ I do refer to “we” and “our” in order to make the work of colleagues visible who have worked together with me wherever applicable. Only if I want to make my own standpoint explicitly visible, I refer to “I” and “my”. In any case and as this is my dissertation, I take the full responsibility for any inaccuracy and standpoint and by no means aim at handing over the responsibility to another person for what is written here. Also, the “we” and “our” can refer to different groups of persons. If I had to specify each time, that would have been a lot more text without specific relevance in the text. The lists of authors with the respective contributions shed light on the various contributions for the scientific publications at least. I aim at showing science as the team work it is by highlighting where I worked together with colleagues as none of my four publications was written by me alone.

reason for the lack of impact on practice is the under-specification of principles. Hence, we identified where exactly existing principles fail to provide clear guidance and proposed a process of refinement. We highlighted that principles need to be rooted in explicit normative standpoints, a domain-specific⁶ analysis of inequalities, and to define the most relevant diversity dimensions as well as evaluation criteria based on the domain-specific analysis of inequalities. The process that we proposed can be used for other domains as well. Our contribution highlights the need to work towards domain-specific regulation on the one hand in order to reach impact on practice and to address intersectional discrimination through additional measures on the other hand. The additional measures against intersectional discrimination seemed necessary to us because preventing intersectional biases in all cases would come with huge additional workload and thereby may lead to an under-use of artificial intelligence systems in physics education. At the same time, intersectional discrimination is not acceptable from the normative point of departure and therefore demands for counter-measures where counter-measures are most efficient.

In our third piece of scholarship and first empirical study, we aimed at contributing empirical evidence to De-Biasing of artificial intelligence systems in physics education. Our goal was to do so in a politically relevant and actionable way in the context of the European Union where regulation is based on a risk analyses and higher requirements are imposed to areas of higher risk (*AI Act*, 2021). We focused on the regulation of training datasets as historically grown inequalities are known to reproduce themselves as bias in artificial intelligence systems through datasets (Baker & Hawn, 2021; Gardner et al., 2019; Latif et al., 2023). We used the student answers and scores from both phases, including 16 teachers who conducted the units in 22 classes and 527 students in total. Six of the classes learned at a so-called “Gemeinschaftsschule” while 16 classes learned at a “Gymnasium”. We found first indications that bias could be reduced but not eliminated through the regulation of training datasets. Instead, some bias tends to remain in the systems even if De-Biasing practices are in place.

Precisely that finding of remaining biases lead to our fourth piece of scholarship: If some bias remains even after De-Biasing, what else can be done in order to prevent a decrease in equity and instead design STEM education where many worlds fit? We turned towards central actors for the STEM identity development of students: teachers. Teachers’ recognition can be a powerful tool to strengthen STEM identity development, especially STEM identity development of under-served students (Carlone & Johnson, 2007; Çolakoğlu et al., 2023). One way to prepare teachers for their contributions towards social justice in education that is well in line with the normative framework of the pluriverse that we built upon in our work is supporting teachers in their development of Critical Consciousness. Critical Consciousness was developed in the Americas and is a well-established field of research and practice (Freire, 1970; hooks, 1994, 2009; Jemal, 2017). However, it remains unclear how applicable existing findings are in the different cultural setting of Northern Europe and the context of artificial intelligence systems. We conducted interviews in the second phase of the data collection in 2023. The relevant interviews for the study were the interviews 1 and 3. We conducted interviews with a total of seven teachers, ending up with 14 interviews in total. The relevant questions for our study did make up around 15 minutes of the total of 90 minutes and were conducted at the end of the interview, after the teachers had already answered questions to the use of the dashboard as well as the resources they

⁶ A “domain” can for example be physics education or – slightly broader – natural science education – in contrast to other domains such as (English) language education.

1 Introduction

had drawn upon in the facilitation of the learning units. Our analyses revealed for all teachers in the sample a certain concern to act non-discriminatory but also many dimensions for improvements. Most potential for improvement was identified for the critical understanding of feminist pedagogical thought, the critical attitude of education as the practice of freedom, and the critical action of reflection. Due to the qualitative character of our study, we cannot draw final conclusions that could inform policy-making but can only report first indications that future research can draw upon. Our indications highlight the potential of Critical Consciousness of physics teachers in Northern Europe to strengthen the STEM identity development of under-served students.

Modern education systems aim to enable all students to participate in society and allow for social mobility. Artificial intelligence systems come with promising potentials to increase participation in STEM education. At the same time, precisely these artificial intelligence systems which enable participation for some students come with the threat of creating a barrier for other students. If we want to invite all students to participate in STEM education, these threats need to be addressed. As De-Biasing seems to have some but in total limited potential, additional structural intervention is needed. Teachers are central actors in the STEM identity development of students due to the recognition they can provide, hence teachers Critical Consciousness holds great potential for such a structural intervention. Ultimately, De-Biasing and Critical Consciousness can be important pieces of STEM education ecosystems which are 'a world where many worlds fit' (Escobar, 2017; Kayumova & Dou, 2022).

Orchesterklang

*Du bist ein wachsender Baum:
Ganz klar scheinst Du aus Deiner Umgebung zu trennen,
Ich meine deutlich Deine Konturen zu sehen,
Doch weiß ich tief in mir, dass kaum
Ein Baum ohne Wiese grünend, Rauschend ohne wehen-
Den Wind oder gar ohne Vöglein vorzustellen
Ich mir mag – denn erkennen
Kann ich Dich nur mit der Umgebung Wellen:*

*Denn ich, das Meer, weiß um der
Wechselwirkung Einfluss
Auf die Identität der Dinge –
Jede Begegnung ist ein Kuss
Von der Seele der Gebenden her,
Auf dass er bei den Beschenkten erklinge
In der Lebensmelodie, die wie Orchesterklang
Nur als Ganzes zu den Lauschenden drang
Und überhaupt dringen kann.*

*Gemeinsam nur sind wir begreifbar,
Zusammen nur ist unser Bild zu malen,
Und was in Form von Zahlen
Erscheint ganz logisch und klar
Ist in Wirklichkeit ein sich einander Prägen
Auf endlos verwobenen und schönen –
Begleitet von laut´ und leisen Tönen –
Sich ständig neu erfindenden Wegen,
Ein durch geteilte Zeit sich öffnender Raum,
Ein Wir, das – verglichen mit Du und ich – ist so viel mehr.*

2 Theory

2.1 STEM Education for all

2.1.1 STEM Education for all and Historically Grown Inequalities in Physics Education

Modern education systems such as the ones in the European Union are envisioned to enable all students to participate in society (*EU Charter*, 2012). Participation does not mean that all students will acquire powerful positions in the economy. Instead, equal opportunities and individual success in education are the fundament of justification for accepted inequalities in a society. In Germany, meritocracy is even defended at the level of constitution in the article § 33 of the constitution by binding public hiring processes to criteria-based selection processes (*Grundgesetz*, 2022). While there are accepted inequalities based on meritocracy, there are inequalities that are not accepted as well. Next to the task of enabling students' participation in society for example, modern education systems have the task to enable social mobility (*EU Charter*, 2012; Muñoz Izquierdo, 2012). Students' individual success in education should not depend on their positions⁷ on gender, race, socio-economic status, religion and beliefs, ability and chronic illnesses, and sexual identity. However, participation in physics-related fields of education is not equal. For example, in engineering, manufacturing and construction in the European Union 72.7 % of all students were male (Eurostat, 2022)⁸. Interestingly, the inequalities cannot be explained by differences in examination results in secondary education. The inequalities are rooted in who does (not) identify with, in our case, physics.

In Germany, socio-economic status is the strongest predictor for starting an academic career or not. While 79 out of 100 children start an academic career when their legal guardians studied, only 27 out of 100 children do so if none of their legal guardians studied (El-Mafaalani, 2021, pp. 66–67; Kracke et al., 2018, pp. 5–6). Students' career aspirations and choices in physics-related areas have been shown to be directly related to their socio-economic status (Archer et al., 2015, p. 939; Avraamidou, 2019, p. 318). The same holds true for other dimensions of diversity: For example in Germany, only 21 % of physics bachelor and 18 % of physics master programme graduates were women (Düchs & Ingold, 2018, p. 36). Inequalities in physics education along dimensions of diversity cannot be explained without considering structural discrimination along dimensions of diversity (Saini, 2017, 2019). In other words: Modern education systems currently fail to live up to the promise of enabling all students to participate in society – an intervention towards social justice is necessary. Social justice does not mean that all students obtain the same educational outcomes- it rather refers to educational outcomes being un-correlated with positions on dimensions of diversity (OECD, 2018, p. 13). When designing interventions against structural discrimination, it is crucial to address the structural discrimination at the level where it operates. For physics education, it has been shown that the high levels of inequalities cannot be explained by differences in the development of competence (OECD, 2016). Instead, the recognition students (do not) receive plays a key role in their development of

⁷ A “position” or “positionality” is what refers to a concrete manifestation for a diversity dimension. On the diversity dimension of gender, a person can position as non-binary, female, or male, for example.

⁸ The available data is binary gendered, information for non-binary students is not even available.

STEM identities and STEM career aspirations (Carlone & Johnson, 2007; Dou et al., 2019; Godwin, 2016; Mujtaba & Reiss, 2013, p. 1824).

STEM identity is a well-established construct within the research fields of science education (Archer et al., 2022; Carlone & Johnson, 2007; Dou et al., 2019). Identity defines who we are and identities are developed through experiences (Brickhouse, 2001). In the process of identity development in out-of-school-contexts, role models and recognition play a key role (Çolakoğlu et al., 2023). For our field of physics education, various historically grown inequalities exist with a high relevance for identity development. For example, characteristics of material are mainly named after *white* men (Götschel, 2015, p. 209). Seeing mainly *white* men in the field of physics can lead students to perceive physics as a field for *white* men and thereby hinder the identity development of students who are already under-served in terms of opportunities for STEM identity development. Precisely this process of strengthened or hindered identity development is one of the processes that lead to the tendency of historically grown inequalities to reproduce themselves without structural counter-measures. In order to break the vicious cycle of reproduction of historically grown inequalities, a focus on identity and, for example, recognition by significant others is needed. When addressing the historically grown inequalities in physics education, normative questions emerge. For example: The inequalities along which dimensions are analysed? Do we seek to avoid additional discrimination only or do we aim at systematically reducing historically grown inequalities?

2.1.2 The Need for a more Concrete Justice Theory

When engaging in questions of justice from a scientific perspective, a concrete and context-relevant definition of justice is needed as a theoretical framework. In the context of physics education, historically grown inequalities exist along dimensions of diversity which can only be explained when structural discrimination is considered. A justice theory therefore needs to be based on an analysis along dimensions of diversity. Next to the definition of the justice for whom, the justice of what needs to be defined more clearly than “STEM education for all” in order to have all relevant boundary conditions defined enough to engage in scientific investigation. Given the historically grown inequalities in physics education, STEM education for all means to address questions of justice at the level where the inequalities operate, at the level of STEM identity development.

In order to address concrete dimensions with clear objectives from a scientific perspective, the matrix of domination was introduced by Black feminist scientist Patricia Hill Collins (1990). The matrix of domination is a concept to explain existing inequalities and to describe lines⁹ along which inequalities are distributed, namely the diversity dimensions. Following the concept of the matrix of domination, striving for social justice along diversity dimensions means to explicitly analyse power relations along these diversity dimensions. Collins proposed race, social class, and gender as diversity dimensions. In the local context of schools in Germany and anti-discrimination law, religion and beliefs, ability, and sexual identity are relevant dimensions of diversity as well. In terms of clear objectives, Patricia Hill Collins frames distributions along these dimensions that do not represent the actual shares of the population as injustice which should not exist. In other words: Students of all positionalities on diversity dimensions should be present in STEM education as they are in the society as a whole. For example, women should make up 51.1 %, their actual

⁹ Such a line can, for example, be drawn on the dimension of gender between the privileged position of male students and the positions that face discrimination, for example female and non-binary students. The line distinguishes a privileged from a discriminated position on a specific dimension.

share in the EU population (Eurostat, 2023), of the students in engineering, manufacturing and construction in the European Union instead of the current 27.3 %.

One central characteristic of diversity dimensions is their interlocked character, their intersectionality (Crenshaw, 1989). Even though intersectionality was already discussed before, for example by women's rights activist Sojourner Truth in 19th century or the Combahee River Collective, Crenshaw coined the term intersectionality in 1989. According to intersectionality, we need to understand the diversity dimensions in the matrix of domination as interlocking systems (Costanza-Chock, 2020; Crenshaw, 1989). This means that belonging to more than one group that is discriminated against leads not only to the sum of discriminations of the single dimensions, but to entirely new forms of discrimination.¹⁰ Consequently, when investigating discrimination for one diversity dimension, the other dimensions need to be considered as well in order to address the full scope of the discrimination that individuals who belong to that one diversity dimension might face. The concept of intersectionality is therefore incredibly important as it connects the anti-discrimination work from various movements and schools of scientific thought. Summarising, a concrete justice theory is needed in order to scientifically engage in questions of justice in physics education that is 1) addressing STEM identity development, 2) focusing on the matrix of domination and diversity dimensions, and 3) recognising the intersectionality of the diversity dimensions within the matrix of domination.

2.1.3 Justice Theory and Focus: The Pluriverse and STEM Identity Development

A justice theory that fulfils the necessary criteria is the pluriverse: 1) Connections to STEM identity development have already been developed and it is rooted in 2) the matrix of domination and 3) intersectional perspectives (Kayumova & Dou, 2022). The pluriverse fits well particularly due to two of its characteristics: a concrete vision combined with clear allocation of responsibility for transformation.

Firstly, the pluriverse provides the vision "a world where many worlds fit" (Escobar, 2017; Kayumova & Dou, 2022; Mignolo, 2007). This vision is rooted in diversity dimensions. As a concept, pluriverse comes with a normative focal point being the standpoint that distributions should not vary over diversity dimensions if these variations lead to discriminatory distributions of, for example, power. In other words: Under-representation in physics-related careers of, for example, women needs to be reduced. To stress that: Reducing under-representation does not mean to neglect the free choice of all students to opt for whatever career they want to. The need to reduce under-representation is founded in an analysis that the current level of inequalities can only be explained when considering structural discrimination. The structural discrimination needs to be reduced in order to enable all students to freely choose a career which will ultimately lead to a reduction of under-representation. Through that problematisation of under-representation, the pluriverse does not only fulfil the three criteria that are necessary to address the inequalities in physics education at the level of their operation, but also provides a clear definition of a goal that can serve as a boundary condition for evaluation in scientific work.

¹⁰ For example, a muslim woman does not only face the sum of the discrimination that muslim men and non-muslim women face. Instead, the combination of being both, muslim and a woman, leads to entirely new forms of discrimination. A muslim woman wearing a hijab might be questioned in her feminist identity – something that would neither happen to a muslim man (he does not wear a hijab) or a non-muslim woman (she does not wear a hijab). However, a muslim woman can wear a hijab and be feminist. The additional questioning of her identity is one example of discrimination that leads to additional stress and workload. It is one example of intersectional discrimination.

Secondly, the pluriverse allocates responsibility not with the under-represented individuals who face discrimination but with the structure in which the individuals live and which should be transformed. This is particular important as there is a lot of research that exposes focusing on the under-represented individuals as deficit-oriented approach which itself reproduces discrimination and inequalities instead of enabling a transformation towards justice (Cheuk, 2021, pp. 8–9).¹¹ Instead of fixing the individuum, the focus in a pluriverse approach is on adjusting the system so that diverse identities can co-exist within it. In our context of artificial intelligence systems that translates, for example, into: We do not investigate how students who face biased algorithms can learn how to cope with them. Instead, we investigate how algorithms can be De-Biased so that the system changes in a way that protects the students from discrimination.

The approach for our work is to reduce under-representations of positionalities on diversity dimensions, so-called “diversity mainstreaming”. We focus on STEM identity development, the level of operation of the inequalities in physics education, and assigning the responsibility with the system. In Figure 2-1, the conceptualisation of the pluriverse with its core elements is shown: The reproduction of historically grown inequalities in physics education that need structural intervention in order to break out from the vicious cycle of reproduction, together with the two central points, a focus on inequalities along diversity dimensions and the assignment of responsibility with the system, all together standing under the vision of the pluriverse – a world where many worlds fit. Our research was conducted in the context of a digital learning environment with automatic evaluation of student answers in real time by artificial intelligence systems. As recognition of students based on competence evaluations plays in key role in the STEM identity development of the students, the issue of potential bias and its influences is at the centre of my dissertation. Concretely, the focus is on avoiding bias through preventive actions, so-called De-Biasing.

¹¹ There is a lot of work that heavily criticises putting specific anti-discrimination work or types of work on those who have to face discrimination already (Gago, 2019, pp. 24–28), which was for example condensed into the phrase “You make me do too much labour” (Paloma, 2023).

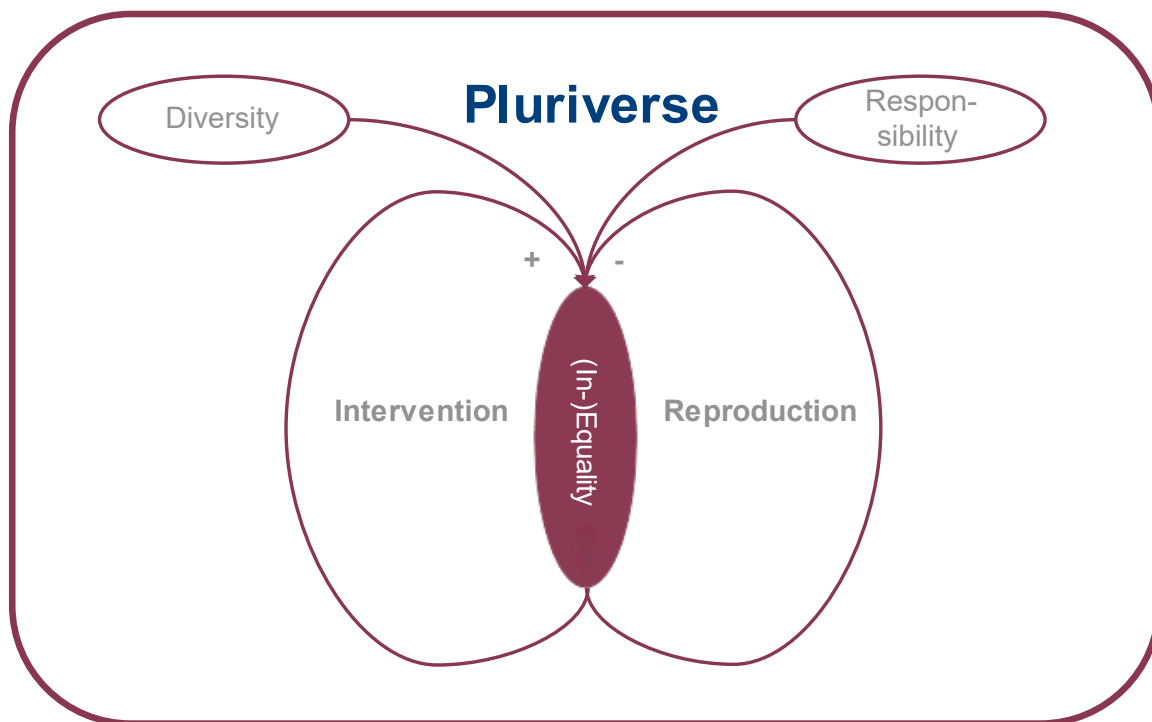


Figure 2-1 - Pluriverse

2.2 De-Biasing and Critical Consciousness

De-Biasing comes with the question: De-Biasing of what? I chose to write about De-Biasing *artificial intelligence systems* within my dissertation. Some authors call to use other terms such as automated decision-making systems instead of artificial intelligence because non-experts tend to fail to understand what artificial intelligence means and even assign unrealistic, almost magic capabilities to artificial intelligence (Matzat et al., 2019, p. 8). The advantage of automated decision-making systems is that it explicitly includes the decision-making model with its algorithm, the datasets, and the entire political parts of the decision. Artificial intelligence system instead focuses on the concrete type of algorithms that are used, thereby obscuring the decisions and political aspects by focusing on technology. We prioritised the technical precision of the term artificial intelligence systems. As the different pieces of scholarship were conducted at different points of time and with different authors contributing, the use of language differs in the papers. No matter which term we used, we always aimed at both, being technically precise and making political decisions explicit.

The first knowledge gap we sought to address was the lack of an explicit theoretical model of how bias in artificial intelligence systems would negatively impact the STEM identity development of under-served students. We already knew that bias is an issue for artificial intelligence systems (Baker & Hawn, 2021; Cheuk, 2021). Additionally, bias often is described to have its origin in inequalities in datasets plus we know about many existing inequalities in physics education (Gardner et al., 2019; Latif et al., 2023; Suresh & Gutttag, 2021). A focus on STEM identity development instead of a focus on competence evaluation only is needed in order to address the historically grown inequalities at the level where they operate. In our first piece of scholarship (Chapter 3 Responsible Learning Analytics and STEM Identities), we aimed at bridging that gap by developing a theoretical model that describes how bias in artificial intelligence system would negatively impact STEM identity development of under-served students. We brought together theoretical contributions from research fields of bias, STEM identity development, and mechanisms of discrimination.

Building on that model, we focused on two normative issues and deduced six suppositions for bias research in the future.

Having connected research on bias and STEM identity development, the second knowledge gap we aimed to address was where exactly existing approaches to De-Biasing fail. We were not the first ones to address De-Biasing of artificial intelligence systems. However, it has been shown that even though guiding principles often are in place (Cerratto Pargman & McGrath, 2021; D'Ignazio & Klein, 2020; Pardo & Siemens, 2014), the impact on practice of designers of artificial intelligence systems in education is little (Kitto & Knight, 2019). One problem that was identified was the under-specification of De-Biasing tasks (Kitto & Knight, 2019, p. 2864). In our second piece of scholarship (Chapter 4 Equity-Focused Decision-Making Lacks Guidance!), we aimed at describing where guidance needs to be more specific and through which processes guidance can be made more specific for a concrete domain and usage scenario.

Building on the theoretical model and the concrete description of which evidence is needed, we sought to address the third knowledge gap of how artificial intelligence systems can effectively be De-Biased. We knew that training datasets and their compositions are both, known to be possible entry points for bias as well as possible to regulate and therefore actionable from a political perspective (Baker & Hawn, 2021; Gardner et al., 2019; Gebru et al., 2018). With our third piece of scholarship (Chapter 5 De-Biasing), we aimed at contributing empirical evidence in order to inform political decision-making on how to regulate the composition of training datasets in a way that is efficient and effective without putting too much requirements on artificial intelligence systems and thereby leading the under-use of artificial intelligence systems in education.

The results from our three pieces of scholarship as well as the literature indicate that De-Biasing is not going to work for a 100 % of all cases. Artificial intelligence systems are not going to be bias-free – instead, bias can be prevented only for some cases effectively and efficiently when potentials of artificial intelligence systems for physics education shall be harvested as well. From a pluriversal perspective on STEM identity development of under-served students, the finding of De-Biased instead of bias-free artificial intelligence systems leads to the question: If we cannot successfully De-Bias our systems completely, what else can be done in order to reach a pluriverse? As STEM identity development highly depends on recognition by significant others, we turned to the potentials of teachers in our context.

In order to prepare teachers for their work of contributing to the reduction of discrimination, various approaches have been developed (Götschel, 2015; Mecheril et al., 2020; Tißberger, 2017). One approach that is well in line with our normative framework of the pluriverse and which has been developed in the Americas as well over decades is Critical Consciousness (Freire, 1970; hooks, 1994, 2009; Jemal, 2017). However, it remained unclear to what extent Critical Consciousness of teachers is present in teachers in the different cultural setting of Northern Europe and in the context of artificial intelligence systems. It seemed necessary to us to understand how applicable findings from the research fields in the Americas are in our context. In our fourth piece of scholarship (Chapter 6 Critical Consciousness), we aimed at developing a qualitative coding manual and exploring the Critical Consciousness of physics teachers in Northern Europe in great qualitative depth.

2.3 Overarching Research Questions

STEM identity development of under-served students takes place (or not) within the context of recognition of both, artificial intelligence systems and teachers. Hence, findings on De-Biasing have an impact on Critical Consciousness and the other way around as well. Therefore, I aimed at discussing the findings of all four pieces of scholarship with its relevance to each of the constructs of De-Biasing and Critical Consciousness asking these two questions:

- How can De-Biasing contribute to structurally address existing inequalities in physics education in the context of the rising use of artificial intelligence systems in Northern Europe?
- How can Critical Consciousness contribute to structurally address existing inequalities in physics education in the context of the rising use of artificial intelligence systems in Northern Europe?

After reflecting on the two constructs separately, I aimed at zooming out a bit more in order to see the entire picture of STEM identity development within a pluriverse. The idea was to break out of the limited perspectives that each construct itself can offer and thereby also to be able to discuss issues beyond the single constructs. I asked:

- What are the promises and potential downfalls of addressing historically grown inequalities in physics education in Northern Europe through Critical Consciousness and De-Biasing for STEM identity development for all students in a pluriverse?

Empathie

*Ich habe Zeit Dir zuzuhören,
Gebe wieder, was Du sagst,
Empathisch topfschlagend versuche ich zu Höh´rem
Verständnis zu gelangen,
Mitschwingendes einzufangen,
Du kannst reden, solange Du magst.*

*Gemeinsam suchen wir zu benennen:
Was brauchst Du? Was wünschst Du Dir?
Das wollen wir erkennen!
Und Stück für Stück entwickle ich
Mehr und mehr Verständnis für Dich,
Wir schaffen Verbindung im Jetzt und Hier!*

*Aufmerksamkeit und Verbundenheit,
Sie klären die Sicht,
Wo vorher viel Nebel war, weit und breit,
Ach, wie schön, dass Du mir zu erzählen wagst!,
Dass so viel Vertrauen prägt unser Wir!,
Sodass erstes Licht ins Dunkel bricht!*

3 Responsible Learning Analytics and STEM Identities

Title. Positioning responsible learning analytics in the context of STEM identities of under-served students

Abstract. Addressing 21st century challenges, professionals competent in science, technology, engineering, and mathematics (STEM) will be indispensable. A stronger individualisation of STEM learning environments is commonly considered a means to help more students develop the envisioned level of competence. However, research suggests that career aspirations are not only dependent on competence but also on STEM identity development. STEM identity development is relevant for all students, but particularly relevant for already under-served students. Focusing solely on the development of competence in the individualisation of STEM learning environments is not only harming the goal of educating enough professionals competent in STEM, but may also create further discrimination against those students already under-served in STEM education. One contemporary approach for individualisation of learning environments is learning analytics. Learning analytics are known to come with the threat of the reproduction of historically grown inequalities. In the research field, responsible learning analytics were introduced to navigate between potentials and threats. In this paper, we propose a theoretical framework that expands responsible learning analytics by the context of STEM identity development with a focus on under-served students. We discuss two major issues and deduce six suppositions aimed at guiding the use of as well as future research on the use of learning analytics in STEM education. Our work can inform political decision making on how to regulate learning analytics in STEM education to help providing a fair chance for the development of STEM identities for all students.

Published. Grimm, A., Steegh, A., Çolakoğlu, J., Kubsch, M., & Neumann, K. (2023). Positioning responsible learning analytics in the context of STEM identities of under-served students. *Frontiers in Education*, 7. <https://doi.org/10.3389/feduc.2022.1082748>

3.1 Introduction

Addressing the challenges of the 21st century, professionals competent in science, technology, engineering, and mathematics (STEM) will be indispensable (FEANI, 2021, pp. 7–8). A stronger individualisation of STEM learning environments is commonly considered a means to help more students develop the envisioned level of competence (National Academies of Sciences, Engineering, and Medicine, 2019). However, research suggests that even the most competent students may not aspire to a STEM career (Taskinen et al., 2013). One reason is that STEM career aspirations do not only depend on students' STEM competence but also on students' developing a STEM identity (Dou et al., 2019, p. 623). Carlone and Johnson (2007) accordingly identify, in addition to the dimension of competence, two more dimensions relevant to the development of a STEM identity: recognition and performance. As we will show, these additional dimensions are specifically relevant to under-served students who face historically grown inequalities due to two mechanisms of discrimination, vulnerability and iterability (Butler, 1990, 2005; Hartmann and Schriever, 2022). Vulnerability is the mechanism that describes that the same situation can have different effects for diverse students. Iterability describes the setting of norms through repetition. The development of a STEM identity is not just a question of preparing enough professionals sufficiently competent in STEM but also of justice and power (see also UN-SDG-Goal 4, 2015). STEM education researchers Waight et al. (2022, p. 19) even advocate for centering equitable perspectives in science education overall. We therefore argue that focusing solely on the development of competence in the individualisation of STEM learning environments is not only harming the goal of educating enough professionals competent in STEM, but may also create further discrimination against those students already under-served in STEM education.

One contemporary approach for individualisation of learning environments is learning analytics. Learning analytics is referring to the collection, analysis and reporting of data about learning and the environment in which learning occurs with the purpose to understand and optimise learning and the learning environment (Society for Learning Analytics Research (SoLAR), 2022). One example for the use of learning analytics is the provision of feedbacks for teachers on the competence development of students. Learning analytics often draw on machine learning techniques; that is, algorithms trained on existing data to monitor students' competence development. However, existing data often reflect historically grown inequalities. Training algorithms with existing data will then lead to the reproduction and, worse, reinforcement of these inequalities. That is, learning analytics would under-serve precisely those students again who already face historically grown inequalities. In order to navigate between the potentials and threats in the field of learning analytics, the concept “responsible learning analytics” has been introduced (Prinsloo and Slade, 2018; Cerratto Pargman et al., 2021). To date, responsible learning analytics are focused on competence development in learning environments, neglecting the relevance of the other dimensions of STEM identity development. However, STEM learning environments often are opportunities for the development of STEM identity as well and multiple historically grown inequalities exist with regard to identity development in STEM (Mujtaba and Reiss, 2013; Rosa and Moore Mensah, 2016; Avraamidou, 2019). As a result, it seems imperative that the concept of responsible learning analytics must be expanded to include STEM identity development as one important aim of individualising STEM learning environments.

In this paper, we propose a theoretical framework that positions responsible learning analytics in the context of STEM identity development, especially of under-served students. We identify two major issues of responsible learning analytics in the context of

under-served students' STEM identity development and derive suppositions to guide future work in this area. Our suppositions are meant to highlight the need for making normative decisions. In doing so, we intend to make normative decisions visible and debatable. The suppositions can guide the use of learning analytics and future research on the use of learning analytics. Our work can inform political decision making on how to regulate learning analytics in STEM education to help providing a fair chance for the development of STEM identities for all students, particularly students from under-served groups. In summary, our paper adds:

- A theoretical framework that positions responsible learning analytics in the context of STEM identities of under- served students,
- A discussion of two major issues that come with learning analytics in the context of STEM identity development, and
- Six suppositions aimed at guiding the use of as well as future research on the use of learning analytics in STEM education.

3.2 STEM identities of under-served students

The concept of identity has been introduced to the field of STEM research with the purpose to help understand why students engage in STEM, how some students are promoted whereas others are marginalized by current STEM education practice and hence a means to work towards more equitable STEM education (Carlone and Johnson, 2007).

3.2.1 STEM identities

Identities are complex constructs through which we bring our experiences and our reflective projections together, define who we are and what influences our learning (Brickhouse, 2001). Identities are shaped in social interaction and new identities need to be negotiated with regard to existing identities which can lead to conflicts (Brown, 2004, p. 811). STEM identities lead to higher career aspirations through “goal setting and behavior” (Dou et al., 2019, p. 632). For STEM identity, various frameworks exist. For our work, we were looking for an understanding of STEM identity that includes clearly operationalized dimensions and decided for the framework proposed by Carlone and Johnson (2007). According to Carlone and Johnson (2007), STEM identity can be understood by its three dimensions: recognition, performance, and competence. Comparable frameworks propose interest as another dimension of STEM identity (Godwin, 2016; Hazari et al., 2020; Mahadeo et al., 2020). While we will argue in the following that learning analytics impact recognition, performance, and competence, we see interest rather impacted by the design of learning opportunities and not by learning analytics.

STEM identity development can create conflicts that emerge from identity negotiations. At the same time, STEM identity development for all students is important. Identity negotiations as well as the role of identities for learning make STEM identity development a relevant context for learning analytics in STEM learning environments. These three dimensions recognition, performance, and competence are a relevant context for learning analytics.

Competence is understood as the empirically testable knowledge, skills, and abilities in a particular domain as well as what an individual says about oneself with regard to the own competence (Carlone and Johnson, 2007, p. 1992).

Performance is what an individual does through concrete actions. The performance definition differs from what often is understood as performance in science education.

Performance as understood by Carlone and Johnson can be based on competence but does not have to be. For example, a person might perform STEM identity by communicating with adequate scientific language in a specific task without a profound understanding of the concepts and thus competence. The repetitive performance of STEM identity can “become patterned and habitual” (Carlone and Johnson, 2007, pp. 1190–1,192). A development opportunity for STEM identity can include or exclude students with different performances of their existing identities, for example their gender or social class performance.

Recognition is considered particularly relevant for the development of STEM identity. While competence and performance are components of STEM identity, STEM identity development is dependent on the recognition of “meaningful others” (Carlone and Johnson, 2007, p. 1992). Meaningful others are persons whose acceptance matters in the context of STEM, for example STEM teachers. In order for STEM identity to become habitual, recognition is a key. This holds particularly true for students who face historically grown inequalities (Carlone and Johnson, 2007, pp. 1887–1992).

3.2.2 Under-served students

STEM identity development is particularly relevant for under- served students who face historically grown inequalities. In order to define who we refer to as under-served students and which inequalities we focus on, we take orientation but do not limit ourselves to so-called protected categories, for example gender or race. Protected categories are identity markers based on which a person should not be discriminated against. There are various words to describe students who face inequalities due to an identity in a protected category but we specifically choose the term “under- served.” From our perspective, the term under-served offers two important features: the normative direction and the allocation of responsibility.

The normative direction points at the students not being served enough, there should be more opportunity and offer. As an example, the OECD (2018) states “that differences in students’ outcomes are unrelated to their background” (p. 13). In this example, the OECD explicitly names the categories socio- economic status, gender, or immigrant and family background. In addition, the OECD offers an analysis of the reality: “There is no country in the world that can yet claim to have entirely eliminated socio-economic inequalities in education” (OECD, 2018, p. 13). The above-mentioned categories make up the analysis lens through which reality is analysed. The analysis results in a normative demand: For the named categories, eliminate the relation to students’ outcomes. Under-served students need to be provided with opportunities to reach better outcomes in the future.

Under-served allocates the responsibility for the change of reality not with the students but with the society and its institutions. It is not the students’ fault or responsibility that reality (does not) change. The society with its institutions is responsible for not providing enough or adequate opportunities for change.

Many students are under-served in terms of these opportunities. As an example for gender in Germany, women are currently opting less for a STEM career (Düchs and Ingold, 2018). As STEM identity is important for career aspirations, female students need to get more and better fitting opportunities for their STEM identity development. Another example is racism that has been found to limit STEM identity development (Avraamidou, 2019): As long as students’ racial identities are related to their STEM identities, STEM identity development opportunities for students facing racism need to be strengthened. Under-servements and historically grown inequalities are well-documented, see for example

(Brown, 2004; Carlone and Johnson, 2007; Rosa and Moore Mensah, 2016; OECD, 2018; Bachsleitner et al., 2022).

3.2.3 Intersectionality and individual needs of students

Students can be under-served from the perspective of diverse protected categories, for example gender and race. What if students are under-served from multiple perspectives at the same time, for example gender and race? For students who are under-served from multiple category perspectives, the under-serving is not only the sum of the under-servings for each category. Multiple categories at the same time can lead to additional, distinct under-servings (Costanza-Chock, 2020, p. 17). A Black female student can experience under-servings that go beyond what a white female and a Black male student experience. Black feminist scholar Crenshaw (1989) gave a name to this phenomenon, intersectionality. Intersectionality is not limited to the categories gender and race but holds true for more combinations of categories as well. In order to provide opportunities for the STEM identity development for all students, diverse needs have to be considered. All students have individual needs. However, the needs of under-served students are particularly relevant due to two mechanisms: vulnerability and iterability.

The first mechanism is based on the dependency of humans on recognition of others (Butler, 2005). This dependency is a core need of every human being, a need for recognition (Hartmann and Schriever, 2022, pp. 95–96). Additionally, being recognised is important for students' STEM identity development (Carlone and Johnson, 2007). However, Hartmann and Schriever emphasize that the recognition humans receive differs heavily. For students, receiving limited recognition can hinder their STEM identity development. There is no certain, quantifiable amount of recognition that students need in order to develop a STEM identity. Nonetheless, having received few recognition in the past makes students more dependent on future recognition. This particular dependency on future recognition of under-served students is called vulnerability.

The second mechanism is based on the norms that are established through repetition (Butler, 1990; Hartmann and Schriever, 2022, p. 95). For example, by 2021 214 out of 218 nobel prizes in physics were given to men. If in physics learning environments successful physicists are men again and again, the combination of physicist and man is established as a norm. The iteration of the combined performance of a STEM identity and a male gender identity leads to the norm: Physicists are men. What does this mean for STEM identity development of under-served students? Whether under-served students start performing their STEM identity depends on whether they perceive this new identity performance as fitting to their performances or not (Taconis and Kessels, 2009). Whether the performances fit or not is a norm. This norm can be exclusive by re-iterating existing inequalities. The norm can also be inclusive by serving currently under-served students with STEM identity performances that fit to their performances. The establishment of norms through iteration of performances over and over again is called iterability.

3.2.4 Summary: STEM identities of under-served students

STEM identities of under-served students need a special focus when analysing development opportunities for STEM identity. We summarise our model of STEM identities of under-served students in Figure 3-1.

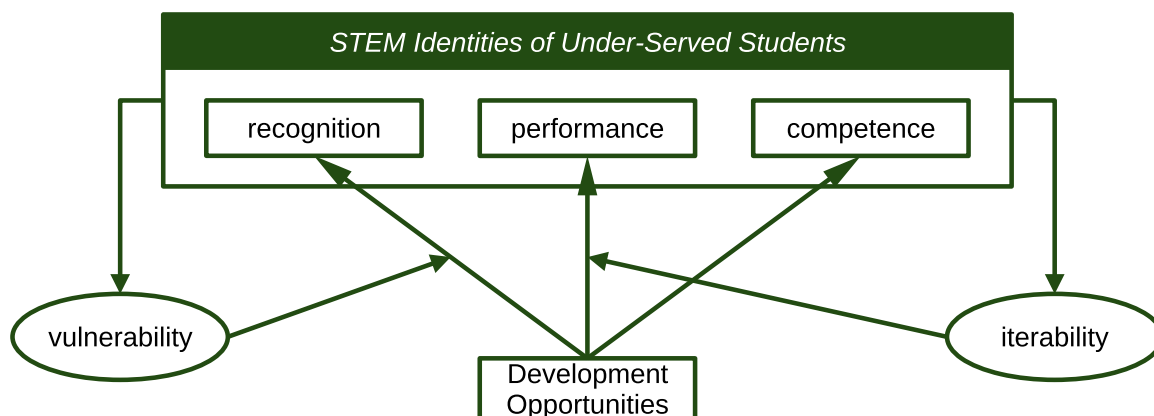


Figure 3-1 - STEM identities of under-served students

STEM identities consist of the three dimensions of recognition, performance, and competence. The needs of under-served students in terms of their development opportunities for STEM identities are very individual and diverse. The vulnerability of under-served students as well as the iterability of their identity performances and whether they (do not) fit in the context of STEM identity moderate the effect of STEM identity development opportunities on the dimensions recognition and performance. In terms of STEM identity development, under-served students have individual needs that often are failed to be addressed.

3.3 Responsible learning analytics

To improve students' competence development, learning environments are individualised (Zhai et al., 2019, p. 1451). One approach to answer individualisation demands are learning analytics. Next to students' needs for individualisation for competence development, students also have individual needs in terms of STEM identity development. We start by discussing learning analytics in the context of competence development. Building on this discussion, we later position learning analytics in the context of STEM identity development.

3.3.1 Learning analytics

Learning analytics are "the measurement, collection, analysis and reporting of data about learners and their contexts, for purposes of understanding and optimising learning and the environments in which it occurs" (Society for Learning Analytics Research (SoLAR), 2022). For example with regard to competence development, students' results on tests can be used for competence diagnosis. Applications of learning analytics are, for example, the provision of individualised and real-time feedback or the empirical analysis of the quality of learning and teaching practices. Learning analytics allow tracking learning trajectories on a task-level and not only through tests before and after learning opportunities. In the Handbook of Learning Analytics, computational analysis techniques and digital data sources are described as "analytics of (1) network structures including actor–actor (social) networks but also actor–artefact networks, (2) processes using methods of sequence analysis, and (3) content using text mining or other techniques of computational artefact analysis" (Hoppe, 2017, p. 23). Learning analytics are most often combined with a focus on institutional strategies and systems perspectives. The learning analytics community has managed to build a huge corpus on equitable perspectives, for example asking questions of power (Wise et al., 2021), questions of geographical coverage around the world (Prinsloo and Kaliisa, 2022), and explicitly demanding for equity-focused research and praxis (Cerratto Pargman and McGrath, 2021).

3.3.2 Learning analytics and responsibility

While learning analytics have great potential for many educational disciplines, they also bring threats that cannot be ignored. These threats are categorized into the under-, over- or miss-use of learning analytics (Floridi et al., 2018, p. 690). Under-use relates to the failure to fully utilise the potential of learning analytics. Over-use is the application of learning analytics in cases where the outcomes do not justify the effort to put learning analytics in place. Miss-use is the application of learning analytics systems in cases that the systems were not made and tested for. The miss-use can, for example, lead to feedbacks that miss-guide decision making. Over- and miss-use are more critical threats than under-use, as they can lead to undesirable outcomes (Kitto and Knight, 2019). In application of learning analytics, the potentials as well as the threats need to be addressed.

In the field of responsible learning analytics, potentials and threats are addressed by a combination of rules and principles to guide the use of learning analytics (Cerratto Pargman et al., 2021, p. 2). Possible potentials are found in a principle that is called the obligation to act – the responsibility to unfold the potentials of learning analytics. The obligation to act addresses the threat of under-use as well as it demands for learning analytics wherever learning analytics may be beneficial. In the context of competence development, learning analytics that lead to higher student outcomes should be used.

Possible threats are addressed by the principle of accountability (Prinsloo and Slade, 2018, p. 3). Staying with the example of competence development, the principle of accountability allocates the responsibility for the threats with the same persons that are responsible for improving the students learning outcomes through the use of learning analytics. Threats in responsible learning analytics in our understanding are also rooted in critical theory which explicitly addresses questions of power and justice (Prinsloo and Slade, 2018, p. 4). This is well in line with the subversive stance on learning analytics as proposed by Wise et al. (2021) as a way of engaging with issues of equity and their interaction with data on learning processes. One example of a threat where issues of equity are relevant is proxy- discrimination (Erden, 2020, p. 85). Proxy-discrimination is the discrimination of a person due to a category that itself is not protected, but related to a protected category. In the example of Erden from the US, an algorithm predicted the duration of staff membership with the category travelling distance to work. Travelling distance itself is not a protected category. However, travelling distance was strongly correlated with race which is a protected category. A responsible implementation of algorithms does not only need to ensure that protected categories are not used for prediction. An analysis of correlations of predictions with protected categories is important as well.

In order to address the potentials and threats, the responsible learning analytics has come up with various general recommendations as well as concrete methods. As general recommendations, there exist for example checklists to address privacy issues (Drachsler and Greller, 2016), policies (Slade, 2016), and principles (Floridi et al., 2018; Phillips et al., 2020). For concrete methods, there exist pre-processing, post-hoc, and direct methods with regard to equity issues (Lohaus et al., 2020). However, the impact of both general recommendations as well as concrete methods for learning analytics practice has been found to be small so far (Kitto and Knight, 2019). Equity issues have been particularly highlighted as a research need in the scientific discourse responsible learning analytics (Cerratto Pargman and McGrath, 2021).

3.3.3 Summary: Responsible learning analytics

In summary, responsible learning analytics acknowledge that learning analytics come with (1) potentials that lead to an obligation to act, and (2) threats that are addressed through accountability. We summarise our understanding of responsible learning analytics in Figure 3-2.

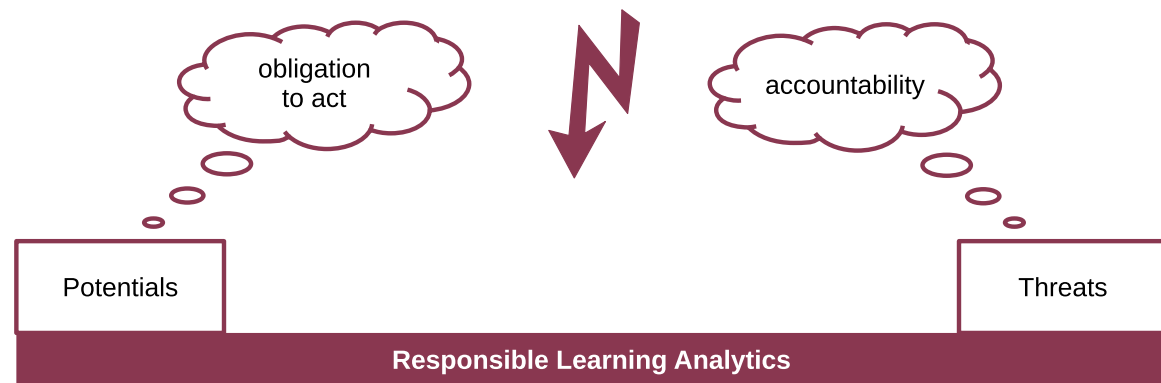


Figure 3-2 - Responsible learning analytics

Responsible learning analytics can be understood as navigating between the obligation to act in terms of potentials and accountability for the threats.

3.4 Responsible learning analytics in the context of STEM identities of under-served students

3.4.1 Proposal of a theoretical framework

Learning analytics are implemented in learning environments to improve competence development. The learning environments are not only competence development opportunities, but also STEM identity development opportunities. Today, historically grown inequalities in STEM identity development exist and some students are under-served in terms of development opportunities in learning environments. Introducing learning analytics in precisely these learning environments adds the known threat of reproducing inequalities through learning analytics systems. In order to be able to address the threat of strengthening inequalities instead of countering them, we propose a theoretical framework that positions responsible learning analytics from Figure 3-2 in the context of STEM identity development of under-served students from Figure 3-1.

Responsible learning analytics' potentials and threats effect development opportunities for STEM identity. We do not see learning analytics as development opportunities themselves because we acknowledge STEM identity as a fairly stable construct. Learning analytics can also be directed at supporting learning on small time scales such as within an instructional task or across a sequence of tasks. STEM development opportunities, however, unfold their effects on the larger scale such as lessons or lesson sets. Learning analytics can support or hinder STEM identity development through a given development opportunity. One example of a development opportunity is a classroom setting and recognition of students through the teacher. Learning analytics can provide competence diagnosis that can trigger teachers to recognise the student in front of other students or even the full class. If learning analytics come with a bias against under-served students, under-served students receive less recognition.

Defining bias in this context is to a huge degree normative. For example, Suresh and Guttag (2021) define an aggregation bias. An aggregation bias is found when a model is trained on full data sets while sub-groups would have needed separate models in order to obtain accurate results. Aggregation bias opens up questions like these: For which sub-groups would the algorithm need to function accurately and be tested? For students from all gender groups? For students from all socio-economic backgrounds? Also intersections of these sub-groups? In order to answer these questions, normative decisions need to be made for all bias analyses. In order to make these normative decisions visible and debatable, we highlight two issues in detail before we deduce suppositions from the theoretical framework.

3.4.2 Issues with normativity: Bias and equity

For bias, many differing definitions exist. Each definition comes with its normative implications. For example, in all papers it is agreed upon that biases should be avoided. We highlight two issues with normativity, bias and equity. We start with our understanding of bias by discussing existing definitions in the research field and positioning ourselves relative to these definitions.

Bias addresses the threats of learning analytics. Traag and Schmeling (2022) understand bias as a “direct causal effect that is unjustified” (p. 1). Our understanding of bias differs from this understanding, as for us the question of direct causality is not in focus. A correlation or an indirect causal effect are biases. This argumentation is also in line with the argumentation of the OECD that students’ outcomes should be unrelated to their backgrounds (OECD, 2018, p. 13).

Suresh and Guttag (2021) define one form of bias that is particularly relevant in the context of STEM identity development, historical bias: Historical bias is the result of the perfect reproduction of a world with existing inequalities and thereby the reproduction of the inequalities (p. 4). In the world as it is, not all persons with diverse gender, race, and class are equally represented. Training algorithms with data from the world as it is without adjusting for equity can result in a historical bias. Mitchell et al. (2021) highlight the fact that this understanding of what they call a societal bias is non-statistical (p. 146). We follow this argumentation and stress that the statistically accurate representation of the world as it is in the context of STEM identity is a bias. At the same time, statistically non-accurate representations of the world can be bias-free – if the non-accuracy is due to adjustments in order to strengthen equity.

Baker and Hawn (2021) identify representation bias, aggregation bias, and testing bias (Baker and Hawn, 2021, p. 9). To give testing bias as an example, missing evaluation for sub-group accuracy can lead to good accuracy for the whole group while the accuracy is very good for men but only partly good for women and non-binary persons. From our point of view, it is usually not feasible to address all potential sources of bias. What matters is to be explicit about which sources are considered and which are not. This leads to an understanding of bias as a limited construct bound to the particular analysis focus. Algorithms that are analysed for the most relevant biases can come with other biases, but have been subject to extensive analyses to reduce bias to the absolute minimum. Which analyses have been done and need to be done is a normative question. Being explicit on the analyses (not) made can help to inform the normative decision making processes.

Equity, in contrast to bias, addresses next to the threats also the potentials of learning analytics. Equity allows to address the demand formulated by the OECD (2018) to strive towards no relation between students’ learning outcomes and protected categories. Equity

allows as well to address the idea formulated in the United Nations (2015) goal to provide education for all students.

The normative issue of equity leads us to ask: How do we want career aspirations to be distributed if not in the way they are distributed today? (Costanza-Chock, 2020, p. 63). We need to specify the protected categories we analyse for and select a distribution that we want. On a macro level, these specifications are the normative direction that the OECD or the United Nations provide. On a micro level, these specifications need to be translated in concrete thresholds in algorithmic training. For example, does an algorithm need work for all female and non-binary students at least as good as for male students? Is it possibly okay to violate this strict rule if the overall accuracy gets a lot better while the accuracy for female and non-binary students only lowers a little bit? (Lohaus et al., 2020). If a small violation of the strict rule is fine, do the designers of learning environments need to make up for this violation at another point in the learning environments through counter-measures?

Our point about the normative issue of equity is not to provide an answer to the aforementioned questions. Our point is to make the decisions on these questions visible and debatable. Without clear normative guidance, the learning analytics can reproduce existing inequalities in terms of STEM identity development. The reproduction of inequalities for STEM identity development can lead to different career aspirations in students. However, these different career aspirations in students are undesired by the OECD and the United Nations and should be prevented. At the same time, learning analytics can be designed with a clear normative direction and intervene in a world with existing inequalities towards more equity. Addressing equity in learning analytics allows to consider potentials and to address cultural change as well as to understand diversity as a value.

3.4.3 Suppositions

Based on our theoretical framework on responsible learning analytics in the context of STEM identity development of under-served students from Figure 3-3 and the two issues with normativity, we deduce suppositions for all potentials and threats.

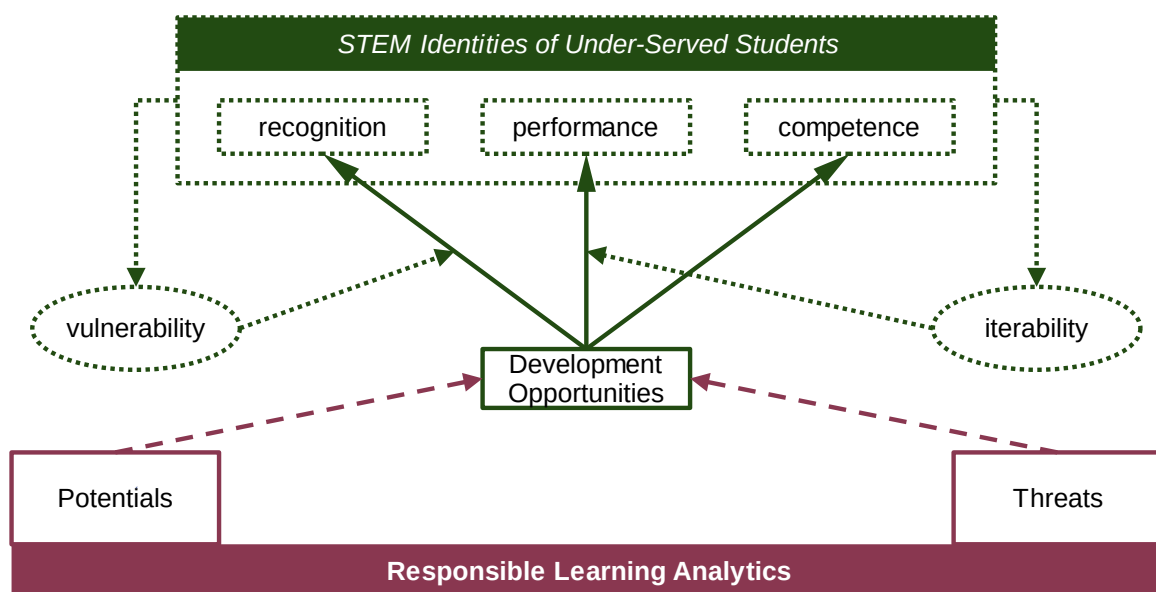


Figure 3-3 - Responsible learning analytics in the context of STEM identities of under-served students

3.4.3.1 *Bias and learning analytics*

Without explicit analyses and careful choice of training data sets, algorithms reproduce existing inequalities as biased algorithms. We analyse the threats of biased algorithms in learning analytics for STEM identity development in its dimensions competence, performance, and recognition.

Bias and competence: Biased algorithms can hinder STEM identity development for under-served students by under-serving them again in terms of competence feedback. Any competence feedback contributes to students' perception of their own competence. An algorithm that provides individualised feedback to students based on what the students write in a task can be trained, for example, with one data set of many student answers to the specific task. If that data set does not represent all students equally well, the choice of this particular data set can lead to the algorithm being biased against those students that are not represented well. When biased algorithms feedback low competence development, this can lead to the student perceiving the own competence in STEM as low. Students' perception of their own competence in STEM is an important piece in building their STEM identity. Therefore, biased competence feedbacks in learning analytics are a threat in terms of STEM identity development.

Bias and performance: Biased algorithms can hinder STEM identity development for under-served students by under-serving them again in terms of performance fitting. This under-serving can further strengthen exclusive norms by iterating them again. For example, STEM can be conceived as a male field in a class already due to many famous male scientists like Isaac Newton or Albert Einstein. If in that situation biased algorithms lead to performance feedback on class level with the male students being more competent than non-binary and female students, STEM as a male field can be further strengthened. In the next STEM identity development opportunity, non-binary and female students with interest in STEM might be questioned in their gender identities. The students would then need to negotiate between their gender and STEM identity which ultimately is another hindrance on the way toward strong STEM identities for these students. This is an example for algorithms that are biased in terms of gender. Other categories as well as their intersections need to be taken into account as well.

Bias and recognition: Biased algorithms can hinder STEM identity development for under-served students by under-serving them again in terms of recognition. This under-serving can further disadvantage precisely the students that are most vulnerable. If, for example, a biased algorithm positions a student with a high competence development through misclassification under low competence development and feeds this classification back to the class as a ranking, the student is not recognised for the high competence development by other students. This effect can also be mediated through a teacher. A learning analytics system that provides feedback to the teacher can trigger the teacher to confront a student in a group with the low competence development classification and thus fail to appreciate a high competence development in a group setting.

3.4.3.2 *Equity and learning analytics*

Instead of focusing on not introducing new biases, equity allows countering historically grown inequalities. Algorithms that lead to stronger recognition for Black students can strengthen equity if this is done in order to make up for missing recognition of precisely these students in other settings. For this redistributive action, an analysis of the reality and the current under-servings in terms of STEM identity development is necessary. To stay with the example from biases, data sets for training of algorithms can have minimum

shares for female and non-binary students if female and non-binary students face historically grown inequalities in the application field. We do not provide an answer to which redistributive action is adequate for which form of discrimination. Nevertheless, we analyse the potentials of equity in learning analytics for STEM identity development in terms of competence, performance, and recognition. Through this analysis, we aim at making the decisions on distributions visible and debatable.

Equity and competence: Responsible learning analytics can support STEM identity development for under-served students by serving them in terms of competence feedback. With the perception of the own competence being important for STEM identity development, this competence feedback can be key to STEM identity development. This holds especially true for contexts in which teachers themselves are biased and stereotyping a lot. Algorithms designed with equity in mind can be an impactful counter-measure here.

Equity and performance: Responsible learning analytics can support STEM identity development for under-served students by serving them in terms of performance fitting. This serving can counter existing disadvantages for under-served students by iterating performances that challenge exclusive norms. Staying with the example from the threats, iterating over and over non-binary and female STEM performances can counter the iteration of male STEM performances through famous male scientists. Finally, this can lead to STEM identity being perceived as inclusive and compatible with all gender identities.

Equity and recognition: Responsible learning analytics can support STEM identity development for under-served students by serving them in terms of recognition. This serving can counter existing disadvantages for under-served students by educating all students to be sensitive to varying levels of vulnerability. With the example of recognition, we propose how equity in learning analytics be implemented: Providing recognition to under-served students can be used by comparing it to historically grown inequalities in terms of recognition, for example the strong gender bias in noble price winners in physics. Making multiple levels of vulnerability visible and explicit in a moment of strong perceived recognition of precisely this vulnerable group of students can lead to them building resilience. In future recognition settings that might be biased, students' resilience can lead to stronger self-confidence of these students as they are aware of different forms of historically grown inequalities and biases.

We do not understand the two issues as the only ones or the most adequate. The decision of which issues are the most important is a political one that cannot be answered by researchers. Instead, we find these issues impactful in terms of STEM identities of under-served students and aim at making normativity explicit and thereby debatable. There are more issues that can be thought of and that would need to be explicitly formulated in order to derive suppositions for future research from them.

3.4.3.3 *Summary: Suppositions*

Responsible learning analytics come with both, potentials and threats, for STEM identity development of under-served students. We position responsible learning analytics in the context of STEM identity development and the dimensions recognition, performance, and competence. In Table 3-1, we present a summary of the suppositions that we identified in the previous sections. The suppositions can give direction to future research.

Table 3-1 - Suppositions

Normative Standpoint	STEM Identity Dimension	Supposition
Bias	Recognition	<p>Learning analytics can hinder STEM identity development for under-served students by under-serving them again in terms of recognition.</p> <p>Learning analytics can further disadvantage precisely the students that are most vulnerable.</p>
	Performance	<p>Learning analytics can hinder STEM identity development for under-served students by under-serving them again in terms of performance fitting.</p> <p>Learning analytics can further strengthen exclusive norms by iterating them again.</p>
	Competence	<p>Learning analytics can hinder STEM identity development for under-served students by under-serving them again in terms of competence feedback.</p>
Equity	Recognition	<p>Learning analytics can support STEM identity development for under-served students by serving them in terms of recognition.</p> <p>Learning analytics can counter existing disadvantages for under-served students by educating all students to be sensitive to varying levels of vulnerability.</p>
	Performance	<p>Learning analytics can support STEM identity development for under-served students by serving them in terms of performance fitting.</p> <p>Learning analytics can counter existing disadvantages for under-served students by iterating performances that challenge exclusive norms.</p>
	Competence	<p>Learning analytics can support STEM identity development for under-served students by serving them in terms of competence feedback.</p>

3.5 Discussion

We proposed a theoretical framework, two issues with normativity, and six suppositions. What do these results mean for STEM identity development of under-served students? How can the results inform a process of transformation of the reality in which the OECD (2018) conclude that no country “can yet claim to have entirely eliminated [...] inequalities in education” (p. 13)? In our discussion, we focus on implications for the research field of responsible learning analytics and its researchers on the one hand. On the other hand, we discuss the role of teachers when interacting with responsible learning analytics systems.

3.5.1 Implications for responsible learning analytics researchers

In learning analytics, many researchers demand for more equity perspectives. For example, Wise et al. (2021) propose a subversive stance on learning analytics. With subversive stance they mean to engage in questions of power and equity. Prinsloo and Slade (2018) advocate for rooting responsible learning analytics in critical theory. So far, there exist various principles on how to enact equity in practice that have been thoroughly reviewed (Sclater, 2014; Prinsloo and Slade, 2018; Cerratto Pargman and McGrath, 2021). However, Kitto and Knight (2019) find that the existing principles often have little impact in practice as they are under-specified. These experiences with principles can support further work on responsible learning analytics in the context of STEM identity development.

In order to make decisions on how to regulate learning analytics in the context of STEM identity development, concrete examples in STEM disciplines are necessary. These concrete examples can then inform political decision making on how to guide and regulate learning analytics practice. Kitto and Knight (2019) propose to create a database with concrete examples on how to navigate between threats and potentials and where guidance is missing. The report on this navigation can be informed by explicit normative decisions as formulated here and point to where these decisions fail to provide clear guidance so far.

For bias analyses, the context of STEM identity development can inform which values to look at and which analyses to make. For example, a purely competence-oriented perspective can lead to other conclusions than a perspective that includes STEM identity development. From a competence perspective, it makes more sense to ensure that learning analytics algorithms identify all students that have not yet understood a concept. A false positive is dangerous because a teacher would assume the student has understood the concept and the teacher would not intervene. A false negative would make the teacher approach the student, notice that the student actually understood the concept and continue with the classes. A STEM identity perspective also makes us ask: How is the sub-group accuracy for different categories of the positive scores? The positive scores can lead to teachers recognising students or not. Recognition is an important dimension of STEM identity, particularly for under-served students. A STEM identity perspective demands for sub-group accuracy in the positive scores and puts a new perspective on which bias analyses need to be made.

Responsible learning analytics researchers should consider the context of STEM identity development when deciding on which bias analyses to make. According to Avraamidou (2020), a special focus on recognition and the emotions linked to recognition processes seems advisable since those two dimensions seem especially important for under-served students.

3.5.2 Connections to critical consciousness of teachers

Next, to learning analytics systems themselves, we want to point at the importance of teachers and their interactions with learning analytics systems. From our point of view, learning analytics can support but do not replace teachers in schools. In reaching equity, teachers play a key role.

One human-centered approach to transform societies with existing inequalities in more equitable societies is critical consciousness. Critical consciousness can be understood as “reflection and action upon the world in order to transform it” (Freire, 1970, p. 51). In the

context of education and for teachers, Baggett (2020) understands critical consciousness as a person's understanding of relevance of and responsibility for categories, for example gender and race, in combination with acting accordingly.

It is crucial to understand how teachers make use of learning analytics in schools in a way that strengthens equity. We need to know what makes teachers critically reflect on the feedback learning analytics systems provide. What do teachers need in order to enact equity in learning analytics for STEM identity development of all students, particularly under-served students? Teachers are key actors in many STEM identity development opportunities and therefore need to be considered in STEM identity development contexts

3.5.3 Conclusions for STEM identity development of under-served students

In order to reach the Sustainable Development Goal 4 on quality education and upward socio-economic mobility, the issues of bias and equity in learning analytics need to be addressed in the context of STEM identity development. Normative decisions need to be made explicit and visible by researchers and designers in order to make their implications debatable. Based on the suppositions we deduced from the theoretical framework, hypotheses need to be formulated as well as empirically tested. These hypotheses need to be as concrete as possible in order to lead to principles for practice that are not under-specified but concrete enough to guide practice. These principles can ultimately lead to STEM identity development opportunities for all students and equally distributed STEM career aspirations.

In addition, under-served students themselves play a crucial role in strengthening equity in practice. To transform societies and to truly change existing inequalities, we need to centre the voices and needs of under-served students and to empower them to be part of this transformation. If we fail to do so, we run into the danger of reproducing exclusion and inequalities. We need to enable students to understand how learning analytics in schools work to create equity through learning analytics. Identifying potential biases in learning analytics might help under-served students to re-evaluate their own scores and to realise that deficits might be rooted in the learning analytics tools and not in their own competencies. This could be a powerful way to empower them to claim potential biases. By doing so, under-served students can be enabled to recognise their own potentials and to construct their own STEM identity based on those potentials. This is what Shanahan (2009) calls agency in terms of STEM identity development.

In this paper, we proposed a theoretical framework that positions responsible learning analytics in the context of STEM identities of under-served students. We discussed two major issues and deduced suppositions that aim at guiding the use of as well as future research on the use of learning analytics in STEM education. With our conclusions for researchers, teachers, and under-served students as well as with our work as a whole we aim at informing political decision making in order to provide STEM identity development opportunities for all students.

Data availability statement. The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

Author contributions. AG, AS, JÇ, MK, and KN contributed to the conception and design of the study and the theoretical framework and main results were created in collective work and discussions with all authors. AG contributed the main work in drafting the sections one

to five. AG and JÇ contributed the main work in drafting the section six. All authors contributed to the article and approved the submitted version.

Funding. This work was supported by the Federal Ministry of Education and Research (BMBF). BMBF project is: 01JD2008.

Acknowledgments. Our work lays on the shoulders of various great thinkers who do not yet receive the visibility they deserve in the context of STEM education from our perspective. We want to highlight especially the work from Black feminist author Kimberlé Crenshaw on intersectionality (1989), the work from nonbinary trans* femme author Sasha Costanza-Chock on design justice (2020), the author from the Global South Paulo Freire on critical consciousness (1970), and Judith Butler with contributions to queer theory (Butler, 1990, 2005).

Conflict of interest. The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note. All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References of the Piece of Scholarship

- Avraamidou, L. (2019). "I am a young immigrant woman doing physics and on top of that I am Muslim": identities, intersections, and negotiations. *J. Res. Sci. Teach.* 57, 311–341. doi: 10.1002/tea.21593
- Avraamidou, L. (2020). Science identity as a landscape of becoming: rethinking recognition and emotions through an intersectionality lens. *Cult. Stud. Sci. Educ.* 15, 323–345. doi: 10.1007/s11422-019-09954-7
- Bachsleitner, A., Lämmchen, R., and Maaz, K. (Eds.) (2022). *Soziale Ungleichheit des Bildungserwerbs von der Vorschule bis zur Hochschule: Eine Forschungssynthese zwei Jahrzehnte nach PISA*. Münster: Waxmann.
- Baggett, H. C. (2020). Relevance, representation, and responsibility: exploring world language teachers' critical consciousness and pedagogies. *L2 J.* 12, 34–54. doi: 10.5070/L212246037
- Baker, R., and Hawn, A. (2021). Algorithmic bias in education. *Int. J. Artif. Intell. Educ.* 32, 1052–1092. doi: 10.1007/s40593-021-00285-9
- Brickhouse, N. W. (2001). Embodying science: A feminist perspective on learning. *J. Res. Sci. Teach.* 38, 282–295. doi: 10.1002/1098-2736(200103)38:3%3C282::AID-TEA1006%3E3.0.CO;2-0
- Brown, B. A. (2004). Discursive identity: assimilation into the culture of Science and its implications for minority students. *J. Res. Sci. Teach.* 41, 810–834. doi: 10.1002/tea.20228
- Butler, J. (1990). *Gender Trouble: Feminism and the Subversion of Identity*. New York: Routledge.
- Butler, J. (2005). *Giving an Account of Oneself*. New York: Fordham University Press.
- Carlone, H. B., and Johnson, A. (2007). Understanding the science experiences of successful women of color: science identity as an analytic lens. *J. Res. Sci. Teach.* 44, 1187–1218. doi: 10.1002/tea.20237
- Cerratto Pargman, T., and McGrath, C. (2021). Mapping the ethics of learning analytics in higher education: A systematic literature review of empirical research. *J. Learn. Anal.* 8, 123–139. doi: 10.18608/jla.2021.1
- Cerratto Pargman, T., McGrath, C., Viberg, O., Kitto, K., Knight, S., and Ferguson, R. (2021). *Responsible Learning Analytics: Creating Just, Ethical, and Caring LA Systems Companion Proceedings LAK21*.
- Costanza-Chock, S. (2020). *Design Justice: Community-Led Practices to Build the Worlds We Need*. Cambridge: The MIT Press.
- Crenshaw, K. (1989). Demarginalizing the intersection of race and sex: A black feminist critique of antidiscrimination doctrine, feminist theory and antiracist politics. *Univ. Chic. Leg. Forum* 1989.
- Dou, R., Hazari, Z., Dabney, K., Sonnert, G., and Sadler, P. (2019). Early informal STEM experiences and STEM identity: the importance of talking science. *Sci. Educ.* 103, 623–637. doi: 10.1002/sce.21499

- Drachsler, H., and Greller, W. (2016). Privacy and Analytics – It's a DELICATE Issue. LAK 16, Edinburgh.
- Düchs, G., and Ingold, G.-L. (2018). Frauenanteil bleibt stabil. *Phys. J.* 17, 32–37.
- Erden, D. (2020). "KI und Beschäftigung: Das Ende menschlicher Vorurteile oder der Beginn von Diskriminierung 2.0?" in *Dann Feministisch*. ed. K. I. Wenn (Berlin: Netzforma* eV), 77–90.
- FEANI (2021). The UN Sustainability Goals: The Role of FEANI/ENGINEERS EUROPE and the European Engineering Community. FEANI (EU STEM Coalition Member). Available at: <https://www.stemcoalition.eu/publications/un-sustainability-goals-role-feani-engineers-europe-and-european-engineering-community> (Accessed October 11, 2022).
- Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., et al. (2018). AI4People—an ethical framework for a good AI society: opportunities, risks, principles, and recommendations. *Minds Mach.* 28, 689–707. doi: 10.1007/s11023-018-9482-5
- Freire, P. (1970). *Pedagogy of the Oppressed*. Penguin Random House UK United Kingdom.
- Godwin, A. (2016). The Development of a Measure of Engineering Identity. 2016 ASEE Annual Conference & Exposition, New Orleans, Louisiana.
- Hartmann, B., and Schriever, C. (2022). *Vordenkerinnen—Physikerinnen und Philosophinnen durch die Jahrhunderte*. Münster: UNRAST Verlag.
- Hazari, Z., Chari, D., Potvin, G., and Brewe, E. (2020). The context dependence of physics identity: examining the role of performance/competence, recognition, interest, and sense of belonging for lower and upper female physics undergraduates. *J. Res. Sci. Teach.* 57, 1583–1607. doi: 10.1002/tea.21644
- Hoppe, H. U. (2017). "Chapter 2: computational methods for the analysis of learning and knowledge building communities," in *Handbook of Learning Analytics*. 1st Edn. eds. C. Lang, G. Siemens, A. Wise and D. Gašević (SoLAR), 23–34.
- Kitto, K., and Knight, S. (2019). Practical ethics for building learning analytics. *Br. J. Educ. Technol.* 50, 2855–2870. doi: 10.1111/bjet.12868
- Lohaus, M., Perrot, M., and von Luxburg, U. (2020). Too relaxed to be fair. *Proc. Mach. Learn. Res.* 119, 6360–6369.
- Mahadeo, J., Hazari, Z., and Potvin, G. (2020). Developing a computing identity framework: understanding computer science and information technology career choice. *ACM Trans. Comput. Educ.* 20, 1–14. doi: 10.1145/3365571
- Mitchell, S., Potash, E., D'Amour, A., and Lum, K. (2021). Algorithmic fairness: choices, assumptions, and definitions. *Ann. Rev. Stat. Appl.* 8, 141–163. doi: 10.1146/annurev-statistics-042720-125902
- Mujtaba, T., and 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. *Int. J. Sci. Educ.* 35, 1824–1845. doi: 10.1080/09500693.2012.762699

3 Responsible Learning Analytics and STEM Identities

- National Academies of Sciences, Engineering, and Medicine. (2019). *Science and Engineering for Grades 6-12: Investigation and Design at the Center*. Washington, DC: The National Academies Press.
- OECD (2018). *Equity in Education*. Paris: OECD.
- Phillips, P. J., Hahn, C. A., Fontana, P. C., Broniatowski, D. A., and Przybocki, M. A. (2020). *Four Principles of Explainable Artificial Intelligence*. Gaithersburg: National Institute of Standards and Technology.
- Prinsloo, P., and Kaliisa, R. (2022). Learning analytics on the African continent: an emerging research focus and practice. *J. Learn. Anal.* 9, 1–18. doi: 10.18608/jla.2022. 7539
- Prinsloo, P., and 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).
- Rosa, K., and Moore Mensah, F. (2016). Educational pathways of black women physicists: stories of experiencing and overcoming obstacles in life. *Phys. Rev. Phys. Educ. Res.* 12:020113. doi: 10.1103/PhysRevPhysEducRes.12.020113
- Sclater, N. (2014). *Code of Practice for Learning Analytics Jisc*, 1–64.
- Shanahan, M.-C. (2009). Identity in science learning: exploring the attention given to agency and structure in studies of identity. *Stud. Sci. Educ.* 45, 43–64. doi: 10.1080/03057260802681847
- Slade, S. (2016). *The Open University Ethical Use of Student Data for Learning Analytics Policy*. The Open University.
- Society for Learning Analytics Research (SoLAR) (2022). *What is Learning Analytics?* Available at: <https://www.solaresearch.org/about/what-is-learning-analytics/> (Accessed August 31, 2022).
- Suresh, H., and Gutttag, 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*. ACM, New York, NY, USA, 1–9.
- Taconis, R., and Kessels, U. (2009). How choosing science depends on students' individual fit to 'science culture'. *Int. J. Sci. Educ.* 31, 1115–1132. doi: 10.1080/09500690802050876
- Taskinen, P. H., Schütte, K., and Prenzel, M. (2013). Adolescents' motivation to select an academic science-related career: the role of school factors, individual interest, and science self-concept. *Educ. Res. Eval.* 19, 717–733. doi: 10.1080/13803611.2013.853620
- Traag, V. A., and Schmeling, L. (2022). Causal foundations of bias, disparity and fairness. *ArXiv*. doi: 10.48550/arXiv.2207.13665
- UN-SDG-Goal 4. (2015). UN. Available at: <https://sdgs.un.org/goals/goal4> (Accessed October 11, 2022).
- Waight, N., Kayumova, S., Tripp, J., and Achilova, F. (2022). Towards equitable, social justice criticality: re-constructing the “black” box and making it transparent for the

future of science and Technology in Science Education. *Sci. & Educ.* 31, 1–23. doi: 10.1007/s11191-022-00328-0

Wise, A. F., Sarmiento, J. P., and Boothe, M. (2021). Subversive Learning Analytics. *Proceedings of the 11th International Conference on Learning Analytics and Knowledge (LAK' 21)*, 12–16 April 2021, Irvine, CA, USA, 639–645.

Zhai, X., Haudek, K. C., Shi, L., Nehm, R. H., and Urban-Lurain, M. (2019). From substitution to redefinition: A framework of machine learning-based science assessment. *J. Res. Sci. Teach.* 57, 1430–1459. doi: 10.1002/tea.21658

3 Responsible Learning Analytics and STEM Identities

Ressourcen-orientiert sein

*Ich möchte Ressourcen-orientiert sein:
Aktivistisch, arbeitend, verbindlich, rundum.
Doch dabei treiben mich Fragen um,
Und die sind nicht gerade klein!*

*Ressourcen-orientiert sein heißt für mich:
Ich bin in Ordnung.
Es ist okay, keine Energie zu haben,
Es ist wichtig, mich auch mal in Genuss zu laben,
Self-Care ist eine notwendige Voraussetzung,
Ich kann nicht pausenlos intervenierend aktiv sein,
Zeit zur Regeneration ist immens wichtig,
Transformation ist ein Ausdauer-Lauf, und der ist nicht zu klein!*

*Doch wenn ich nicht einschreite,
Weil mein Energie-Level zu gering ist, heute,
Kann ich dann abends noch feiern gehen?
Kann ich mir im Spiegel in die Augen sehen,
Wenn ich Wochen-lang entspanne, genieße,
Ja mit meiner Enthaltsamkeit gar den Weg bereite
Für noch mehr Leid in uns'rer Welt, hieße
Das nicht ausblenden, wegsehen – gar das sich das Böse mit einläute?*

*Die Grenze zwischen Ressourcen-orientiert
Und ignorant-unschuldig ist verschwommen;
In diesem Konsent seh' ich 'nen Versuch unternommen
Sie einzufangen, sicht- und diskutierbar zu machen, sie für mich präzisiert:*

*Ich verurteile mich und andere nicht,
Wenn wir eine Kreuzfahrt machen,
Das in-den-Urlaub-Fliegen genießend lachen,
Wenn wir für uns festlegen, schlicht:
Ich habe dafür keine Kraft,
Ich muss Räume nicht schützen, wenn das mich dahinrafft.*

*Was ich verurteile, und zwar scharf,
Das ist Leid ignorieren,
Das ist auf vermeintliche Intention fokussieren
Statt auf Wirkung und Betroffene, das darf
Uns nicht egal sein, uns unschuldig-ignorant lassen,
Dieser Auslegung möchte ich einen Tritt in den Arsch verpassen!*

*Es purzelt als Konsent für mich aus der Reflexion:
Solidarität mag mich leiten in meinem Handeln,
Ich erlaube mir Pausen und zurückgezogen zu wandeln,
Zentral ist unser gemeinsam-entschlossener Schritt gen Transformation!*

4 Equity-Focused Decision-Making Lacks Guidance!

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

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.

Published. Grimm, A., Steegh, A., Kubsch, M., & Neumann, K. (2023). Learning Analytics in Physics Education: Equity- Focused Decision-Making Lacks Guidance! *Journal of Learning Analytics*, 10(1), 71–84. <https://doi.org/10.18608/jla.2023.7793>

4.1 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 cases¹² 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.¹³

4.2 Theory

4.2.1 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.

¹² 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.

¹³ 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).

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.

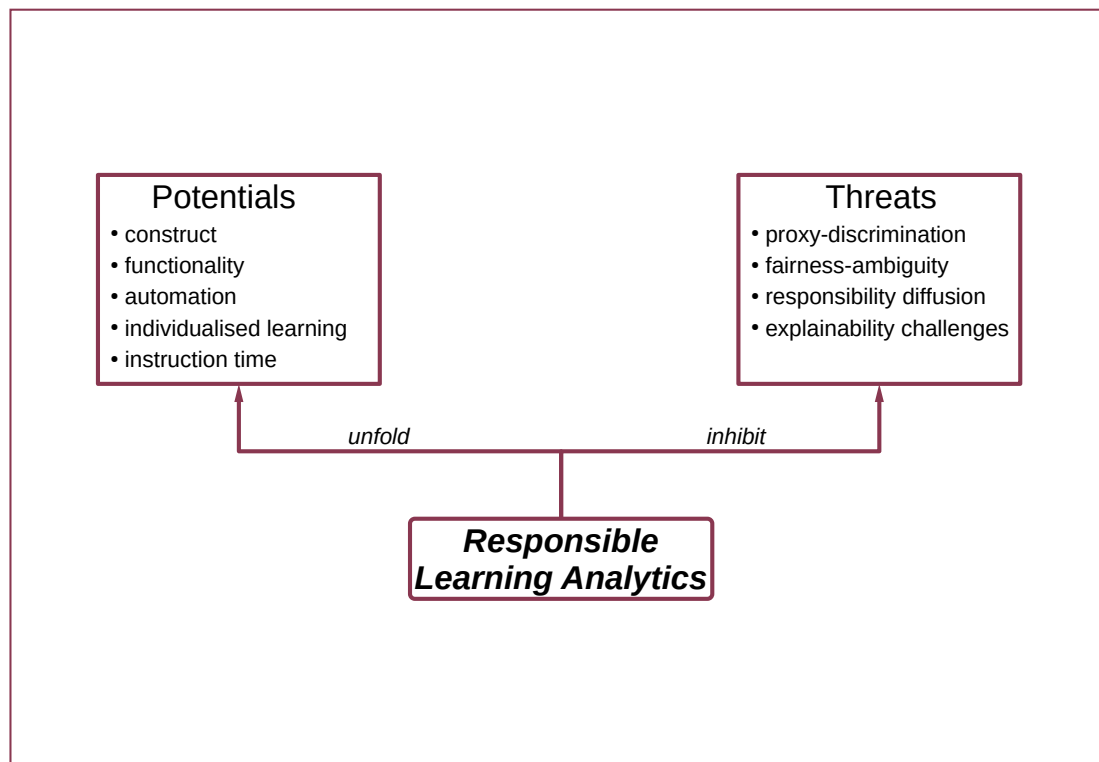


Figure 4-1 - Responsible learning analytics — practice between potentials and threats

The potentials of learning analytics are summarized on the left side of Figure 4-1. 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 4-1. 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

4 Equity-Focused Decision-Making Lacks Guidance!

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.

4.2.2 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). Costanza-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.

4.2.3 Historically Grown Inequalities in Physics Education in Germany

There is a broad range of threats when applying learning analytics in physics education. In Figure 4-2, 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).”¹⁴ 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.”¹⁵

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), 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.

¹⁴ The quote was translated from German into English by the authors.

¹⁵ The quote was translated from German into English by the authors.

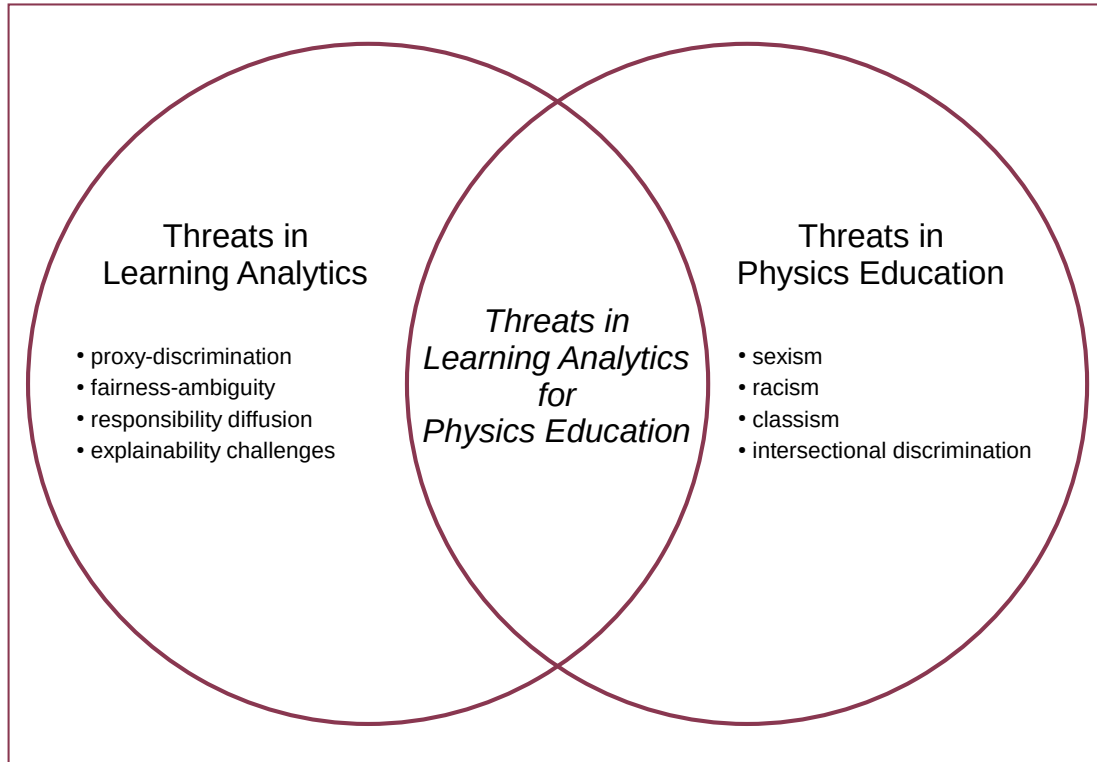


Figure 4-2 - Threats in learning analytics for physics education

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

4 Equity-Focused Decision-Making Lacks Guidance!

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.

4.2.4 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 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.

4.3 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.

4.4 Edge Cases

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

4.4.1.1 *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 4-3, 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?

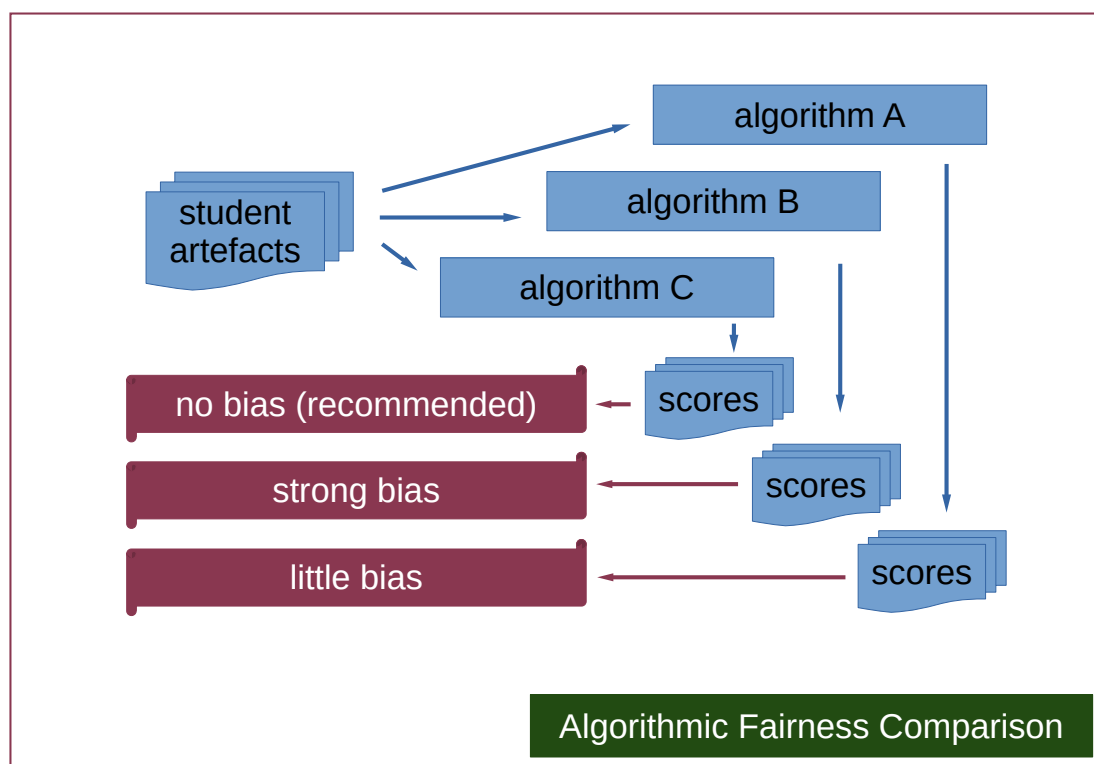


Figure 4-3 - Bias analyses

4.4.1.2 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.

4.4.1.3 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.

4.4.1.4 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.¹⁶

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 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.

4.4.1.5 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.

¹⁶ 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.

4 Equity-Focused Decision-Making Lacks Guidance!

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.

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

4.4.2.1 *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 4-4. 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 consciousness⁶ 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?

4.4.2.2 *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).

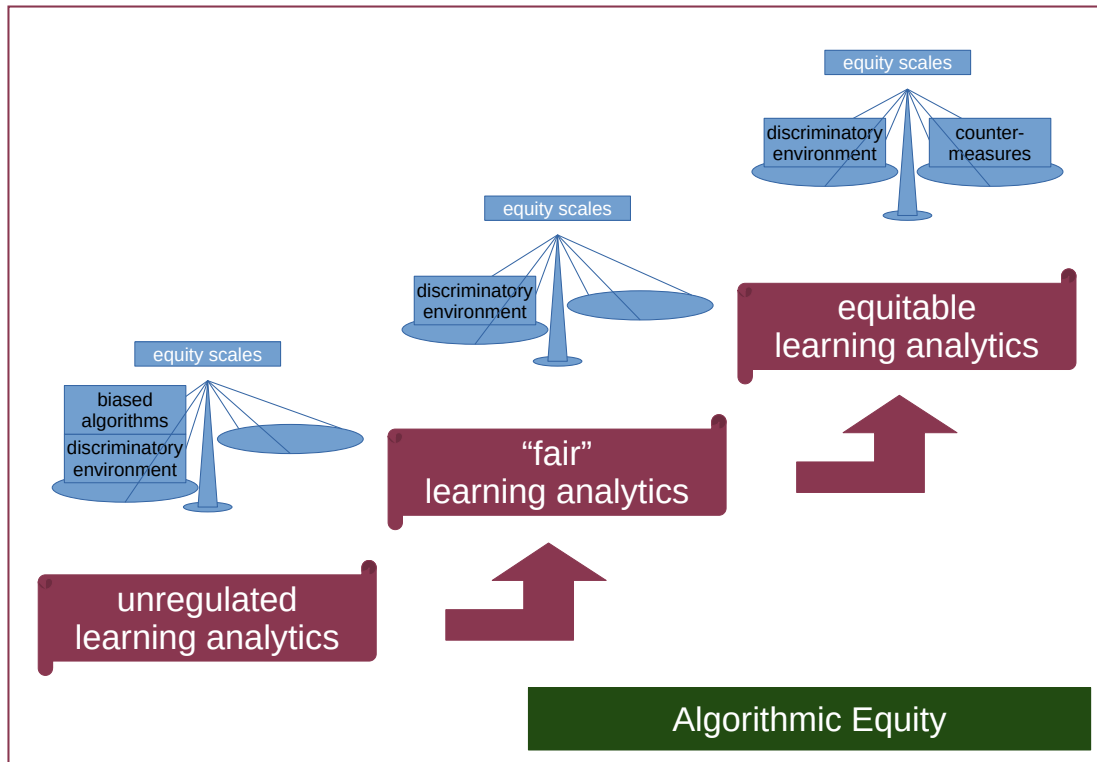


Figure 4-4 - Measures to counter discrimination

4.4.2.3 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?

4.4.2.4 Difficulties

Before designing counter-measures, we needed a strong normative grounding to provide legitimization. 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

4 Equity-Focused Decision-Making Lacks Guidance!

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.

4.4.2.5 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 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.

4.5 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.

4.5.1 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.

4.5.2 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 of the Piece of Scholarship

- Aikins, M. A., Bremberger, T., Aikins, J. K., Gyamerah, D., & Yıldırım-Calıman, 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 Successful 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/rrfa9szd/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>

4 Equity-Focused Decision-Making Lacks Guidance!

- Kitto, K., & Knight, S. (2019). Practical ethics for building learning analytics. *British Journal of Educational Technology*, 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](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 sense of belonging influences decisions to major in 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>

4 Equity-Focused Decision-Making Lacks Guidance!

Die Leere danach

Da ist sie wieder, diese Leere.
Als hätte jemand den Stöpsel gezogen.
Ohnmacht. Überrollt.
Unterlegen. Chancenlos.
Wo kommst Du bloß
Her? Gehst Du wieder, von allein?
Was ist falsch mit mir?
Kommst Du aus meinem Alltag ins Hier?
Muss ich wohl mein
Leben grundlegend ändern, sollt'
Ich reflektieren, um dann gründlich abgewogen
Etwas loszulassen, weil ich dann wieder bewegt wäre?

Was brauch' ich gerade?
Nein, ich kann aus dieser Leere
Noch nicht heraus,
Will hier erst verweilen,
Will schwach sein, mir erlauben in diesen Seilen
Zu hängen,
Nicht gleich nach Lösungen suchen,
Nicht meine Gefühle verfluchen,
Nicht drängen;
Denn meine Leere ist Zeugin geraderaus
Für meinen Kompass im Angesicht von Schwere,
Sie ist für meine Integrität die Parade.

Was brauch' ich danach?
Gehört werden.
Schreiben. Worte finden.
Kürzer arbeiten,
Laufend durch frische Winde gleiten,
Mich durchpusten lassen,
Kalten Regen auf der Haut verspüren,
Singend meine Gefühle aus voller Brust berühren,
Meine Leere als Haltung neu erfassen;
Dieses Licht kann mich von meinen Fragen entbinden,
Kann machtvoll mich erden,
Kann liebevoll lindern meine Schmach.

5 De-Biasing

Title. Algorithmic Justice in Education Through De-Biasing: Towards Politically Actionable Evidence That is Rooted in Identity Theory and De-Colonial Thought

Abstract. As Artificial Intelligence (AI) algorithms are increasingly used in education, research shows that the use of these algorithms is not without cost. Instead, AI algorithms are prone to biases which are discussed a lot in various domains. The core strength of this contribution is to anchor the discussion of biases in the specific domain of physics education and to discuss the biases in front of a description of the domain-specific inequalities along physics identity development of students. The database consists of the written answers of 527 students to around 30 items from a five-week-period of physics classes in a digital learning environment. Two concrete biases of AI algorithms in physics education and possible approaches to identify and reduce these biases are investigated quantitatively. In a critical discussion from a feminist and de-colonial perspective, it is highlighted that the chosen approaches seem to have promising potentials to mitigate negative effects on under-served students' physics identity development and relevant limitations that demand for additional counter-measures in order to break out of the vicious cycle of reproduction of historically grown inequalities in physics education as well. The domain-specific analysis can serve orientation for other domains as well in order to tackle the challenges of AI algorithmic bias effectively and efficiently.

Submitted. Grimm, A.; Gombert, S.; Armbrüster, S.; Kubsch, M.; Steegh, A.; Navarro Camacho, M.; Kolbe, H.; Tautz, S.; Petersohn, K.; Holst, V.; Karademir, O.; Bohm, I.; Neumann, K. (2025). Algorithmic Justice in Education Through De-Biasing: Towards Politically Actionable Evidence That is Rooted in Identity Theory and De-Colonial Thought. Frontline Learning Research

5.1 Introduction

5.1.1 Relevance and Contribution

As Artificial Intelligence (AI) algorithms are increasingly used in education, research shows that the use of these algorithms is not without cost. A growing number of studies have revealed, for example, that AI algorithms are prone to racist biases (Cheuk, 2021; Erden, 2020). Hence, mitigating such biases – that is, de-biasing – of algorithms has become increasingly important. From early on, researcher communities have acknowledged the risks related to biases and proposed principles for “ethical use” (Slade, 2016), “non-discrimination” (Bergmann et al., 2019), “fairness” (Diakopoulus et al., 2021), “justice” (Floridi et al., 2018, pp. 696–700), or “equitable treatment of all people” (Cerratto Pargman et al., 2021, p. 2). Despite the existence of such principles, in education, algorithms are rarely de-biased in practice (Kitto & Knight, 2019). In fact, a systematic literature review revealed that most scholarship on bias is rather focused on describing biases than developing strategies to address them, highlighting a significant research gap (L. Li et al., 2023). In other words: We know about biases but do not know how to mitigate them. Not knowing how to mitigate biases is problematic because the mere knowledge of existence of biases is not actionable, neither in the practice of algorithm design nor when it comes to policies for algorithm use.

Previous work has identified substantial challenges when it comes to de-biasing algorithms. Lohaus et al. (2020), for example, demonstrated a major influence of the very definition of bias by showcasing how an algorithm de-biased according to one definition can be understood as biased according to another definition; and Erden (2020) found that algorithms can reproduce historically grown inequalities even if the training data contain no information about race – through the mechanism of proxy discrimination. In educational contexts many historically grown inequalities exist. In the European Union only 28 % of the 2018 graduates in engineering, manufacturing, and construction were women. In Germany, where 79 % of the students from academic households start an academic career, only 12 % of the students whose parents of no professional qualification do so (El-Mafaalani, 2021, pp. 66–67). The reproduction of these inequalities does not take place in the arena of student achievement, in which de-biasing is currently performed (Düchs & Ingold, 2018; OECD, 2016). Instead, scholars suggested that the historically grown inequalities are reproduced in the arena of student identity development (Archer et al., 2015; Avraamidou, 2019; Dou et al., 2019). Global South and Black feminist scholars have stressed that the reproduction of historically grown inequalities is a systemic issue and hence requires a systemic perspective beyond the isolated case in order to address it (Collins, 1990; Crenshaw, 1989; Escobar, 2017; Freire, 1970; Mignolo, 2007). We hence recognise the necessity 1) to conceptualize bias in the context of identity theory and 2) to approach de-biasing from a feminist, de-colonial perspective to make them practically and politically actionable.

This paper showcases how to de-bias algorithms in STEM education by i) focusing on the specific biases relevant in the context of historically grown inequalities in STEM identity development and ii) taking a feminist, de-colonial perspective in the de-biasing of algorithms to account for the underrepresentation of groups subject to historically grown inequalities in training data. As relevant historically grown inequalities, we consider gender, social class, and their intersections. In our analysis we draw on two different methods: 1) a training dataset analysis in which we examine the extent to which the training data indeed exhibit bias by predicting the positionalities along diversity dimensions based on the training data and 2) a training dataset slicing analysis with slices made along

positionalities along diversity dimensions to offset underrepresentation of groups subject to bias. In the discussion, we highlight how conceptualizing bias in the context of identity theory is mandatory to obtain a valid definition of bias in the respective context and how a de-colonial feminist perspective can help finding the right strategies to mitigate bias in the training data. Precisely these bias definitions and strategies are needed to reach politically relevant and actionable conclusions. With our research we seek to contribute an exemplar of a de-biasing process rooted in a strong theoretical framework and a sound methodological procedure.

5.1.2 Authors' positions based on feminist standpoint theory

"Feminist standpoint theory recognizes that all knowledge is situated in the particular embodied experiences of the knower" (Costanza-Chock, 2020, p. 9). Acknowledging the relevance of standpoints, we position ourselves as authors on diversity dimensions in order to make our standpoints explicit and thereby the positionalities of researchers in certain areas transparent – and in order to open up an opportunity to identify a lack of diversity where diversity would be necessary. At the same time, we value the privacy of each author and therefore do not provide detailed statistics for all diversity dimensions. As a research team, we carry many privileges: We are cis-gendered *white*¹⁷ women and men. We all hold European citizenships and are able-bodied. We acknowledge that these positionalities shape, for example, our perspectives on our data and the questions we ask.

5.2 Theoretical Background

All students should have equal access to education; that is, the opportunity to engage in learning, develop competence and, more importantly, a respective identity (OECD, 2024, p. 9). However, education is subject to a broad range of historically grown inequalities, especially in STEM and physics education. In STEM, for example, although students who identify as female show similar levels of competence as students who identify as male, the latter show substantially stronger aspirations to pursue a career in this domain (OECD, 2016); that is, develop a stronger STEM identity. A growing number of researchers attribute the diverging identity development to STEM education that marginalizes students who identify as female or, more broadly, students from underrepresented groups (Hazari et al., 2020) and warn that this development is further amplified by the growing use of Artificial Intelligence (AI) algorithms to monitor student learning (Cheuk, 2021).

5.2.1 Identity Development

The concept of identity and its development has gained traction in science education (Hazari et al., 2020) as issues such as the leaky pipeline, i.e. the disproportionate loss of capable women from STEM disciplines, were not solely explained by established constructs such as interest (Hazari et al., 2010). Identity theory represents a more comprehensive frame for understanding why students engage in STEM, how some students are promoted while others are marginalised and hence a means to work towards more equitable STEM education (Carlone & Johnson, 2007). In principle, STEM identity refers to the ways in which a person navigates the meaning of STEM and how the person positions itself with respect to STEM (Çolakoğlu et al., 2023). Identity development is understood as the process of negotiating the multiple identities that a person possesses (Gee, 2000); including disciplinary identities such as a STEM identity, social identities such

¹⁷ We set white in italics to emphasize it as a privileged position in the structure of racism rather than a skin colour as done and argued by Black German author Tupoka Ogette in her book "exit racism" (Ogette, 2019, p. 14).

as gender (i.e., I am identifying as female, physicists are rarely female), or personal identities such as being a loner (i.e., I am a loner, people liking physics are often loners) (Hazari et al., 2010). The process of developing an identity is driven by students' experiences from prior and current experiences in STEM (Calabrese Barton et al., 2013; Shanahan, 2009). In turn, the development of a STEM identity is considered the major pre-requisite for further engagement in STEM. That is, the process of developing a STEM identity can be described as an iterative process, in which engagement in STEM is a prerequisite for the development of an identity in STEM, and the development of a STEM identity drives further engagement.

Research has repeatedly documented that a substantial number of students are under-served in terms of developing an identity in STEM or, more specifically, physics. Examples include students who identify as female (Ladewig et al., 2020; Traxler et al., 2016), with low socio-economic status (Bachleitner et al., 2022; El-Mafaalani, 2021), or of colour (Rainey et al., 2018), as well as students located in the intersections of these dimensions (Avraamidou, 2019; Rosa & Moore Mensah, 2016). A growing body of research examined the mechanisms underlying groups of students being under-served in terms of STEM identity development (for an overview see (Çolakoğlu et al., 2023)). Although the development of a STEM identity is a complex process and the reasons for which a student may not develop a STEM identity are typically complex and highly individual, several principal mechanisms have emerged. A principal mechanism relevant in the context of this study, iterability, is driven by historically grown inequalities (Butler, 1990; Hartmann & Schriever, 2022). Many researchers in physics have been (white) men, the role of non-white men or women is often marginalised. By the year 2021, a total of 214 out of 218 Nobel prizes in physics have, for example, been awarded to men; in some cases, in plain ignorance of the contribution of women. As these Nobel prize winners are commonly glorified in physics education, female students are more likely to perceive physics as a male discipline; barring them, from developing a physics, or more broadly, a STEM identity. The second mechanism, vulnerability, is fuelled by discrimination (Avraamidou, 2019, 2020). Being recognised is not only a core need of every human being (Hartmann & Schriever, 2022), but also crucial for developing a STEM identity (Carlone & Johnson, 2007). Especially students from marginalised groups are commonly not sufficiently recognised in their efforts to engage in science. Instead, these students are often faced with stereotypical expectations because they belong to a group traditionally marginalised in STEM. If Black students are repeatedly confronted with unusual surprise when they engage in science activity, the development of their STEM identity is hindered. Both mechanisms are also deeply intertwined. Historically grown inequalities may signal students STEM is not for them, but at the same time fuel stereotypes that will in turn lead to fewer or false recognition. In combination, both mechanisms can mutually reinforce each other. Bodnar and colleagues (2020), for example, found that while girls, in general, had lower science aspiration scores than boys, Black girls scored lowest in that category.

Historically grown inequalities pose a particular threat for the development of an identity in STEM, leaving affected students with the conclusion that STEM is not for them and a likely decision for a career outside of STEM. A career in STEM, however, is commonly leading to higher (financial) power and (societal) recognition compared to other occupations. Hence, to strengthen diversity and broaden access to education for all students it is necessary to find ways to counter historically grown inequalities with their processes of self-reproduction into future inequalities.

5.2.2 Learning Analytics

Recent developments in the field of learning analytics (i.e., the intelligent analysis of data related to student learning) promise a way forward to address historically grown inequalities. AI algorithms provide the opportunity to monitor student learning or learning-related constructs, predict to which extent students will meet the overarching educational goals and offer targeted opportunities to support students further learning (Karademir et al., 2024). More specifically, AI algorithms can provide students with immediate and personalised feedback about their progress at scale (Dennis et al., 2016; Pardo et al., 2019), allowing for all students to engage in the performance of STEM practices together and experience competence and recognition – supporting the development of a STEM identity (Carlone & Johnson, 2007; Hazari et al., 2010, 2020).

However, AI algorithms also pose a number of threats that may reinforce historically grown inequalities. In fact, AI algorithms are not simply just, ethical or even functioning as intended (Uttamchandani & Quick, 2022). Instead, a rapidly growing amount of research shows how AI algorithms can be racist (Dressel & Farid, 2018), biased towards socio-economic status (Fletcher et al., 2021) or reproduce gender stereotypes (Bolukbasi et al., 2016). This behaviour of AI algorithms is particularly concerning when it affects vulnerable groups; for example, students from groups historically underrepresented. If AI algorithms used for learning analytics assessed female students' learning systematically worse than male students, the female students, who typically struggle with developing a physics identity anyhow, will experience insufficient recognition and subsequently are less likely to develop a physics identity; effectively reducing the number of female students in physics and thus further increasing historically grown inequalities. A range of recently published studies suggest that this effect is in place for a broad spectrum of vulnerable student groups (Bolukbasi et al., 2016; Cheuk, 2021; Costanza-Chock, 2020; Sha et al., 2022; Traag & Waltman, 2022). Hence, the use of AI algorithms for the purposes of learning analytics without a critical reflection is likely to not only not live up to its potential for a more equitable education but bears substantial threats to make education less equitable by reinforcing existing discriminatory practices (Uttamchandani & Quick, 2022).

As a basis for a more critical reflection in the use of AI algorithms for learning analytics purposes, we proposed a framework that describes how learning analytics can affect STEM identity development (Figure 5-1; Grimm et al., 2023). The framework links STEM identity development to (responsible) learning analytics. At the center of the framework are STEM identity development opportunities. These opportunities must address three dimensions of STEM identity development: recognition, performance, and competence (Carlone & Johnson, 2007). Recognition and performance are particularly relevant to under-served students due to two mechanisms of discrimination, vulnerability and iterability (Grimm et al., 2023). Learning analytics poses potentials and threats to STEM identity development opportunities. Navigating both, potentials and threats, is what “responsible learning analytics” are about (Prinsloo & Slade, 2018). One particular threat arises from the use of AI algorithms for learning analytics that exhibit substantial bias; bias that reflects a misinterpretation of student performance leading to students perceiving their competence wrongly and receiving less recognition.

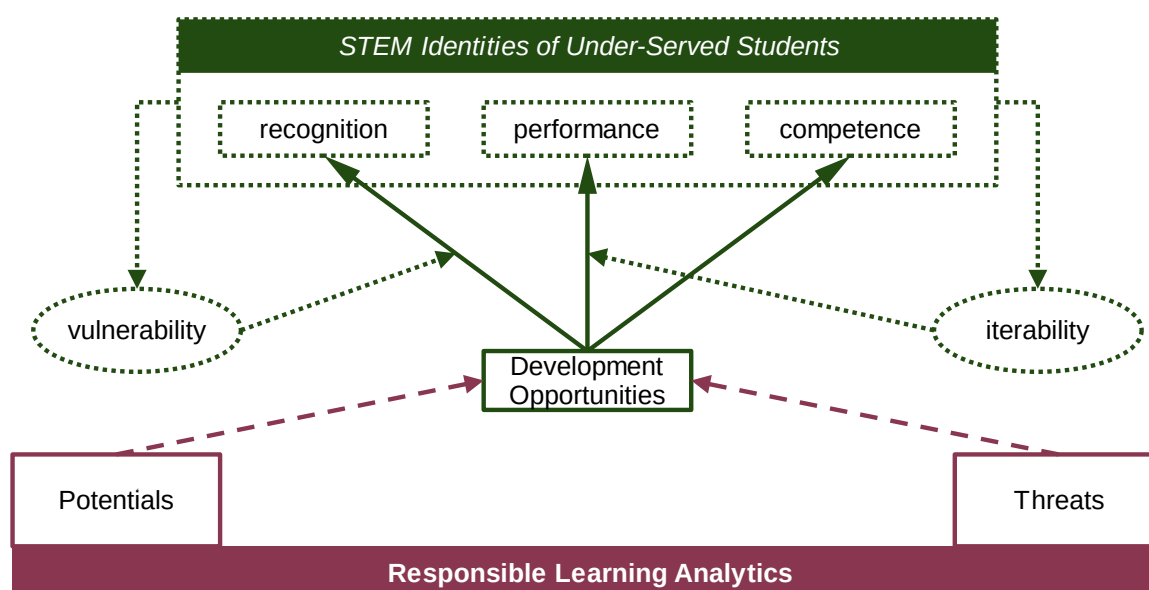


Figure 5-1 - STEM Identities of Under-Served Students and Responsible Learning Analytics (Grimm et al., 2023)

5.2.3 Bias in Learning Analytics

Issues of equity in learning analytics have been approached in terms of the absence of algorithmic bias (Uttamchandani & Quick, 2022). In principle, algorithmic bias refers to situations in which AI algorithms yield a result that advantages or disadvantages specific groups; that is, where decisions made based on the results from the algorithm are discriminatory, violating norms of justice and equity (Kordzadeh & Ghasemaghahi, 2022, p. 388). Algorithmic bias can result, for example, from bias in training data (i.e., specific groups being insufficiently represented in the training data) or the bias(es) of those who develop the algorithms (i.e., specific prejudices, stereotypes or other, often unconscious, attitudes of the developers). Biased algorithms can lead to unjust perceptions, policies, and practices of oppressing underrepresented groups (Uttamchandani & Quick, 2022). Erden (2020) found, for example, that algorithms can reproduce historically grown inequalities even if the training data contain no information about race – through the mechanism of proxy discrimination. Further examples include bias along the dimension of ability (Hutchinson et al., 2020) or intersectional biases (Guo & Caliskan, 2021; Tan & Celis, 2019).

Computational scientists have developed mathematical techniques to detect and mitigate biases in algorithms and learning analytics researchers have acknowledged the relevance of bias and started to explore bias in AI algorithms used for learning analytics (Cerratto Pargman & McGrath, 2021; Doroudi & Brunskill, 2019; W. Li et al., 2019; Prinsloo & Kaliisa, 2022; Prinsloo & Slade, 2017). Alexandron and colleagues (2019), for example, find that learning analytics algorithms supposed to measure students' performance on massive open online courses can be biased by users misusing the system. In a similar notion, Riazzy and colleagues (2020) report that learning analytics algorithms designed to predict course outcome and, more specifically, identify students at risk, reproduce bias against some groups (for example women), even if these groups are equally represented in the (training) dataset. If biases in training datasets exist, AI algorithms have been shown to come with the threat of bias amplification (Zhao et al., 2017). Overall, bias in learning analytics algorithms and the threats that come with it are well described (Baker & Hawn, 2021; Erden, 2020; Lohaus et al., 2020; Phillips et al., 2020; Yeung, 2019). However, there is little research to date on how to mitigate bias in learning analytics. If bias can be reduced,

removed completely and hence if ultimately equitable analytic-based decision making is possible is an open question.

One reason for a lack of research on how to mitigate bias in learning analytics is likely the very definition of bias. Lohaus and colleagues (2020), for example, demonstrated that an algorithm de-biased according to one definition can be understood as biased according to another definition. Bias has in fact been defined in multiple ways (Baker & Hawn, 2021; Gardner et al., 2019; Mitchell et al., 2021; Suresh & Guttag, 2021). In the context of our aim to achieve a more equitable access, especially with respect to the development of identity, we define bias as an algorithm yielding “undesirable [...] behaviours or properties” (Cheuk, 2021, p. 2) that lead to a preference for an already privileged group, such as the preference of men through an algorithm, or a discrimination against an already underprivileged group. Note that discrimination against a privileged group (i.e. white men) can serve as a justifiable means to achieve greater equity and thus would not fall under bias in our definition. This is in line with Constanza-Chock (2020) who argues that for example racial hierarchies “can only be dismantled by actively antiracist systems design, not by pretending they don’t exist” (p. 62). Accordingly, it is not enough, to explore and describe bias, but it is necessary to identify ways by which bias can be counteracted; that is by which learning analytics algorithms can be de-biased.

5.2.4 De-Biasing Learning Analytics

De-biasing learning analytics algorithms requires attending to the different forms of bias that may occur prior to and during the development as well as the use of an algorithm. In Figure 5-2, which is heavily informed by the work of Baker and Hawn (Baker & Hawn, 2021, p. 9), we present an algorithmic life cycle along with entry points of bias; that is, points that de-biasing strategies may be directed at.

Before an AI algorithm enters its use phase, many steps need to be taken and in each of these steps, bias can enter the algorithm. The world as it is already might contain biases, for example historically grown inequalities in the case of physics education. The online world is no perfect representation of the world as it is and the choice of the platform heavily depends on the task an algorithm is set out to perform. Instead of a perfect representation, some parts of the population might not be present on a particular platform. However, an algorithm can only be trained with digital data available. Hence, the data preparation can suffer of underrepresentation of specific subgroups, the measurement process can introduce biases, and when data is scored as in our context the annotation may introduce biases. The model training itself can introduce new biases when there is no mechanism in place to assure the same performance over all subgroups, for example. Finally, the algorithm may have bias in the use phase in the case of a deployment bias: If an algorithm is trained in a setting with high socio-economic status and digital competence but is then used in a different context, the algorithm might work worse and thereby the mismatch of training and use population can introduce a bias. Two of these biases, namely representation and evaluation bias with the entry points data preparation and model training, bear special relevance to STEM identity development, and hence our work.

Representation bias results from the underrepresentation of specific sub-groups of the target population in the training data that will be used to train the algorithm. Representation data can originate, for example, from how the data was collected or be an inert part of the data when the sub-groups represent minority groups within the target population (Shahbazi et al., 2023). Since there is fewer information to train the algorithm with, predictions made by the algorithm may be less accurate and hence decisions made based on the predictions can negatively affect students’ identity development.

Evaluation bias occurs when the criteria used for evaluating the algorithm differ across sub-groups of the target population. One source of evaluation bias is that the criteria used for evaluation advantage or disadvantage a specific subgroup. In case of evaluation bias, the algorithm may be found to function accurately across all kinds of groups but show high error rates when used with a specific subgroup, with the same effect as representation bias; namely, that students' identity development will be impeded.

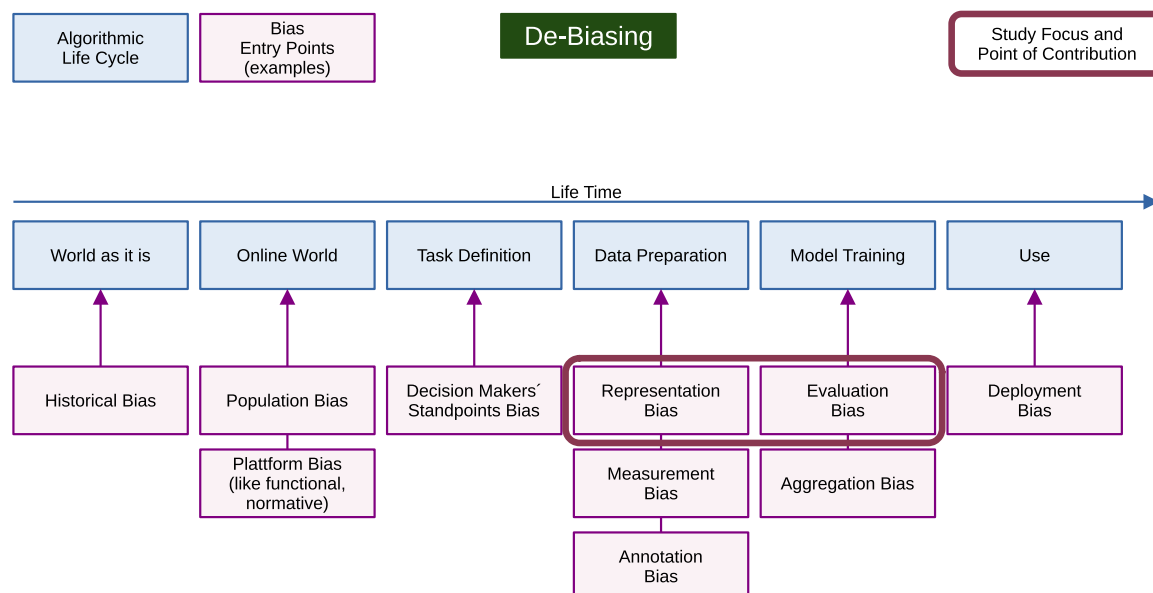


Figure 5-2 - Bias Entry Points in Learning Analytics and Focus on This Paper (inspired and informed by (Baker & Hawn, 2021, p. 9))

Li and colleagues (2023) identify three groups of de-biasing strategies for machine learning algorithms: 1) pre-processing (i.e. sampling or transforming of the training data to reduce underrepresentation), 2) in-processing (i.e., designing the algorithm to reduce potential discrimination), and 3) post processing (i.e., adjusting the prediction outcomes to ensure fairer decisions for those disadvantaged). Pre-processing strategies include for example omitting information about the under-represented group, representing features such that the under-represented group becomes indistinguishable, creating a more balanced training dataset. Creating a more balanced training dataset through oversampling underrepresented groups has been found to enhance accuracy of the algorithm for the underrepresented group while not negatively affecting accuracy of the algorithm for well represented groups (Sha et al., 2022). In-processing strategies include adding fairness requirements as a regularization term or constraint to the objective function, using an adversary model to minimize the impact of sensitive attributes, or fair representation learning (i.e., acquiring fairer embedding-based representations of the entities involved). Post-processing strategies refer to a broader range of strategies specific to the respective task the algorithm is trying to perform (L. Li et al., 2023, pp. 504–508). The authors also note from their review of the literature that most work had considered (representational) bias in the training data as the main cause for inaccurate decision making.

In our de-biasing analyses, we aim at informing how datasets for algorithmic training need to be set up with respect to diversity dimensions to mitigate representation and evaluation bias and hence ensure an accurate prediction and subsequently decision making process.

5.2.5 Research Questions

The underlying vision of our paper is the vision of a pluriverse; that is, a world where many worlds fit (Escobar, 2017; Kayumova & Dou, 2022; Mignolo, 2007). We envision a world in where all students have equal access to education, in particular physics education. This vision is rooted in critical theory meaning that it demands for analyses from a justice perspective, for example through diversity dimensions. As a concept, pluriverse comes with a normative focal point being the standpoint that distributions should not vary over diversity dimensions if these variations lead to discriminatory distributions of, for example, power. In the context of STEM education, a pluriverse needs to consider STEM identities of under-served students. For our study, we focus on de-biasing learning analytics algorithms in an example for learning the energy concept and in order to strengthen STEM identities.

In order to address evaluation bias, we aim at identifying threats of biases before the algorithms are actually trained. In order to do so, we train the same algorithmic architecture with the same student answers to predict positionalities on diversity dimensions, for example to predict students' gender. The idea is simple: If the algorithm is able to predict gender better, the student answers contain gendered patterns. These patterns could be used by the algorithm in the actual training as well. The greater the gendered patterns, the greater the potential threat of bias. If the algorithm is not able to predict gender, there is no certainty that no gendered patterns exist. However, we expect the probability for bias to be lower. We expect the bias of the actual algorithm used for learning to be higher for the diversity dimensions where the prediction of the positionality on the respective dimension is higher. We ask:

To what extent can threats of bias be identified by training an algorithm to predict positionalities on diversity dimensions?

In order to address representation bias, we seek to find a way to reduce bias based on training dataset configuration. From a feminist and de-colonial standpoint, we do not only ask whether the positionalities on diversity dimensions are represented as in the target population. Instead, we aim at finding out how training datasets need to be configured in terms of positionalities on diversity dimensions in order to end up with de-biased algorithms. For example, if we have a lot more students who predominantly speak German instead of another language at home, we do not control for actual target group population shares in our training datasets. Instead, we seek to find out how the representation needs to be configured in the training dataset in order to end up with an algorithm that works at least equally well for students who do not predominantly speak German at home. We ask:

How does dataset slicing effect the prediction results when grouping based on gender, most spoken language at home, or educational background of legal guardians?

General guidelines to address bias are developed (Cerratto Pargman & McGrath, 2021; Khalil et al., 2022; Sclater, 2014). However, Kitto and Knight have shown that these general guidelines often are not applied in practice (2019). In a theoretical work, we have shown where and how the guidelines are not concrete enough to be applied for the example of physics education (Grimm et al., 2023b). In this study, we aim at filling precisely that gap through concrete approaches to address evaluation and representation bias. Our goal is to inform effective and efficient ways of identifying threats of bias and reducing bias. Ultimately, we aim at informing evidence-based decision making on regulation and standardisation of de-biasing.

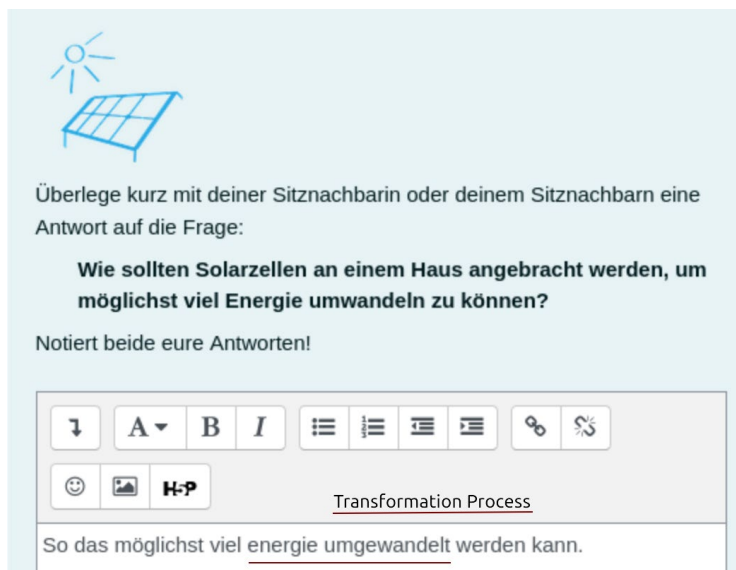
5.3 Methodology

In this work we focus on “under-served students”, meaning students who face historically grown inequalities and to whom the current physics education systems do not provide enough opportunities to develop their STEM identities. This lack of opportunities provided is what we refer to as under-served in contrast to well-served students with enough opportunities. In the case of our project, a bias in the selected algorithms would lead to incorrect labels for students’ competence. These wrong labels would then be displayed to teacher dashboards and guide interventions in biased directions. A false negative prediction could, for example, result in unwarranted negative feedback from teachers to female and non-binary students possibly leading to a lack of recognition and a weaker science identity of these students. A false positive prediction could result in a lack of support that would be necessary for effective learning. Therefore, we decide to use a bias measure that includes correct prediction of both, negative and positive cases.

5.3.1 Research Design

In order to address our research questions, we utilize data from the project “Learning Progression Analytics - Analyzing and Fostering Learning for the Development of Competence”. In the project, we implemented one of two so-called curriculum replacement units over five weeks in classes of grade seven or eight within northern Germany. Both units are implemented in a digital learning environment, have energy transformation as topic, and connect two domains within physics through the energy concept. The units differ in context: One unit is about solar cells with 36 items connecting optics and radiant energy with electricity and electric energy. The other unit is about laptops with 31 items connecting thermodynamics and thermal energy with electricity and electric energy. Each unit follows instruction in line with project-based learning and has one driving question which is split up into three sub-questions. For each sub-question, there is an experiment. The three experimental sessions are framed by an introductory session in the beginning and a summarising session in the end. Before and after the units, we conducted competence tests with the students.

In order to assess students’ competence with our items, we followed evidence-centered design. We formulated a student model with the competence that we want to assess split into sub-dimensions of knowledge, skills, and learning performances. From there, we formulated which evidence we would accept from a student’s answer to an item in order to label the answer with the respective competence element. The important point here is that we ended up by a set of labels on that we scored each student answer – depending on the item there might be one or multiple labels. The students’ answers were then scored by human raters. With the set of students’ answers and scores, we finally trained our algorithms. In Figure 5-3, you can see an example item from the unit on solar cells. For those who are further interested in the data collection and scoring process, the software scripts or the software versions we used, we provide detailed documentation in the supplemental material.



Überlege kurz mit deiner Sitznachbarin oder deinem Sitznachbarn eine Antwort auf die Frage:

Wie sollten Solarzellen an einem Haus angebracht werden, um möglichst viel Energie umwandeln zu können?

Notiert beide eure Antworten!

Transformation Process

So das möglichst viel energie umgewandelt werden kann.

Figure 5-3 - Example item, answer, label, and score – the item “How should solarcells be installed on a roof in order to transform as much energy as possible?” with the answer “in a way that as much energy can be transformed as possible” with the label “transformation process” scored positively as transformation process identified

We assessed gender and social class in the very end of the unit after conducting the post-test in order to not activate stereotype threats. We assessed gender as a diverse item, including options for male, female, non-binary identities, the option to write freely as how the student identifies, and the option to prefer not to answer (gender [gen]). For social class, we approached the diversity dimension with two items. We are aware that this is not a valid or reliable form in order to assess social class as a whole. However, we believe that the two items we chose are relevant in our context and can be used as valuable examples of whether our methods work or not in the context of social class, even though final conclusions on the relevance of social class as a whole cannot be drawn. Our two items include the most spoken language at home, assessed by providing options such as Turkish, German, and a free text field (most spoken language at home [lan]), and whether at least one of the legal guardians of the students holds an academic degree or not (educational background of legal guardians [edu]).

5.3.2 Data Base

The units were enacted in 22 classes with a total of 527 students. Since some of the teachers we recruited for participation in the original project enacted multiple units in different classes, the 22 classes were located at twelve different schools. The schools were representing the two main tracks in the German school system, which heavily relies on tracking based on academic achievement. Two schools (6 classes) were “Gemeinschaftsschulen” (englisch: community schools), representing the lower achieving of the two tracks, 10 schools (16 classes) were Gymnasium schools, representing the higher achieving and university-aiming track.

The student answers were scored by 4 researchers with interrater reliabilities’ median of Cohen’s Kappa values for all relevant labels bigger than 0.8. As not all students gave answers to items on diversity dimensions and/or items in the course, the full dataset is reduced for our use case. The number of students who answered the items on each diversity dimension are shown in Table 5-1 and the respective number of students’ answers that we have for each diversity dimension is shown in Table 5-2.

Table 5-1 - Numbers of Students per Diversity Dimension

diversity dimension	well-served	under-served	total
gender (binary included only)	130	133	263
social class (most spoken language at home)	242	40	282
social class (educational background of legal guardians)	159	51	210

Table 5-2 - Numbers of Student Answers per Diversity Dimension

diversity dimension	well-served	under-served	total
gender (binary included only)	1,350	1,391	2,741
social class (most spoken language at home)	2,606	381	2,987
social class (educational background of legal guardians)	1,724	560	2,284

We scored the student answers on nine different labels in total. Seven labels trace whether a student answer contains a certain knowledge element or not. Two labels trace whether a student answer contains a successful learning performance or not. For further explanations on the labelling process, please refer to the coding book in the supplemental materials.

For our slicing analysis, we train our models label-specific. One model is trained for each label. Additionally, we perform the slicing analyses specific for each of the three dimensions, one dimension for each item. Hence, we end up with 27 slicing analyses resulting from three diversity dimensions times nine labels. In Table 5-3, we show the available student answers on the level of each of that 27 models. We further distinguish the available positive and negative samples for the labels for each group, well-served and under-served students in the respective diversity dimension.

5 De-Biasing

Table 5-3 - Numbers of Student Answers per Diversity Dimension Well-Served/Under-Served per Label Positive/Negative; Abbreviations: lan – Language, gen – Gender, edu – Educational Background; ws – Well-Served, us – Under-Served; pos – Positive, neg – Negative; MEE – Manifestation Electric Energy, MEv – Manifestation Electric variable, MRE – Manifestation Radiant Energy, MRv – Manifestation Radiant variable, MTE – Manifestation Thermal Energy, MTv – Manifestation Thermal variable, TP – Transformation Process, M_lp – Manifestation Learning Performance, T_lp – Transformation Learning Performance, colour scheme groups relevant numbers for one model training (27 models in total)

Case	MEE	MEv	MRE	MRv	MTE	MTv	TP	M_lp	T_lp
lan- ws- pos	171	95	319	695	68	199	247	165	192
lan- ws- neg	63	651	820	443	131	50	815	1158	855
lan- us- pos	491	9	51	103	5	6	37	22	30
lan- us- neg	189	87	134	83	4	6	119	171	124
gen- ws- pos	434	47	147	340	25	89	120	74	93
gen- ws- neg	146	328	446	255	68	23	420	607	441
gen- us- pos	411	49	179	367	40	107	141	93	110
gen- us- neg	128	347	432	242	67	29	439	616	461
edu- ws- pos	119	62	234	462	40	132	175	119	143
edu- ws- neg	37	412	491	265	95	36	506	729	525
edu- us- pos	796	22	73	153	17	37	60	46	43
edu- us- neg	265	147	181	100	24	14	174	246	188

5.3.3 Data Analysis Procedure

We have one algorithmic architecture with two biases to address and three criteria that we can use to represent diversity dimensions – ending up with two times three analyses to address bias. In order to understand the individual datasets, we first provide project

context, then describe the full dataset, and finally specify which part of the full dataset we used for the slicing analyses and the training dataset analyses respectively.

5.3.3.1 Identifying Threats of Bias: Training Dataset Analysis

In order to address evaluation bias, we implement training dataset analysis. The idea is to train the same model architecture, in our case a transformer model, not on predicting student's performance but on predicting the diversity dimension. As exemplary shown in Figure 5-4, we predict the diversity dimension gender instead of predicting the knowledge about electric energy as part of students' competences. If we find the algorithm to be able to predict gender, we know that the students' answers to contain what we call gendered patterns. In our case, female students' answers would contain patterns that distinguish them from male students' answers.

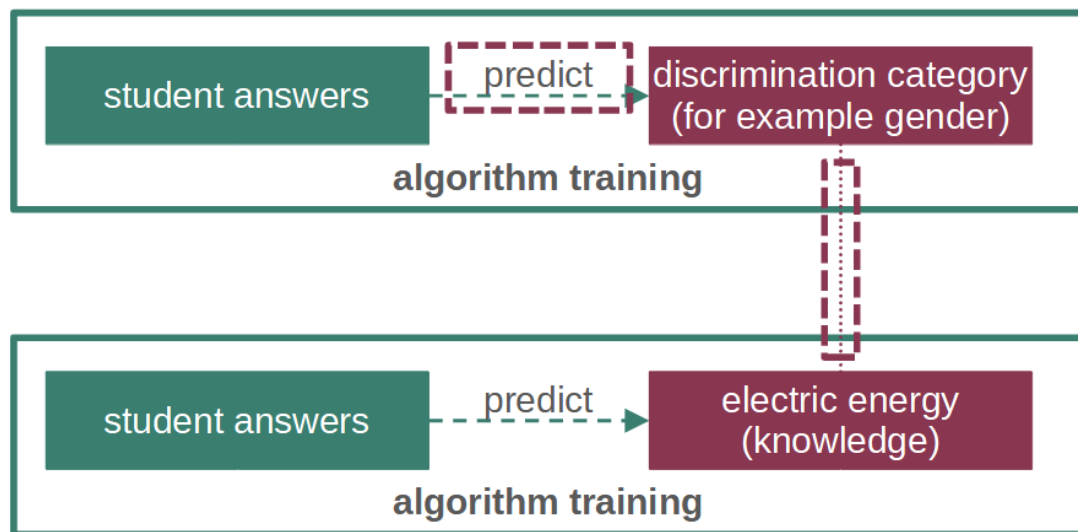


Figure 5-4 - Identifying Threats of Bias: Training Dataset Analysis

For the training dataset analysis, we use the student answers only and aim at predicting the diversity dimensions. We split the dataset in five splits and then train five models – each split is four times part of the training data and one time forms the testing data. We balance the splits so that each split has as many under- as well-served student answers. For balancing, we use a combined method with over-sampling for the under-served and under-sampling for the well-served student answers. We do so as the combined method for balancing yielded the most reliable predictions on unseen data, thereby preventing overfitting the best. With our method, we yield four result scores which are Cohen's kappa, F1-scores¹⁸, precision¹⁹, and recall²⁰. We use Cohen's kappa here instead of quadratic weighted kappa²¹ for its more intuitive interpretability – which is more important for a first

¹⁸ F1-scores are a mix of precision and recall: two times precision times recall over the sum of precision and recall.

¹⁹ Precision is an answer to the question: From all the identified positives, how sure can I be that the artefact is actually positive? It is calculated as true positives over the sum of true and false positives.

²⁰ Recall is an answer to the question: From all actually positive artefacts, how many does my model identify? It is calculated as true positives over the sum of true positives and false negatives.

²¹ Quadratic weighted kappa is a score to measure how much prediction outperforms random guessing for a specific dataset and task. It is calculated as difference between empirical probability

risk analysis than it is in the more in-depth slicing analysis where quadratic weighted kappa is already well established. We report the arithmetic means for the five models of Cohen's kappa, F1-score, precision, and recall. The kappa scores allow us to interpret whether a prediction beyond mere guessing is possible and how reliably we get that prediction. The F1-score, precision, and recall qualify then how well which student answers are predicted. The interpretation of the scores of all sub-folds remains similar and only varies in effect sizes which is why we do not report them in detail and only mention them here as indicator for reliable results.

5.3.3.2 *Reducing Bias: Slicing Analysis*

In order to address representation bias, we implement a slicing analysis inspired by the proposal of Gardner and colleagues and as shown in Figure 5-5 (2019). The goal is to evaluate how the training datasets need to be set up and who needs to be represented with which share in order to yield satisfactory results. With regard to representation, we focus on diversity dimensions through the examples of gender and social class which we operationalised by three criteria:

- gender
- language that is spoken most at home
- educational background of legal guardians, if at least one of them holds an academic degree

With “satisfactory results”, we refer in our example to de-biased for the specific bias under observation. For all criteria, we split the data in under-served and well-served students – reflecting whether the group faces historically grown inequalities and the system-inherent self-reproduction of these inequalities as discrimination or as privilege. For the example of gender, we check whether training the algorithm only with answers of well-served (here: male) students effects the accuracy for prediction results for under-served (here: female) students. We vary the percentage of under- and well-served students in the training dataset from 0 % to 100 % in 10 %-steps. For slicing, we decide to split on the student level. As one student has more than one answer and training is done on an answer-level, the datasets might vary in size. The training datasets could be further stratified for number of student answers or, as another example, the rating student who could have a decision makers' standpoint bias, which we do not do in our analyses in order to keep the variations more understandable and interpretable.

of agreement and expected agreement with random assignment over the difference between one and the expected agreement with random assignment.

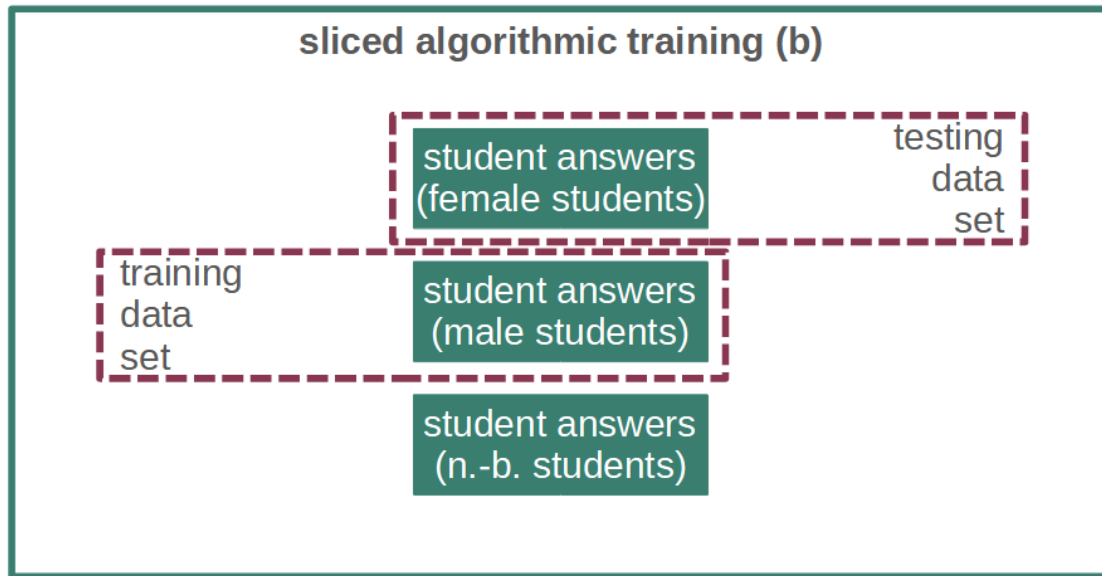


Figure 5-5 - Reducing Bias: Slicing Analysis

For each of these 27 models, we performed the slicing analyses with eleven variations of the training and testing datasets. We split the datasets eleven times for each diversity dimension with under-served students' share in the training dataset reaching from 0 % up to 100 % in 10 % steps. 27 models times eleven splits result in 297 cases for bias evaluation. We stratified label distributions in terms of positive and negative scores for all splits. For each case, we quantified F1-scores, precision, recall, and quadratic weighted kappa for the whole group as well as the sub-groups of well- and under-served students. For bias evaluation, we refer to bias on F1-scores – using the other scores for further grounding of our interpretations and discussions where needed.

5.3.4 Algorithmic Architecture

In terms of the EU AI Act, we use both general-purpose AI and high-risk systems. In terms of general purpose, we use pre-trained language models and fine-tune these models to our specific use case. For choice of algorithmic architecture, we used the criteria that 1) the models and libraries are frequently used, and 2) the models and libraries are used in many application fields. We built on previous experience in our research project and could therefore use the architecture that we had actually used outside of the context of diversity research (Gombert et al., 2022). We simplified the architecture and lost a bit of cutting-edge technology in order to meet our criterion of frequently used models and libraries. As we are in the application of natural language processing, the models are used across all educational domains and hence are well suited for our second criterion. For those readers who are interested in more details, we used the libraries G-BERT-large (Chan et al., 2020) and Huggingface transformers (Wolf et al., 2020).

5.4 Results

We decide to kick off the results with a very relevant limitation of our findings: We cannot generalise from our evidences alone. Our main contribution is the clear and transparent definition of the process. Our evidences can be first indications and together with many more evidences, they can inform decision making on regulation. However, we provide evidence only for 1) physics education, 2) a specific region of the world, 3) the diversity dimensions gender and (partly) social class, 4) 7th and 8th grade in school, and 5) the

specific topic of energy education – to name some examples. We believe results to be generalisable to a certain extent, as mechanisms can be similar. Nonetheless, this one contribution is far from being enough to draw solid conclusions for political decision making. As our training dataset analysis shall inform the identification of threats of bias, we start with a presentation of our results on the slicing analysis and then present our training dataset analysis, including a discussion in front of the findings from the slicing analysis.

5.4.1 Reducing Bias: Slicing Analysis

In this sub-section, we present our evidences to the research question:

How does dataset slicing effect the prediction results when grouping based on gender, most spoken language at home, or educational background of legal guardians?

The input of 297 models for training resulted in 132 models that reached a quadratic weighted kappa of bigger than 0.6. Hence, the results presented in the following rely on 132 models only. The quadratic weighted kappa is low for the labels where 1) we have less student answers (for example thermal energy and thermal variables), 2) the label itself holds more content than others (whereas electric energy contains energy constructs only, electric variable contain electric current, voltage, and power – the same holds true for radiant and thermal variable), and 3) the label is of a more complex nature (learning performances). Having small quadratic weighted kappas for these values is thus explainable. Still, the remaining 132 models allow for meaningful slicing analyses in the context of our research question.

5.4.1.1 Gender

In Figure 5-6, all 48 bias values for gender are shown with regard to prediction accuracy. We looked at the biases in prediction accuracy as differences in F1 scores between well- and under-served students with positive values representing higher F1 scores for the well-served students.

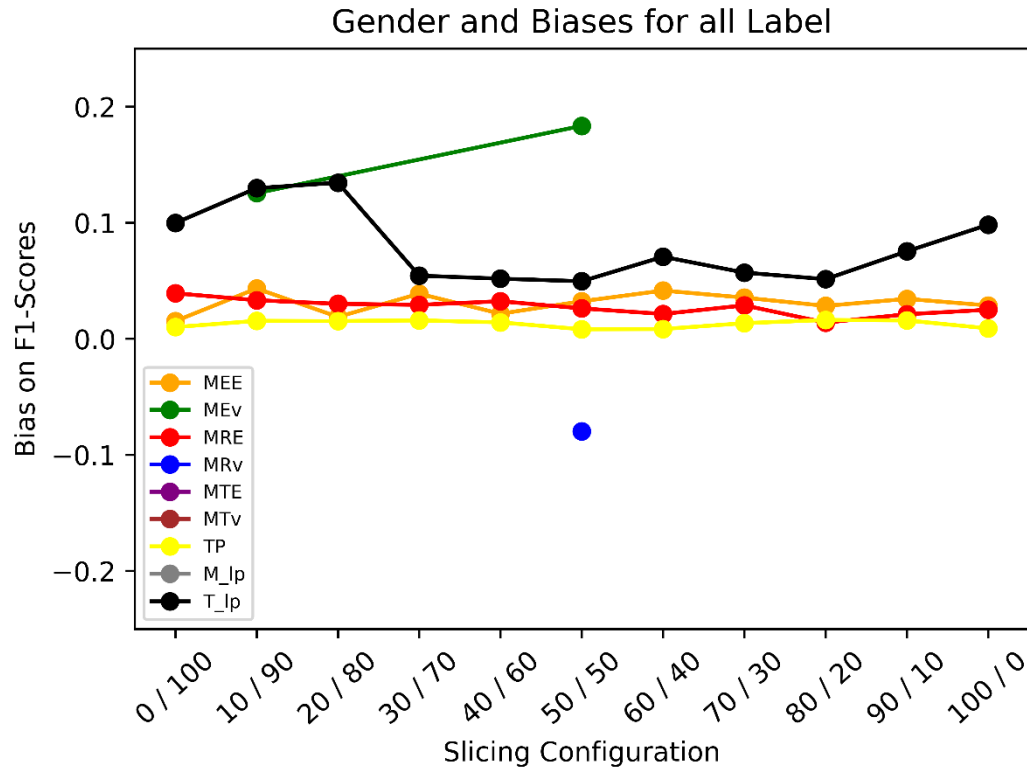


Figure 5-6 - Gender and Biases for all Label; MEE – Manifestation Electric Energy, MEv – Manifestation Electric variable, MRE – Manifestation Radiant Energy, MRv – Manifestation Radiant variable, MTE – Manifestation Thermal Energy, MTv – Manifestation Thermal variable, TP – Transformation Process, M_lp – Manifestation Learning Performance, T_lp – Transformation Learning Performance

Biases on F1-scores along all slicing configurations for gender reach from -0.080 up to 0.261 with a mean of 0.045 and a median of 0.030. Gender is the only diversity dimension in which we found quadratic weighted kappas bigger than 0.6 for the label of “MEv – Manifestation Electric variable” and “MRv – Manifestation Radiant variable” – though there are still only three values for both labels in total. High kappa values indicate a well-functioning prediction of the respective label and are outside of bias considerations a measure to decide whether a model works or not. Having no values with kappa bigger than 0.6 for the other diversity dimensions means that gender is the only diversity dimension for which we can perform a bias analysis for these labels. Additionally, within the diversity dimension gender also is the only value among all 132 values for biases along all diversity dimensions under observation that is negative – according to our definition of bias and further discriminating under-served students, the only model without bias. In other words: From 297 calculated models, only 132 are considered to work and therefore used for bias analyses. From these 132 models, 131 have a bias: The prediction accuracy for the well-served students is better than the one for the under-served students. No matter what the slicing configuration is, it cannot completely prevent biases in our cases, neither for gender nor for the other diversity dimensions.

There are no clear effects of gender-based dataset slicing on the prediction results in terms of bias on F1-scores. This gender-based evidence supports the claim that balancing training data does not necessarily prevent bias. Differing from Latif and colleagues, our evidence for gender-based slicing does not indicate that balanced datasets outperform imbalanced datasets in terms of bias (2023). Instead, our evidence on the diversity dimension gender indicates that training dataset configuration is not the most relevant

screw to prevent bias. The finding our evidence indicates is: Balancing training datasets does neither prevent nor introduce biases.

However, we find clear indication for existing biases that cannot be explained with the slicing configuration: With a mean of 4.5 % and a median of 3.0 %, our evidence clearly indicates existing biases along the diversity dimension gender. Together with the historically grown inequalities in the world as it is (El-Mafaalani, 2021; Rosa & Moore Mensah, 2016), our evidence indicates a need to further investigate the prevention of bias beyond training dataset configuration.

5.4.1.2 Social Class – Most Spoken Language at Home

In Figure 5-7, all 40 bias values for language are shown.

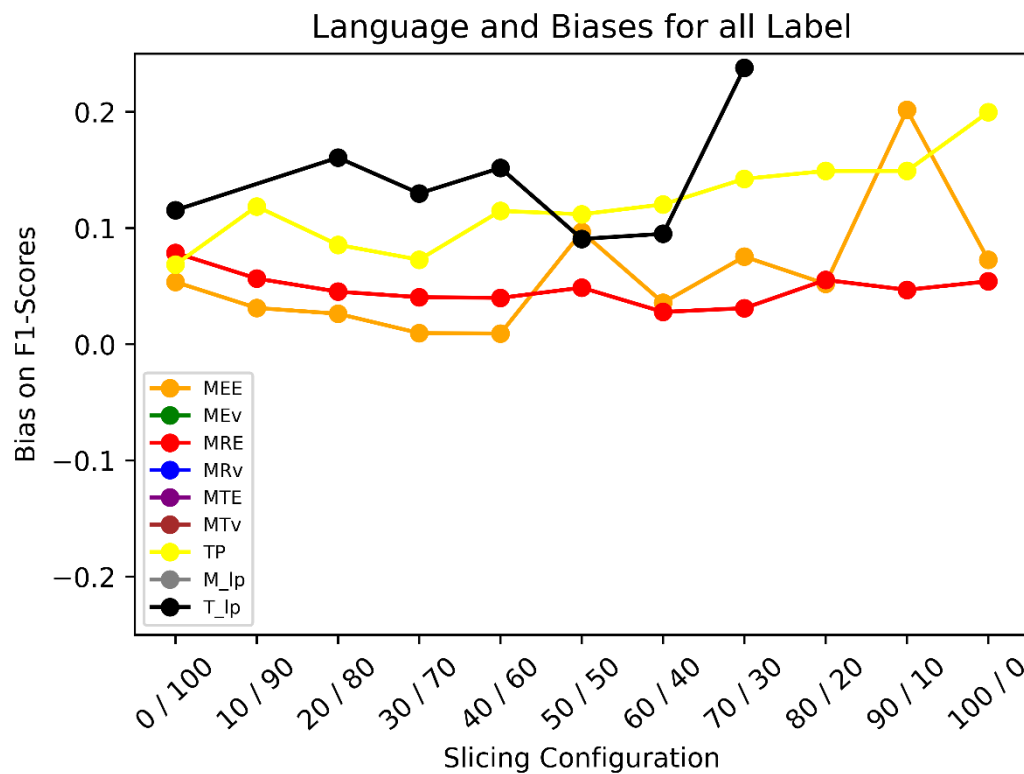


Figure 5-7 - Language and Biases for all Label; MEE – Manifestation Electric Energy, MEv – Manifestation Electric variable, MRE – Manifestation Radiant Energy, MRv – Manifestation Radiant variable, MTE – Manifestation Thermal Energy, MTv – Manifestation Thermal variable, TP – Transformation Process, M_lp – Manifestation Learning Performance, T_lp – Transformation Learning Performance

Biases on F1-scores along all slicing configurations for most spoken language at home reach from 0.009 up to 0.238 with a mean of 0.087 and a median of 0.074.

Differing from our evidence from gender-based slicing, language-based dataset slicing has an effect on the prediction results in terms of bias on F1-scores. None of the slicing configurations leads to the elimination of bias. Still, slicing configurations yield increasing scores: For example, the average means of the slicing configurations 0 / 100 to 20 / 80, 40 / 60 to 60 / 40, and 80 / 20 to 100 / 0 increase for MEE (0.037, 0.047, 0.109), TP (0.091, 0.116, 0.166), and T_lp (0.135, 0.141, -) while only those for MRE do not show that trend (0.060, 0.039, 0.052). For MEE as an example, there is a difference in bias from 3.7 % to 10.9 % for the average mean of the respective three slicing configurations. Our evidence here is well in line with the findings from Latif and Colleagues (2023) that balanced datasets outperform imbalanced datasets in terms of bias and that algorithmic biases can be

prevented by thoughtful configuration. The finding our language-based slicing indicates is: Balancing training datasets prevents some biases while not introducing new biases.

Similar to our findings on gender-based slicing, we find clear indication for existing biases that cannot be explained with the slicing configuration: With a mean of even 8.7 % and a median of 7.4 % (compared to 4.5 % and 3.0 % for gender-based slicing), our evidence clearly indicates existing biases depending on the most spoken language at home. Our evidence indicates a need to further investigate the prevention of bias beyond training dataset configuration.

5.4.1.3 Social Class – Educational Background of Legal Guardians

In Figure 5-8, all 44 bias values for educational background are shown.

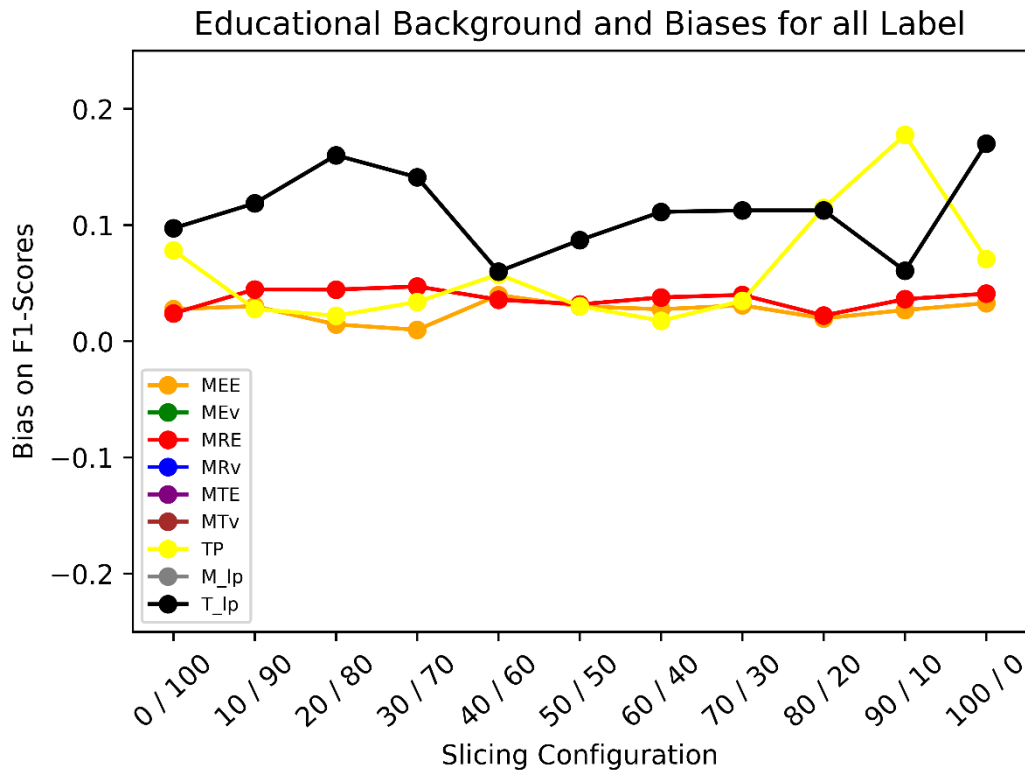


Figure 5-8 -Educational Background and Biases for all Label; MEE – Manifestation Electric Energy, MEv – Manifestation Electric variable, MRE – Manifestation Radiant Energy, MRv – Manifestation Radiant variable, MTE – Manifestation Thermal Energy, MTv – Manifestation Thermal variable, TP – Transformation Process, M_lp – Manifestation Learning Performance, T_lp – Transformation Learning Performance

Biases on F1-scores along all slicing configurations for educational background of legal guardians reach from 0.010 up to 0.178 with a mean of 0.059 and a median of 0.039.

There are no clear effects of education-based dataset slicing on the prediction results in terms of bias on F1-scores. As for gender-based slicing, our findings indicate that balancing training datasets does neither prevent nor introduce biases.

As for both gender- and language based-slicing, we find clear indication for existing biases that cannot be explained with the slicing configuration: With a mean of 5.9 % and a median of 3.9 %, our evidence clearly indicates existing biases depending on the educational background of legal guardians. Our evidence indicates a need to further investigate the prevention of bias beyond training dataset configuration.

5.4.2 Identifying Threats of Bias: Training Dataset Analysis

In this sub-section, we present our evidences to the research question:

To what extent can threats of bias be identified by training an algorithm to predict positionalities on diversity dimensions?

5.4.2.1 Gender

Table 5-4 - Results for Gender in Training Dataset Analysis

Score – Arithmetic Means	Under-Served Students	Well-Served Students
kappa	0.126	
F1	0.492	0.588
precision	0.612	0.548
recall	0.456	0.670

In Table 5-4, the results for gender in our training dataset analysis are shown. The kappa value between 0.0 and 0.2 indicates success beyond mere guessing – even if the prediction is not reliable.

Prediction capability beyond 0.0 tells us that gendered patterns exist in the student answers. Bias is possible. At the same time, gendered patterns in input data do not necessarily lead to a bias in the output. The presented effects also remained in testing with unseen data.

In the light of the slicing analysis where we found biases as well, the need for special care needs to be underlined. In our case, it seems that the algorithm uses the gendered patterns even if trained for label prediction. Hence, the training dataset analysis would have been a valuable risk management measure. However, we need to stress that this case is not generalisable in terms of all high kappa and F1-scores in training dataset analyses need to correlate with biased algorithms. It is well possible that algorithms do not use the gendered patterns and have no bias even though strong gendered patterns exist. Also, a careful dataset configuration would not have led to a de-biased algorithm in our case – other measures need to be found.

5.4.2.2 Social Class – Most Spoken Language at Home

Table 5-5 - Results for Language in Training Dataset Analysis

Score – Arithmetic Means	Under-Served Students	Well-Served Students
kappa	0.789	
F1	0.896	0.892
precision	0.855	0.941
recall	0.943	0.849

In Table 5-5, the results for language in our training dataset analysis are shown. The kappa value between 0.6 and 0.8 indicates success beyond mere guessing – with remarkably reliable predictions.

Again, prediction capability tells us that language patterns exist in the student answers. Bias is possible but not a necessary consequence. Given the remarkably well predictions, special care for not using precisely these patterns is necessary. However, we need to stress that the remarkably well predictions might be due to overfitting as the evaluation with unseen data revealed prediction capability that was worse but still existing. The testing with unseen data is an indication and no proof as we tested with very little unseen data (below 100 student answers) which is why we still carefully interpret the results presented in Table 5-5.

In the light of the slicing analysis where we found slicing effects for language-based slicing, the need for special care needs to be underlined. In our case, it seems that the algorithm uses the language patterns even if trained for label prediction. An intervention in terms of careful dataset configuration can de-bias the algorithm. We need to stress that this case is not generalisable in terms of all high kappa and F1-scores in training dataset analyses need to correlate with impact of slicing analyses. It is well possible that algorithms do not use the language patterns and have no bias even though strong language patterns exist. What we want to highlight is that in our case, the risk assessment through training dataset analyses could have effectively informed where further slicing analyses are needed. That is precisely the decision that we want to inform: Would a political regulation for training dataset analyses make sense in terms of directing further risk assessment and use of resources? Our evidence can only be seen as a first sign and should not be used to draw conclusions already. Nonetheless, it is a first indication. Whether that indication can be backed by further evidences or not needs to be shown in the future.

5.4.2.3 Social Class – Educational Background of Legal Guardians

Table 5-6 - Results for Educational Background in Training Dataset Analysis

Score – Arithmetic Means	Under-Served Students	Well-Served Students
kappa	0.160	
F1	0.348	0.724
precision	0.627	0.617
recall	0.266	0.886

In Table 5-6, the results for educational background in our training dataset analysis are shown. The kappa value between 0.0 and 0.2 indicates success beyond mere guessing – even if the prediction is not reliable.

Prediction capability beyond 0.0 tells us that patterns along educational background exist in the student answers. Bias is possible. At the same time, patterns along educational background in input data do not necessarily lead to a bias in the output. We also need to stress that prediction capability vanished when predicting unseen data for educational background which is an indication for overfitting. It is an indication and no proof as we tested with very little unseen data (below 100 student answers) which is why we still carefully interpret the results presented in Table 5-6.

Comparing our results of training dataset analyses with our slicing analyses, we find the same pattern as before: Existing biases in the slicing analyses could be traced back in patterns in the student answers along educational background. Little reliability of prediction in the training dataset analyses correlates with little impact of slicing configurations. This evidence is another indication that first risk assessment through training dataset analysis can work.

5.5 Discussion of Implications

In this section, we make our results tangible to three different communities by discussing implications specifically for 1) education researchers, 2) political decision makers, and 3) learning analytics researchers.

5.5.1 Implications for Education Researchers

We need education researchers who onboard to discourses around de-biasing algorithmic decision making. Algorithmic decision making comes with great potentials for learning. By no means we aim at downplaying this fact. A good education is a basic right protected in Germany for example through the constitution and at UN level through a Sustainable Development Goal. As community, we need to unpack this potential. At the same time, algorithmic decision making introduces new threats that can reproduce and strengthen historically grown inequalities. However, algorithmic decision making is not inherently biased – how biases can be prevented is an empirical question that can be answered (Latif et al., 2023). We know from empirical evidence that de-biasing does not necessarily sacrifice predictive accuracy (L. Li et al., 2023, p. 506). Our results indicate that balancing training datasets can be a valuable contribution but alone cannot address all bias issues. Within the education research communities, we build on strong theories to address inequalities, for example identity research in STEM education (Kayumova & Dou, 2022). The theory of the pluriverse with its vision of “a world where many worlds fit”, the clear allocation of responsibility with the structures and not the individuum, as well as the reconfiguration screws are useful tools in the discussion around de-biasing as we have shown in our paper. It is crucial and with the pluriverse concept possible to build these discussions on theory established by Black feminist scholars such as the matrix of domination (Collins, 1990), intersectionality (Crenshaw, 1989), and feminist standpoint theory (Costanza-Chock, 2020, pp. 9–10). These theoretical contributions need to be added from education research communities to the discourses around de-biasing and added to the existing critical discourses in the field.

5.5.2 Implications for Political Decision Makers

We need political decision makers who aim at setting a regulatory frame for de-biasing algorithmic decision making. We see many funding programmes and grants to unfold the potentials of algorithmic decision making – which we highly appreciate. However, de-biasing practice and the impact of ethical principles have not found their way into practice until today (Kitto & Knight, 2019). In a previous study we have outlined where exactly politically given guidance is missing in order to reach practice: a clear definition of bias, explicit diversity dimensions to analyse for, and a well-defined process of evaluation including evaluation criteria (Grimm et al., 2023b). This lack is problematic, as protection against discrimination is an equally strong individual right of each student as the right on education itself. It gets increasingly problematic in domains such as STEM education where already today various historically grown inequalities exist. Positions on diversity dimensions have a strong impact on whether you become a STEM person or not, whether

you build a STEM identity or not, and whether you choose a STEM career or not. Algorithmic decision making comes with threats to precisely reproduce these historically grown inequalities. At the same time, there are strong economic interests articulated by powerful, global companies to bring learning analytics to schools. If we do not want to risk to unfold the threats, we need to turn towards concepts such as responsible learning analytics and start building a regulatory frame for de-biasing that addresses the problems where they occur. For example, representation and evaluation bias could be addressed by methods such as slicing and training dataset analyses as proposed in our study. The findings from this study can be first evidence as well as a valuable methodological and theoretical contribution. Nonetheless, much more evidence is needed – for other constructs than energy in physics education, for other domains, for other age groups, and for more diversity dimensions. Political decision makers need to 1) provide funding for and give direction to research to gather further evidence where regulation is most useful and 2) start building up a regulatory frame that brings de-biasing into practice.

5.5.3 Implications for Learning Analytics Researchers

We need learning analytics researchers to focus on bringing de-biasing into practice. The learning analytics from its beginnings has a long and strong history in addressing ethical issues (Prinsloo & Slade, 2017). The most pressing task for the community, from our perspective, is to bring ethics into practical political decision making and to manifest it in a regulatory framework. The focus needs to be on de-biasing strategies and their success in comparison. Li and colleagues already reviewed the work on de-biasing which they call “fairness-enhancing strategies”, and we need to build on the existing work in the future (2023). It is the task of the community to provide evidences that inform political decision making, that holds true for de-biasing as well. As a community, we need to provide conclusions based on a set of different evaluation criteria and bias definitions that political decision makers can choose from according to their political preferences. We cannot make the political decisions – but we can inform for a given combination of evaluation criteria and bias definition which are the best regulation strategies. That is an empirical question. However, the big global economic players are not going to address them unless they need to and next to our research communities, not many actors exist who have the resources to inform political decision making on the regulation of de-biasing. There are questions that need to be answered: How much diversity need the training datasets for successful de-biasing? How many students per position on each diversity dimension do we need for valid and reliable results? Which diversity dimension comes with the biggest risk of bias for a specific domain? Is the regulation of training datasets 1) a successful and 2) the most effective way to prevent biases? Additionally, we need to highlight the limits of our research. De-biased algorithms are not bias-free. There exist multiple bias entry points and we never address all of them – that is not even the goal. The goal from a pluriversal standpoint is to address enough of them in order to reach de-biased outcomes on a macro level. For STEM education as an example, that goal translates as “STEM identity development does not depend on a position on any diversity dimension”. We need to carefully reflect and decide how to assess our data, positions on diversity dimensions for instance. Gender can be assessed – as in our example – with more than the binary options. If we assess multiple gender identities, for how many of them do we need to de-bias? As a learning analytics community, we do not need to answer this political question. However, we need to point at the existence of this political question and describe the implications it has. Most important, we need to measure our success as research community not only in terms of theoretical contributions but to closely monitor and steer our impact on practice.

5.6 Conclusions

5.6.1 De-biasing through regulation of training datasets – a promising starting point

Regulating training datasets seems to be a promising starting point. Given the few existing evidences, it is too early for final conclusions or recommendations. Nonetheless, existing evidence as well as the evidence in our study support the claim that regulating training datasets can successfully prevent biases in algorithmic decision making. However, our results indicate existing biases that cannot be addressed by the regulation of training datasets alone. Additionally, our findings do not answer the question whether the regulation of training datasets is the most effective way in terms of preventing bias and enabling learning. De-biasing through regulation of training datasets seems to be a promising piece of a regulatory framework that most effectively prevents threats in terms of discrimination and enables both, potentials in terms of learning and diversity mainstreaming in physics education.

5.6.2 6Needs Addressing Researchers and Politicians

We need 1) education researchers who onboard to discourses around de-biasing algorithmic decision making and contribute strong theoretical frameworks such as the pluriverse and identity development. We need 2) political decision makers who aim at setting an evidence-informed regulatory frame for de-biasing. Finally, we need 3) learning analytics researchers who focus on bringing de-biasing into practice through a) methodological innovation and aiming at actionable results in terms of political decision making, and b) listening for evaluation criteria beyond competence A/B-testing that allow for successfully reaching diversity. Such evaluation criteria can be theoretical contributions formulated by education researchers with identity development and pluriversal perspectives.

Author Contributions

Authorship – the earlier the name the bigger the contribution for the particular point if no specific contributions are listed: Adrian Grimm (A.G.), Sebastian Gombert (S.G.), Silvio Armbrüster (S.A.), Marcus Kubsch (M.K.), Anneke Steegh (A.S.), Marianela Navarro Camacho (M.N.C.), Hannah Kolbe (H.K.), Simon Tautz (S.T.), Karoline Petersohn (K.P.), Valentin Holst (V.H.), Onur Karademir (O.K.), Isabell Bohm (I.B.), Knut Neumann (K.N.)

- Conceptualization: A.G., M.K., A.S., K.N.;
- Project Operational Coordination Work: A.G., O.K., I.B., S.G.;
- Data Collection: A.G., O.K., I.B., H.K., S.T., K.P.;
- Data Scoring for Algorithmic Training: H.K., K.P., V.H.;
- Documentation: H.K., S.T., K.P., A.G., S.A., S.G.;
- Discussions about Relevant Theory: A.G., A.S., M.K., K.N., M.N.C.;
- Methodology & Results Preparation: S.A. (training dataset analyses), S.G. (slicing analyses), A.G. (data pre- and post-processing, further code-commenting);
- Original Draft Preparation: A.G.;
- Writing-Review and Editing: A.G., K.N., A.S., M.K., M.N.C.;
- Mentoring: A.S., M.K., M.N.C.;
- Project Management: A.G., M.K., K.N.;
- Funding Acquisition: K.N., M.K.;

All authors have read and agreed to the submitted version of the manuscript.

Acknowledgments

We understand our work as a tiny contribution to the massive house of science built by scientists before us who we admire, especially: pluriverse thinkers such as Escobar and Mignolo as well as Black feminist thinkers such as hooks and Collins. Also, we want to point out that our work is as quantitatively measurable output that strongly builds on the support of administrative staff at IPN who often is neither mentioned nor receives the merits their work is worth. Finally, we want to highlight our privilege of having both access to so much scientific work and the support of a large administrative department – a too often forgotten privilege, from our perspective.

References of the Piece of Scholarship

- Alexandron, G., Yoo, L. Y., Ruipérez-Valiente, J. A., Lee, S., & Pritchard, D. E. (2019). Are MOOC Learning Analytics Results Trustworthy? With Fake Learners, They Might Not Be! *International Journal of Artificial Intelligence in Education*, 29(4), 484–506. <https://doi.org/10.1007/s40593-019-00183-1>
- Archer, L., Dawson, E., DeWitt, J., Seakins, A., & Wong, B. (2015). “Science Capital”: A Conceptual, Methodological, and Empirical Argument for Extending Bourdieusian Notions of Capital Beyond the Arts. *Journal of Research in Science Teaching*, 52(7), 992–948. <https://doi.org/10.1002/tea.21227>
- 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>
- Avraamidou, L. (2020). Science identity as a landscape of becoming: Rethinking recognition and emotions through an intersectionality lens. *Cultural Studies of Science Education*, 15, 323–345. <https://doi.org/10.1007>
- Bachsleitner, A., Lämmchen, R., & Maaz, K. (Eds.). (2022). *Soziale Ungleichheit des Bildungserwerbs von der Vorschule bis zur Hochschule: Eine Forschungssynthese zwei Jahrzehnte nach PISA*. Waxmann. <https://doi.org/10.31244/9783830996248>
- Baker, R., & Hawn, A. (2021). *Algorithmic Bias in Education*. <https://doi.org/10.1007/s40593-021-00285-9>
- Bergmann, U., Bonefeld-Dahl, C., Dignum, V., Gagné, J.-F., Metzinger, T., Petit, N., Steinacker, S., Van Wynsberghe, A., & Yeung, K. (Eds.). (2019). *Ethics Guidelines for Trustworthy AI*. European Commission - High-Level Expert Group on Artificial Intelligence. <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>
- Bodnar, K., Hofkens, T., Wang, M.-T., & Schunn, C. (2020). Science Identity Predicts Science Career Aspiration Across Gender and Race, but Especially for Boys. *International Journal of Gender, Science, and Technology*. <https://www.semanticscholar.org/paper/Science-Identity-Predicts-Science-Career-Aspiration-Bodnar-Hofkens/ecb0782dd8eb07357c22d37d1eb0af13474d93b1>
- Bolukbasi, T., Chang, K.-W., Zou, J. Y., Saligrama, V., & Kalai, A. T. (2016). Man is to Computer Programmer as Woman is to Homemaker? Debiasing Word Embeddings. *Advances in Neural Information Processing Systems*, 29. https://papers.nips.cc/paper_files/paper/2016/hash/a486cd07e4ac3d270571622f4f316ec5-Abstract.html
- Butler, J. (1990). *Gender Trouble: Feminism and the Subversion of Identity*. Routledge.
- Calabrese Barton, A., Kang, H., Tan, E., O'Neill, T. B., Bautista-Guerra, J., & Brecklin, C. (2013). Crafting a Future in Science: Tracing Middle School Girls' Identity Work Over Time and Space. *American Educational Research Journal*, 50(1), 37–75. <https://doi.org/10.3102/0002831212458142>
- Carlone, H. B., & Johnson, A. (2007). Understanding the Science Experiences of Successful 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
- Chan, B., Schweter, S., & Möller, T. (2020). German's Next Language Model. Proceedings of the 28th International Conference on Computational Linguistics, 6788–6796. <https://doi.org/10.18653/v1/2020.coling-main.598>
- 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>
- Çolakoğlu, J., Steegh, A., & Parchmann, I. (2023). Reimagining informal STEM learning opportunities to foster STEM identity development in underserved learners. *Frontiers in Education*, 8, 1–16. <https://doi.org/10.3389/feduc.2023.1082747>
- Collins, P. H. (1990). *Black Feminist Thought: Knowledge, Consciousness and the Politics of Empowerment*. <https://doi.org/10.4324/9780203900055>
- Costanza-Chock, S. (2020). *Design justice: Community-led practices to build the worlds we need*. The MIT Press.
- Crenshaw, K. (1989). Demarginalizing the Intersection of Race and Sex: A Black Feminist Critique of Antidiscrimination Doctrine, Feminist Theory and Antiracist Politics. *University of Chicago Legal Forum*, 1989(8), 139–167.
- Dennis, M., Masthoff, J., & Mellish, C. (2016). Adapting Progress Feedback and Emotional Support to Learner Personality. *International Journal of Artificial Intelligence in Education*, 26(3), 877–931. <https://doi.org/10.1007/s40593-015-0059-7>
- Diakopoulus, N., Friedler, S., Arenas, M., Barocas, S., Hay, M., Howe, B., Jagadish, H. V., Unsworth, K., Sahuguet, A., Venkatasubramanian, S., Wilson, C., Yu, C., & Zevenbergen, B. (2021). Principles for Accountable Algorithms and a Social Impact Statement for Algorithms. *FAT/ML*.
<https://www.fatml.org/resources/principles-for-accountable-algorithms>
- Doroudi, S., & Brunskill, E. (2019). Fairer but Not Fair Enough On the Equitability of Knowledge Tracing. Proceedings of the 9th International Conference on Learning Analytics & Knowledge, 335–339. <https://doi.org/10.1145/3303772.3303838>
- 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>
- Dressel, J., & Farid, H. (2018). The accuracy, fairness, and limits of predicting recidivism. *Science Advances*, 4(1), eaao5580. <https://doi.org/10.1126/sciadv.aao5580>

- Düchs, G., & Ingold, G.-L. (2018). Frauenanteil bleibt stabil. *Physik Journal*, 17(8/9), 32–37.
- 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>
- Escobar, A. (2017). *Designs for the Pluriverse: Radical Interdependence, Autonomy, and the Making of Worlds*. Duke University Press.
<http://www.jstor.org/stable/j.ctv11smgs6>
- Fletcher, R. R., Nakeshimana, A., & Olubeko, O. (2021). Addressing Fairness, Bias, and Appropriate Use of Artificial Intelligence and Machine Learning in Global Health. *Frontiers in Artificial Intelligence*, 3, 561802.
<https://doi.org/10.3389/frai.2020.561802>
- Floridi, L., Cows, 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>
- 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>
- Gee, J. P. (2000). Identity as an Analytic Lens for Research in Education. *Review of Research in Education*, 25, 99. <https://doi.org/10.2307/1167322>
- Gombert, S., Di Mitri, D., Karademir, O., Kubsch, M., Kolbe, H., Tautz, S., Grimm, A., Bohm, I., Neumann, K., & Drachsler, H. (2022). Coding energy knowledge in constructed responses with explainable NLP models. *Journal of Computer Assisted Learning*, jcal.12767. <https://doi.org/10.1111/jcal.12767>
- Grimm, A., Steegh, A., Çolakoğlu, J., Kubsch, M., & Neumann, K. (2023). Positioning responsible learning analytics in the context of STEM identities of under-served students. *Frontiers in Education*, 7. <https://doi.org/10.3389/feduc.2022.1082748>
- Grimm, A., Steegh, A., Kubsch, M., & Neumann, K. (2023b). Learning Analytics in Physics Education: Equity- Focused Decision-Making Lacks Guidance! *Journal of Learning Analytics*, 10(1), 71–84. <https://doi.org/10.18608/jla.2023.7793>
- Guo, W., & Caliskan, A. (2021). Detecting Emergent Intersectional Biases: Contextualized Word Embeddings Contain a Distribution of Human-like Biases. *Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society*, 122–133. <https://doi.org/10.1145/3461702.3462536>
- Hartmann, B., & Schriever, C. (2022). *Vordenkerinnen—Physikerinnen und Philosophinnen durch die Jahrhunderte*. UNRAST Verlag.

- Hazari, Z., Chari, D., Potvin, G., & Brewe, E. (2020). The context dependence of physics identity: Examining the role of performance/competence, recognition, interest, and sense of belonging for lower and upper female physics undergraduates. *Journal of Research in Science Teaching*, 57(10), 1583–1607. <https://doi.org/10.1002/tea.21644>
- Hazari, Z., Sonnert, G., Sadler, P., & Shanahan, M. (2010). Connecting high school physics experiences, outcome expectations, physics identity, and physics career choice: A gender study. *Journal of Research in Science Teaching*, 47(8), 978–1003. <https://doi.org/10.1002/tea.20363>
- Hutchinson, B., Prabhakaran, V., Denton, E., Webster, K., Zhong, Y., & Denuyl, S. (2020). Social Biases in NLP Models as Barriers for Persons with Disabilities. *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 5491–5501. <https://doi.org/10.18653/v1/2020.acl-main.487>
- Karademir, O., Borgards, L., Di Mitri, D., Strauß, S., Kubsch, M., Brobeil, M., Grimm, A., Gombert, S., Rummel, N., Neumann, K., & Drachsler, H. (2024). Following the Impact Chain of the LA Cockpit: An Intervention Study Investigating a Teacher Dashboard's Effect on Student Learning. *Journal of Learning Analytics*, 1–14. <https://doi.org/10.18608/jla.2024.8399>
- Kayumova, S., & Dou, R. (2022). Equity and justice in science education: Toward a pluriverse of multiple identities and onto-epistemologies. *Science Education*, 106, 1097–1117. <https://doi.org/10.1002/sce.21750>
- 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 Technology*, 50(6), 2855–2870. <https://doi.org/10.1111/bjet.12868>
- Kordzadeh, N., & Ghasemaghaei, M. (2022). Algorithmic bias: Review, synthesis, and future research directions. *European Journal of Information Systems*, 31(3), 388–409. <https://doi.org/10.1080/0960085X.2021.1927212>
- Ladewig, A., Keller, M., & Klusmann, U. (2020). Sense of Belonging as an Important Factor in the Pursuit of Physics: Does It Also Matter for Female Participants of the German Physics Olympiad? *Frontiers in Psychology*, 11, 2685. <https://doi.org/10.3389/fpsyg.2020.548781>
- Latif, E., Zhai, X., & Liu, L. (2023). AI Gender Bias, Disparities, and Fairness: Does Training Data Matter? *arXiv*. <https://doi.org/10.48550/arXiv.2312.10833>
- Li, L., Sha, L., Li, Y., Raković, M., Rong, J., Joksimovic, S., Neil, S., Gašević, D., & Chen, G. (2023). Moral Machines or Tyranny of the Majority? A Systematic Review on Predictive Bias in Education. *LAK23: 13th International Learning Analytics and Knowledge Conference*, 499–508. <https://doi.org/10.1145/3576050.3576119>

- Li, W., Brooks, C., & Schaub, F. (2019). The Impact of Student Opt-Out on Educational Predictive Models. *Proceedings of the 9th International Conference on Learning Analytics & Knowledge*, 411–420. <https://doi.org/10.1145/3303772.3303809>
- 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>
- Mignolo, W. D. (2007). Delinking. The rhetoric of modernity, the logic of coloniality and the grammar of de-colonialityFootnote. *Taylor & Francis Online*, 21(2–3), 449–514. <https://doi.org/10.1080/09502380601162647>
- 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>
- OECD. (2016). Excellence and equity in education (Volume I; PISA 2015 Results). OECD.
- OECD. (2024). Education at a Glance 2024: OECD Indicators. OECD. <https://doi.org/10.1787/c00cad36-en>
- Ogette, T. (2019). Exit racism (5th ed.). unrast-Verlag.
- Pardo, A., Jovanovic, J., Dawson, S., Gašević, D., & Mirriahi, N. (2019). Using learning analytics to scale the provision of personalised feedback. *British Journal of Educational Technology*, 50(1), 128–138. <https://doi.org/10.1111/bjet.12592>
- 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., & Kaliisa, R. (2022). Learning Analytics on the African Continent: An Emerging Research Focus and Practice. *Journal of Learning Analytics*, 1–18. <https://doi.org/10.18608/jla.2022.7539>
- Prinsloo, P., & Slade, S. (2017). Chapter 4: Ethics and Learning Analytics: Charting the (Un)Charted. In *Handbook of Learning Analytics* (1st ed., pp. 49–57). SoLAR. <https://doi.org/10.18608/hla17.004>
- 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.
- Riazy, S., Simbeck, K., & Schreck, V. (2020). Fairness in Learning Analytics: Student At-risk Prediction in Virtual Learning Environments: *Proceedings of the 12th International Conference on Computer Supported Education*, 15–25. <https://doi.org/10.5220/0009324100150025>
- 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
- Sha, L., Rakovic, M., Das, A., Gasevic, D., & Chen, G. (2022). Leveraging Class Balancing Techniques to Alleviate Algorithmic Bias for Predictive Tasks in Education. *IEEE Transactions on Learning Technologies*, 15(4), 481–492. <https://doi.org/10.1109/TLT.2022.3196278>
- Shahbazi, N., Lin, Y., Asudeh, A., & Jagadish, H. V. (2023). Representation Bias in Data: A Survey on Identification and Resolution Techniques. *ACM Computing Surveys*, 55(13s), 1–39. <https://doi.org/10.1145/3588433>
- Shanahan, M.-C. (2009). Identity in science learning: Exploring the attention given to agency and structure in studies of identity. *Studies in Science Education*, 45(1), 43–64. <https://doi.org/10.1080/03057260802681847>
- 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>
- 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>
- Tan, Y. C., & Celis, L. E. (2019). Assessing Social and Intersectional Biases in Contextualized Word Representations. *Advances in Neural Information Processing Systems*, 32.
- Traag, V. A., & Waltman, L. (2022). Causal foundations of bias, disparity and fairness. arXiv. <https://doi.org/10.48550/arXiv.2207.13665>
- Traxler, A. L., Cid, X. C., Blue, J., & Barthelemy, R. (2016). Enriching gender in physics education research: A binary past and a complex future. *Physical Review Physics Education Research*, 12(020114), 1–15. <https://doi.org/10.1103/PhysRevPhysEducRes.12.020114>
- Uttamchandani, S., & Quick, J. (2022). An Introduction to Fairness, Absence of Bias, and Equity in Learning Analytics. In C. Lang, G. Siemens, & A. F. Wise (Eds.), *The Handbook of Learning Analytics* (2nd ed., pp. 205–212). SOLAR. <https://doi.org/10.18608/hla22.020>
- Wolf, T., Debut, L., Sanh, V., Chaumond, J., Delangue, C., Moi, A., Cistac, P., Rault, T., Louf, R., Funtowicz, M., Davison, J., Shleifer, S., Von Platen, P., Ma, C., Jernite, Y., Plu, J., Xu, C., Le Scao, T., Gugger, S., ... Rush, A. (2020). Transformers: State-of-the-Art Natural Language Processing. *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, 38–45. <https://doi.org/10.18653/v1/2020.emnlp-demos.6>
- 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>
- Zhao, J., Wang, T., Yatskar, M., Ordonez, V., & Chang, K.-W. (2017). Men Also Like Shopping: Reducing Gender Bias Amplification using Corpus-level Constraints. *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, 2979–2989. <https://doi.org/10.18653/v1/D17-1323>

Intersectional Feminist

*Historically grown inequalities
 Along the matrix of oppression
 Make our standpoints matter –
 Gender, race, class, abilities,
 Age, religion and belief, and sexual identities –
 Critically thinking individuals need to take control for a better
 Future by forming mass-based movements, united in intersectionalities.*

*Within our struggle we center
 The voices of those most affected, we shed
 Light on voice and experience when
 Demanding for our right to live in a system
 Without oppression, when living up to our duty
 To participate, contribute, reflect, and render
 A new system configuration, a well distributed power set.*

*We take up agency for our cause,
 Live our lives as practice of freedom, and
 We change and transform striving towards a pluriverse:
 A world where many worlds fit because
 The system is responsible to grand
 Fair chances and distributions along diver-
 Sity dimensions – resulting in a world full of beauty!*

6 Critical Consciousness

Title. Learning from the Global South: Informing Professional Development Needs for the Development of Teachers' Critical Consciousness

Abstract. Teachers play a crucial role in reducing persistent historically grown inequalities in Science, Technology, Engineering, and Mathematics (STEM) fields. One well-established approach from the Americas to prepare teachers for reducing inequalities is to support teachers in building critical consciousness. In our study, we explored the critical consciousness of teachers in the context of Northern Europe and the use of artificial intelligence systems. We conducted a type-building deductive qualitative analysis on 14 interviews with a total of seven physics teachers. We found a certain concern that professional developments for critical consciousness in the Northern European context can build upon but also many dimensions for potential improvements. Improvement potentials were especially identified for the critical understanding of feminist pedagogical thought, the critical attitude of education as the practice of freedom, and the critical action of reflection. These potential improvements indicate a promising potential of critical consciousness to reduce inequalities in STEM fields in Northern Europe.

Submitted. Grimm, A.; Navarro Camacho, N.; Steegh, A.; Grosenick, E.; Mena León, C. L.; Holst, V.; Hott, J.; Karademir, O.; Neumann, K. (2024). Learning from the Global South: Informing Professional Development Needs for the Development of Teachers' Critical Consciousness. *Journal for Educational Research Online*

6.1 Introduction

Science, Technology, Engineering, and Mathematics (STEM) fields are characterised by persistent historically grown inequalities across various dimensions, for example gender (Düchs & Ingold, 2018). In Europe, for example, certain social identities continue to encounter significant barriers in STEM fields, leading to the systematic exclusion of groups based on factors like religion, gender identity, and migration history (Avraamidou, 2019). Teachers play a crucial role in this process, as teachers' recognition – or lack thereof – of students can either foster inclusion or perpetuate exclusion in STEM fields (Archer et al., 2015; Carlone & Johnson, 2007; Çolakoğlu et al., 2023). Breaking this vicious cycle can be challenging for teachers as the education system itself is not a neutral entity; it tends to uphold or even reinforce existing inequalities without teachers who actively challenge existing inequalities and mechanisms of exclusion (hooks, 1994, p. 30). In this landscape, the increasing use of artificial intelligence systems in STEM education, for example in order to automatically evaluate student answers for teachers, introduces new risks of bias that may reinforce existing inequalities (Baker & Hawn, 2021; Lohaus et al., 2020; Yeung, 2019). However, artificial intelligence systems may also help to challenge existing inequalities (Costanza-Chock, 2020; D'Ignazio & Klein, 2020). Much like STEM education itself, artificial intelligence systems are not neutral; they either reinforce or disrupt existing power dynamics (Prinsloo & Slade, 2018, p. 4). Despite this, artificial intelligence systems are often viewed as a more 'neutral' or 'objective' means of assessing competence, which can influence how teachers recognize students' potential. Given the critical importance of teacher recognition in addressing entrenched inequalities, understanding teachers' perception of and enhancing teachers' preparation for the use of artificial intelligence systems is more vital than ever.

To prepare teachers to address historically grown inequalities, several approaches have been established in the Americas (Freire, 1970; hooks, 1994; McCausland & McDonald, 2024). Critical consciousness, for example, is supposed to help teachers reflect whether female and non-binary students in comparison to their male peers (do not) get to speak and suffer from more experiences of shame in classroom situations (hooks, 1994, 2003). The critical consciousness of a teacher is the combination of actions, attitudes, and understandings that teachers can use to challenge historically grown inequalities in the classroom (Diemer et al., 2015; Jemal, 2017; Watts et al., 2011). For the context of Northern Europe, comparable approaches exist such as critical whiteness studies by Tißberger (2017), migration pedagogics informed by perspectives such as those of Mecheril and colleagues (2020), or queer physics by Götschel (2015). Critical Consciousness can complement the existing approaches in Northern Europe well as 1) it is rooted precisely in the objective of addressing persistent inequalities and on the other hand already is an established field of research and practice in the Americas that can be built upon and learned from (Freire, 1970; hooks, 1994; Jemal, 2017). For that learning process, it is crucial to understand how applicable existing findings are in the different cultural context, especially with the new opportunities and challenges introduced by artificial intelligence systems. In order to understand to what extent critical consciousness can be used in another cultural context, a qualitative in-depth exploration is needed. In the light of the pressing challenges, the qualitative insights can inform the design of future professional developments to effectively address the historically grown inequalities.

In our study, we seek to explore the critical consciousness of physics teachers in Germany. We interviewed teachers who were involved in a project with a digital learning environment and artificial intelligence systems. In the digital learning environment, students' free text answers were automatically evaluated and the results visualised in a dashboard for

teachers, with teachers having the opportunity to provide direct feedback to students. The interviews were analysed to explore teachers' critical consciousness, using a coding framework developed collaboratively by researchers from Costa Rica and Germany. Based on this analysis, we identified different types of resources and obstacles for the development of critical consciousness among teachers. This typology can serve as a guide for future professional development. Ultimately, our goal is to inform the creation of a STEM education that effectively invites all students, "a world where many worlds fit" (Escobar, 2017; Kayumova & Dou, 2022).

6.2 Theory

6.2.1 Navigating Meritocracy and Social Justice: Inequalities in Physics Education

Modern education systems typically fulfil the task to enable participation and social mobility for all students as a member of the respective society (EU Charter, 2012; Muñoz Izquierdo, 2012). Participation is guaranteed by formulating objectives for what students need to learn in order to successfully participate in their society. The objectives of modern education systems are commonly outlined in so called educational standards (KMK, 2004; MBWK SH, 2019; MNC, 2021). These standards typically contain universal formulations of what a student should learn and thus be tested on. The result of such an examination and the respective degree of education is one important justification for social mobility based on meritocracy as well as for existing inequalities – for example higher salaries for positions that require a higher degree of education. At the same time, the inequalities should be independent of students' position on gender, race, socio-economic status, religion and beliefs, ability and chronic illnesses, and sexual identity. In other words, modern education systems contribute to processes of selection but have the task to do so in a just way that allows for social mobility and strengthens social justice. However, modern education systems systematically exclude some groups of students from careers that prepare for high salary and power positions. In Germany, women make up only 23 % and 22 % of graduates in bachelor and master degree programmes in physics (Düchs & Ingold, 2018, p. 36). For non-binary students, information is not even available. Interestingly, the inequalities cannot be explained by differences in competence development and respectively worse results in examinations. The inequalities rather find their origin in who does (not) identify with, in our case, natural sciences.

In order to understand why students opt for a career in natural sciences, the concept of identity has been introduced into the field of natural science education (Archer et al., 2022; Carlone & Johnson, 2007; Dou et al., 2019). Identities are complex constructs that help us to define who we are, rooted in our experiences and reflections (Brickhouse, 2001). STEM identity provides an answer to the question: Am I a STEM person? Students develop a STEM identity through various processes such as the recognition from their STEM teachers and the experience of being competent in classroom situations (Carlone & Johnson, 2007; Çolakoğlu et al., 2023). At the time of potential STEM identity development, students have already developed various identities such as gender identities and class identities. New identities have to be negotiated with existing identities (Brown, 2004). A STEM identity needs to fit with a student-specific combination of – staying with the example – gender and class identities. If STEM fields are considered to be for men, that results in an obstacle for female students to develop their STEM identity. Worse, when teachers recognise male students more than female students in classroom situations, this process of recognition can explain differences in STEM identity development according to gender identity. Teachers do not necessarily need to believe that men tend to be more likely to be natural science persons in order to recognise more male than female or non-

binary students. Teachers may also rely on biased information about their students' competence provided by artificial intelligence systems.

Artificial intelligence systems are introduced more and more into learning environments with many promising potentials, among them the automation of tasks (Zhai et al., 2019, p. 1145). For example, in the context of our study, teachers are provided with real-time evaluation of student artefacts visualised on their desktop in order to free up time for interactions with students and provide tailored feedbacks in the digital learning environment to the students. When an artificial intelligence system informs a teacher that a student has performed well, this can lead to the teacher recognising the student, thereby fostering the student's STEM identity development. However, it has been shown that artificial intelligence systems carry the risk of bias, for example to end up in racial profiling because postal code is used for prediction of criminality and postal code is highly correlated with race (Cheuk, 2021; D'Ignazio & Klein, 2020; Erden, 2020). Since artificial intelligence systems are trained with existing data, the artificial intelligence systems tend to reproduce historically grown inequalities, in our case to work better for male than for female students. The existing inequalities in STEM fields, the need to foster STEM identity development especially for students who face the inequalities, and on top of that the introduction of artificial intelligence systems make it increasingly necessary to prepare teachers. If teachers are prepared, they can provide recognition and competence feedback to students who need that feedback the most in order to counter historically grown inequalities.

6.2.2 Critical Consciousness as a Means of Preparing Teachers

Critical Consciousness was developed by the scholar Paulo Freire in the Global South and since then its use has spread around the world (Diemer et al., 2015; Freire, 1970; hooks, 1994; Jemal, 2017). Paulo Freire understood critical consciousness as the "reflection and action upon the world in order to transform it" (1970, p. 51). In the context of education, Freire rooted education systems in justice theories and a wider humanist perspective with the explicit task to reduce inequalities (1970, p. 45). Building on Freire's understanding, the Black feminist scholar bell hooks elaborates on how critical consciousness can be taught and conceptualises critical consciousness with three categories (2003, p. 181). Next to critical action (body), bell hooks divides reflection in what we call critical understanding (mind) and critical attitude (heart) (2003, p. 181). Critical understanding is the way teachers see the world, for example to describe the impact of colour-evasiveness on existing inequalities. Critical attitude refers to, based on a critical understanding, positioning oneself with regard to, for example, the extent to which a teacher takes up responsibility to address existing inequalities. Critical action is the concrete behaviour of a teacher to reduce inequalities. In a literature review, Jemal can distinguish conceptualisations of critical consciousness as being made out of one, two, or three of these categories (2017). As we aim at informing the professional development of critical consciousness and see in bell hooks' work (hooks, 1994, 2003, 2009) a strong foundation on that, we decided to conceptualise critical consciousness with all three categories of critical understanding, attitude, and action.

Critical Consciousness with its three categories of critical understanding, attitude, and action can help to reduce persistent inequalities. Students' STEM identity development depends on many factors, for example on having role models (Archer et al., 2022) and being recognised by teachers (Rahm & Moore, 2016). In natural sciences exist historically grown inequalities and hence there are students under-served with role models, for example physics units are almost exclusively named after white men such as Volta,

Newton, and Ampère. In order to break the vicious cycle of self-reproduction of inequalities, critical consciousness can provide the needed intervention. On the one hand, teachers can, for example, actively make physics classes more inviting especially for under-served students by inviting daily life experiences (hooks, 1994, p. 150) and emotions (hooks, 1994, pp. 15, 154) to the classroom. On the other hand, critically conscious teachers can enable under-served students to not only acknowledge existing barriers but also build resilience in order to face these barriers (O'Connor, 1997). The knowledge about a collective effort towards reducing inequalities can increase under-served students' sense of agency instead of perceiving themselves as victims of structural barriers only (O'Connor, 1997). Both, reducing barriers through critical action as well as preparing students to navigate barriers themselves, are examples for relevant contributions teachers in natural science classes can make if they are prepared to do so.

6.2.3 Professional Development Needs for Critical Consciousness in Northern Europe

So far, most of the research on critical consciousness has been conducted in the Americas. In order to advance the research in Europe, a global perspective (“#LaInternacionalFeminista”) as proposed by Global South feminist scholar Verónica Gago is well suited (2019). A global perspective allows to navigate and to find a position between localism and universalism. In contrast to localism, a global perspective enables 1) the possibility of mutual learning for researchers from different regions instead of ignoring established research fields in other regions with the argument of their irrelevance due to different contexts, and 2) the creation of connections and a common language of established research fields in different regions of the world (Gago, 2019, pp. 191–218). In contrast to universalism, a global perspective 1) does not aim for one shared solution or consensus, 2) acknowledges, respects, and allows for differences based on local contexts, and thus 3) creates a need for local contextualisation without preventing learning from experiences from elsewhere. In our research on critical consciousness, these characteristics are very important: We work with a concept developed mainly in the Americas and which might not be fully valid and reliable in our context, but can be a powerful knowledge base on which to build. In order to bring critical consciousness into praxis in Northern Europe, it is highly relevant to understand well where natural science teachers stand in terms of their critical consciousness today. As we want to inform professional development needs of teachers, we aim at understanding resources and obstacles that can be addressed in professional development (Hott, 2024). In other words, we need to explore resources and obstacles in teachers' personal resources that can be the foundation for professional development in the specific context of Europe.

Synthesising, we contextualise the goal of interventions towards social justice in natural sciences: The most relevant contexts are STEM identity development and career aspirations, which correlate strongly with access to power. This access to power is the reason why we do not focus on the achievement of basic competence levels in natural sciences. Instead, we anchor our conceptualisation of critical consciousness in STEM identity development. The normative framework is to analyse inequalities along positions on diversity dimensions and to assign responsibility for transformation to the system. The objective is a representation of the actual proportions of the student population in the proportions of students choosing natural science careers. Since we root the need for critical consciousness development in the need for STEM identity development for all students, the action category of critical consciousness is of particular relevance to us. We aim at preparing teachers to strengthen STEM identity development in a way that is equally inviting to students of all positionalities on diversity dimensions. In order to make our

findings transferable to praxis, it is necessary to make findings actionable for professional developments. We ask:

- Which professional development needs for critical consciousness can be identified in the different cultural setting of Northern Europe?
- What kind of different types of teacher characteristics can be identified based on professional actions in the classroom?
- How can resources and obstacles of different types of teacher characteristics inform the design of professional development pathways?

6.3 Methodology and Data

6.3.1 Setting the Context: Semi-Structured Interviews within the Project

We conducted our research as part of the project “Learning Progression Analytics - Analyzing and Fostering Learning for the Development of Competence”. In the project, we worked with a total of seven teachers and their classes over a period of five weeks with one 90-minute session per week on a learning unit on the topic of energy. The learning unit was accompanied by a digital workbook implemented in a digital learning environment. The digital learning environment included pre-trained artificial intelligence systems that analysed the students’ responses and provided information on students’ competence or progression in developing competence respectively for teachers in real time through a dashboard. The dashboard also had a function that allowed teachers to provide students with feedback. Teachers were supported by the researchers in two ways: 1) the teachers received a training on the design principles behind the learning unit and how to use the dashboard prior to facilitating the unit themselves, and 2) the researchers provided technical support in the classroom during the first session and video conferences were scheduled after the second, third, and fourth session.

Our main aim for this study was to gather information for designing professional development opportunities to foster teachers’ critical consciousness and to describe culture-specific challenges for the development of critical consciousness in the context of Northern Europe. We conducted semi-structured interviews that allowed the interviews to have both, a theoretically informed structural core as well as spontaneous adjustments diving deeper when necessary. In order to make sure that our findings are robust against day-specific moods and the concrete teaching experience in the week, we decided to conduct two interviews at different points of time with a similar structure, ending up with 14 interviews in total. We decided to conduct the interviews after the first and the third video conference in which we provided support in using the dashboard in order to 1) have the relevant context present, and 2) have as much time as possible between the interviews. The interview guides can be found in the supplemental materials.

6.3.2 Type-Building Deductive Qualitative Analysis

We built on a solid theoretical foundation built in the Americas and therefore decided to conduct a deductive instead of an inductive qualitative analysis. In order to reveal specific needs for future professional developments, we conducted a type-building analysis on top of the deductive qualitative analysis. We used teachers’ resources and obstacles that a professional development for critical consciousness can build upon to build types. The clear advantage from our point of view was that teachers’ resources and obstacles provided us with actionable results that can be used directly for designing professional development programmes and transferring the empirical findings into praxis. Given our

decision to use a type-building analysis, we chose the methodology developed by Kuckartz due to its explicit and detailed description of the type-building process (2018).

6.3.2.1 *Deductive Qualitative Analyses for Critical Consciousness*

We chose one the interview as the unit of analysis instead of both interviews of one teacher as we conducted the two interviews in order to be robust against day-specific moods and the concrete teaching experience in the week. We believe the analysis of each interview to be valid as we expect the development of critical consciousness to be minimal based on the small amount of time combined with the intervention being only questions for relevance, responsibility, and action. Coding single interviews, we obtained two units per teacher. We defined a coding unit as one sequence of a speaker until the interviewer said something. Defining the unit of coding in this way allowed 1) for great precision in terms of inter-coding analysis as there was no ambiguity to where a coding unit ends or begins, and 2) for clearer results in terms of which codes frequently occurred together, which was relevant for our resources and obstacles analysis. For example, if a particular critical understanding frequently co-occurred together with a critical action, but the critical action never occurred without that critical understanding, this indicates that the critical understanding might be an important personal resource for the development of the critical action. Finally, we defined the unit of context as all other necessary parts of the interview that were relevant in order to understand the unit of coding.

The coding manual was developed in four steps. 1) A literature-based coding manual was created, drawing heavily on the works of bell hooks and enriched with relevant concepts from other scholars for our specific context of natural science education (hooks, 1994, 2003, 2009). 2) With that initial coding manual, four interviews were coded by two researchers from Germany in order to sharpen the explanations of the codes, to validate the functioning of the codes, and to define first boundary cases that made the limits of the codes clearer. 3) Equipped with this initial coding manual and first codes, the coding manual was discussed from a global perspective with researchers from Costa Rica and Germany. In the course of this discussion from a global perspective, new codes were added, explanations were adjusted, and some codes were merged based on the enriching theoretical perspectives of the researchers from Costa Rica. 4) The updated coding manual was then used by the two German researchers in order to code the interviews and was subsequently only slightly updated in the testing phase when discussions revealed imprecise or incorrectly updated explanations. The main categories and their sub categories from the final version of the coding manual are shown in Table 6-1. Next to the codes themselves, we coded weights identifying whether the unit of coding indicates critical consciousness or violation of its principles. We coded positive, negative, or neutral weights. In addition to these categories, we also coded some formal information and tried to identify factors that enable the development of critical consciousness. The full coding manual including the explanations can be found in the supplemental material.

Table 6-1 - Coding Manual on Critical Consciousness

Mind: Critical Understanding	Heart: Critical Attitude	Body: Critical Action
feminist pedagogical thought	commitment	building groups of action
identity development	education as the practice of freedom	dialogic conversation

Mind: Critical Understanding	Heart: Critical Attitude	Body: Critical Action
intersectionality	pluriverse	embodiment and engaged pedagogy
matrix of domination	positionality on culture	initial action
mechanisms of discrimination	positionality on identity	natural science culture
transformation narrative	willingness to change	passion of experience
		reflection
		resilience

In addition to the main and sub categories, we coded positive, negative, or neutral weights. With these weights we make visible whether a statement is in line with the sub category or violating a sub category's principle. For example, if a teacher claims that students' gender identities do not play any role at all in their physics classes, we label matrix of domination with a negative weight. The teacher's answer contains information about the perception of historically grown inequalities and in that specific case, the teacher neglects any role of the historically grown inequalities. On the other hand, when the teacher claims that gender identities play a role for the students' learning we also label matrix of domination but assign a positive weight. For example, an interview with ten codes for matrix of domination can have a positive weight balance (more positive than negative weights), a negative weight balance (more negative than positive weights), or a neutral weight balance. For balancing, the weights can be aggregated per main category or separately for each sub category. The weights allow us to analyse not only where teachers demonstrate critical consciousness but also where teachers have, for example, understandings that violate critical understandings. These violations can be valuable information for the design of a professional development as the existing understandings need to be addressed and actively deconstructed in order to develop critical consciousness.

Inter-rater reliability was determined in multiple steps. The two researchers from Germany each coded two interviews from the same teacher. The inter-rater reliability was calculated for each code as share of words coded by both researchers out of words coded by at least one researcher. The researchers met in order to discuss differences, reach agreements, and jointly define boundary cases. The main purpose of testing with the initial coding manual was to make explanations less ambiguous and to gain a common understanding of the main categories. This common ground was very helpful when the testing phase with the updated coding manual began. The updated coding manual was discussed together before starting the testing phase again with interviews that were not coded with the initial coding manual. The first two interviews with the updated coding manual revealed new differences in interpretations of the researchers, which were solved by discussion until agreement. The following two interviews still revealed differences in coding. However, two things were achieved in this step: 1) a great agreement on the coding of main categories, and 2) the researchers were able to recognise in almost all cases why the other researcher had coded a code, even if they themselves had not assigned the corresponding code. These achievements can be seen as a major step towards reliable inter-rating, especially in the light of the coding manual's very fine-grained codes. For example, a critical attitude

can be coded as commitment if the teacher shows concern for gender equality. However, the same unit of coding can be coded as pluriverse if the teacher assigns the responsibility for gender equality to the school in general, which goes beyond the teacher's own commitment. When one researcher assigns commitment and the other researcher assigns pluriverse, a rigorous analysis of inter-rater reliability yields no agreement. Nevertheless, agreement on the main category has already been reached and two codes close to each other were assigned. We therefore interpreted the inter-rater reliability as clearly increasing already after the second round of analysis. We continued in that same way, discussing disagreements and adding boundary cases. We analysed the final inter-rater reliabilities on the level of main categories as well as sub categories in terms of numbers in addition to the procedural description that we offer here, both of which was considered sufficient and can be found in the supplemental material.

6.3.2.2 Type Building based on the Main Category of Critical Action

The goal of the type-building was to inform future professional developments for critical consciousness of natural science teachers by identifying resources and obstacles in qualitative depth. For the type-building we used the same unit as for the deductive analysis for critical consciousness, one interview.

We defined the summarising characteristics of each interview oriented on the codes from our deductive analysis. For critical understandings, attitudes, and actions, we characterised the tendency on the main category as a whole as well as the contributions of the single sub categories with their weight contributions. We did so as we were not sure whether to expect types rather on, for example, the level of more or less action or on the level of different action. Analysing for both allowed our type-building process to be evidence-informed where we had no theory-informed hypothesis.

To build our typology, we first wrote a summary of the characteristics for each of the 14 interviews, trying to stick as closely as possible to what the teachers had actually said. These summaries were written by one of the researchers who conducted the deductive analysis. Based on the summaries, the two researchers who conducted the deductive analysis created groups and decided on an appropriate number of groups. The aim here was 1) to create as homogenous groups as possible that are as clearly distinguishable as possible from each other – high intra-group homogeneity and high inter-group heterogeneity – and 2) to build action-oriented groups. In contrast to Kuckartz, we also decided to assign all interviews to one group at this stage due to the small number of interviews (2018). Once the groups had been defined, we developed a name for each of the groups. The names should reflect the respective characteristics as well as possible. In the end, we had a typology with each type having a name and each interview being assigned to a type and having a summary of characteristics.

After the creation of the typology, we wrote a description of each type as well as the typology as a whole. To validate the descriptions, we went back through the summaries of each interview's characteristics, checking for consistency with the type description and updating the type descriptions where necessary. We sought to describe each type in qualitative depth with the respective resources and obstacles that can be the foundation for the design of professional development programmes.

6.3.3 Reflections on the Methodology

Discussing with researchers from Costa Rica and Germany, one researcher with a formation in theology and philosophy made an interesting observation. Our coding manual with its use of heart, body, and mind has some interesting parallels with analyses that the

researcher knows from theology and philosophy: It is based in visualisations that are rooted in the human being in full integrity. The visualisation of heart, body, and mind was very helpful for that researcher when actually coding text, especially because we used quite complex and fine-grained codes. We found that observation relevant for us as social science researchers and worth sharing with our research communities.

Analysing from a global perspective, we made another interesting observation: We took different cognitive routes in the deductive analysis. The researchers from Germany coded each interview three times, starting with one main category and then moving on to the next main category. In contrast, the researchers from Costa Rica decided to take some time in the beginning to memorise the complete coding manual and then went through the interviews only once. Since we used the same coding manual but applied it to different interviews (the interviews in this article were only coded by the researchers from Germany), we could not analyse for possible effects. However, we found the observation interesting and consider further investigation to be relevant. Such an investigation could draw on theories on subjectivity and the role of the researchers' subjective gazes and ways of observing the world.

6.4 Results: Professional Development Needs

6.4.1 Critical Consciousness in the Context of Northern Europe

We start off with our general observations over the entire sample in order to reflect on our first question. From a main category perspective, every single interview has positive balances on critical actions, attitudes, and understandings. In other words: There is no single interview without any concern or commitment for diversity. In all interviews, teachers demonstrate critical consciousness to a certain extent, no teacher actively works against diversity main-streaming. The finding of positive averages indicates that there would be no general need for deconstruction in professional developments of understandings, attitudes, and/or actions that work against critical consciousness.

However, the overall number of sub categories with statements that have positive weights is quite low on average. For critical actions, from eight possible sub categories on average 1.9 sub categories have at least one positive score. For critical attitudes out of six possible sub categories on average 1.6 sub categories have at least one positive score. And for critical understanding out of six possible sub categories on average 3.1 have at least one positive score. Many categories without at least one positive score indicate little use of diversity unfolding analyses and action options that would be available.

Negative weight balances on average over all interviews only result for three sub categories: 1) feminist pedagogical thought, 2) education as the practice of freedom, and 3) reflection. Negative weight balances in these sub categories indicate a need to actively deconstruct in professional development understandings, attitudes, and actions that work against critical consciousness. Also, negative weight balances should guide our analyses for obstacles and resources that might lead and/or prevent negative weights in these sub categories in order to meaningfully deconstruct them in professional developments.

Asking the teachers whether they believe the artificial intelligence systems to work equally well for students with different gender identities or differences in socio-economic status, teachers did not feel comfortable to judge. Teachers indicated either that they lack the competence to judge, the time to investigate, the necessary information on their students' socio-economic status, or that the responsibility should be with someone else and that only systems that are tested to be non-discriminatory should be given to them. This finding

emphasises the need to on the one hand sensitise teachers for possible biases in artificial intelligence systems and on the other hand the need to establish effective structures to counter the negative effects of biased artificial intelligence systems on students.

Finally, four sub categories were not even coded once: 1) intersectionality, 2) positionality on identity, 3) positionality on culture, and 4) dialogic conversation. No statement regarding these sub categories indicates that they are invisible as categories of understanding, attitude, and action for the teachers in the interviews.

6.4.2 Typology

6.4.2.1 *Deciding on Two Types: Initial Action and First Critical Action*

Critical consciousness is necessary in order to enable STEM identity development for all students. As the impact shall be on students, we believe the action component of teachers' critical consciousness to be of particular relevance. The general observation is that all interviews yield a positive weight score in critical action. However, when zooming in to the sub categories of critical action, we notice that teachers in seven interviews only score in initial action while teachers score in the other seven interviews in several sub categories of critical action that go beyond initial action. Initial action means that teachers do something in order to specifically address gender, for example – however, the action is so small that it cannot be considered a critical action. Watching on gender identities in building working groups for some minutes, for example, is such an initial action. Other sub categories of critical action go beyond that – for example embodiment and engaged pedagogy means to actively invite emotions to be present in the classroom. Building the types based on action beyond initial action makes sense because 1) action is what has impact on STEM identity development, 2) the further development of critical consciousness has other necessities based on the practices already established, and 3) interviews can be clearly grouped based on any action beyond initial action or not. The scores beyond initial action are not on all sub categories and rather scarce than many: The critical action beyond initial action reported in the interviews is of a first critical action character rather than of a bold and well-established type. Creating the typology, we have two interesting findings from an evidence-perspective: There are neither teachers in our sample who do not care for diversity at all or wilfully work against it nor teachers who have a well-established and profound critical action in the terms that we understand it. Instead, the two types that we can distinguish are initial and first steps which is why we name the types 1) Initial Action, and 2) First Critical Action as shown in Figure 6-1.

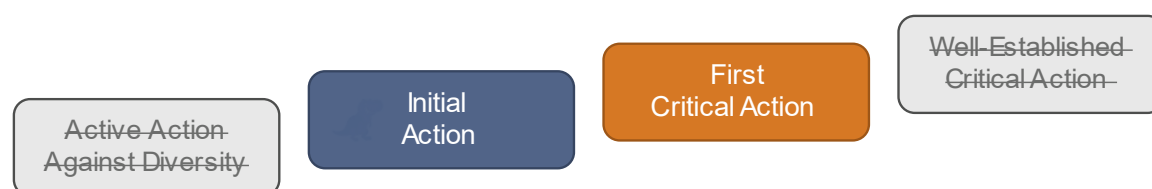


Figure 6-1 - Typology built by analyses for Critical Action: 1) Initial Action (n=7) and 2) First Critical Action (n=7)

From a professional development perspective, the non-existence of active action against diversity or questioning of diversity as an attractive or necessary goal is interesting. In our sample, there can be no need derived from any interview to address the necessity to strive towards diversity from a critical consciousness perspective. In other words: Professional developments can build on a normative demand for diversity. At the same time, all interviews indicate a necessity to address critical understandings and attitudes as relevant principles are violated – we coded negative weights. It cannot be built on common

understandings or attitudes that are in line with our definition of critical consciousness and which we would have called Well-Established Critical Action. Therefore, we end up with two types: Initial Action and First Critical Action.

6.4.2.2 Type-Comparative Observations

Comparing the two types with regard to the remaining two main categories of critical attitude and critical understanding, the differences mainly manifest in number of positive scores for a sub category. Both types score on a similar number of sub categories on critical understanding and critical attitude. Also, both types score similarly on negative scores for sub categories. This finding is well in line with our typology as Initial Action and First Critical Action instead of Well-Established Critical Action: In all interviews we find violation of sub category understandings and attitudes that would be in line with critical consciousness. However, in the interviews with stronger steps towards Well-Established Critical Action also more instances of positive sub categories are found. The detailed findings from a difference perspective for all sub categories and positives, negatives, and balances are shown in Table 6-2.

Table 6-2 - Comparing Initial Action (IA) and First Critical Action (FCA)

FCA with more than double of <u>positives</u> than IA for...	FCA with less than half of <u>negatives</u> than IA for...	FCA with more than double of <u>balances</u> than IA for...
feminist pedagogical thought	-	-
identity development	-	identity development
mechanisms of discrimination	mechanisms of discrimination	mechanisms of discrimination
education as the practice of freedom	-	-
-	pluriverse	pluriverse
willingness to change	-	willingness to change
building groups of action	-	building groups of action
embodiment and engaged pedagogy	-	embodiment and engaged pedagogy
natural science culture	-	natural science culture
passion of experience	-	passion of experience
reflection	-	-
resilience	-	resilience

The comparison of the two types offers interesting conclusions for professional development as well. More positive instances indicate a higher degree of reflection and ability of explicit naming for the interviews that are grouped in First Critical Action. At the same time, both types seem to have comparable needs with the origin in negative scores: Only codes for mechanisms of discrimination and the perspective on responsibility of the

system instead of the individual as coded by the pluriverse differ. Finally, the differences could also be used for hypotheses generation in terms of how individual teachers can move from Initial Action to First Critical Action. Obviously, no such trajectory can be derived from our data. However, using the empirical evidence on differences can be a starting point to empirically investigate trajectories.

6.4.3 Resources and Obstacles for Professional Developments

6.4.3.1 Initial Action

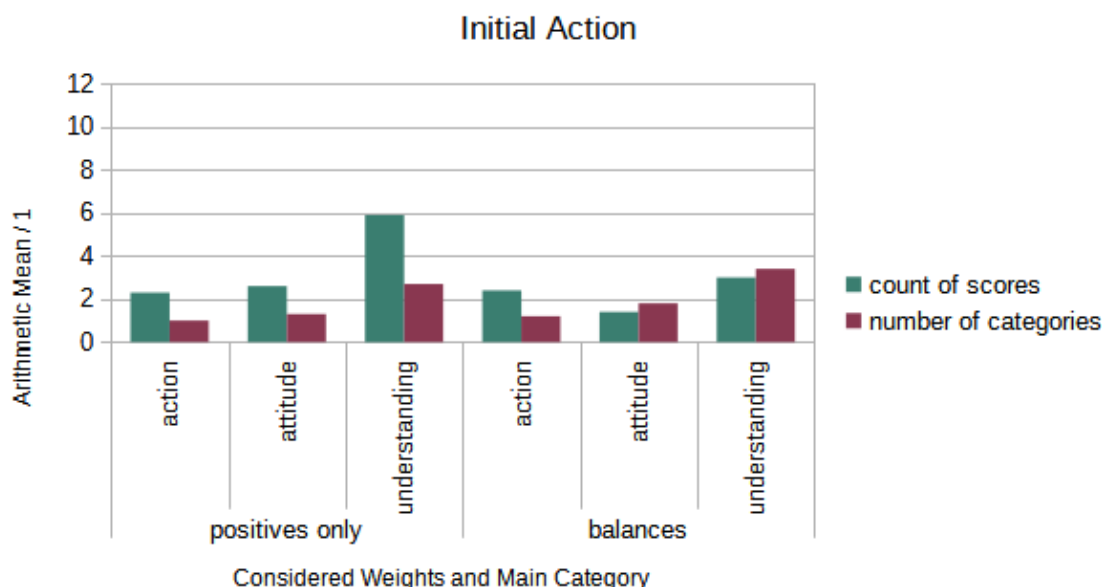


Figure 6-2 - Results for Initial Action on Critical Consciousness with its Main Categories

The results for the seven interviews of the type Initial Action are shown in Figure 6-2. There is no interview with no action related to a diversity dimension registered. However, the arithmetic mean of positive scores is 2.3 critical actions per interview only with all of them being initial actions instead of more profound critical actions. The attitudes and understandings both have a higher positive than negative count resulting in the balance score being smaller than the positive count but still positive. Finally, no interview of the type Initial Action contains positive weights in more than three sub categories of attitudes (1.3 on average) or in more than four sub categories of understandings (2.7 on average).

"Ich sehe generell in der Schule sehr deutlich, dass Jungs und Mädchen anders funktionieren."

"In school in general, I witness that boys and girls function in distinct ways."

"Insofern, dass man beim Thema Geschwindigkeiten auf der einen Seite einen Fußball hat, der geschossen wird und auf der anderen Seite ein Pferd, das reiten kann."

"With the topic of velocity, on the one hand you can have a football being shot and on the other hand a horse for riding."

"Ich gebe mir größte Mühe, alle gleich zu behandeln und keine Unterschiede zwischen Geschlechtern zu machen."

"I really try to treat every student in the same way and to make no differences between the genders."

From a resource perspective, we find strong positive weights with an arithmetic mean of 1.0 and bigger in the critical understanding of 1) matrix of oppression (“I witness that boys and girls function in distinct ways”), and 2) identity development (“the topic of velocity”), the critical attitude of 3) commitment and in 4) initial action (“I really try to”). In other words: There are understandings, attitudes, and actions where professional developments can build upon. For the identity development positive weights are mainly found due to mentions of the relevance of context as shown in the quote about the topic of velocity and two contexts – which is an important part of identity development, but only one part of it and not the full construct.

"Ich finde, so wie es im Moment formuliert ist, ist es geschlechtsneutral und das finde ich in Ordnung."

"I believe that in the ways its formulated currently it is gender neutral and I believe that is fine."

"Aber wenn die Familien das anders sehen [als ich] und sich da mehr oder weniger beschweren, werde ich da nicht hinterherrennen und sagen, du solltest dich aber bitte auch beschweren [wie bei deinem Sohn], wenn deine Tochter schlechte Noten [in Physik] schreibt. Das ist nicht meine Aufgabe."

"But when the families have a different position [than me] and start complaining more or less, then I am not going to bring the topic up again and again, you should complain also [as you did with your son], when your daughter brings home bad grades [in physics]. That is not my task."

From an obstacle perspective, we find negative weights for the understandings 1) feminist pedagogical thought, 2) matrix of domination, and 3) mechanisms of discrimination, for the attitudes 4) education as the practice of freedom and 5) pluriverse, and for the critical action 6) reflection. Negative weights are indicators for a violation of the principles defined for the sub categories in critical consciousness. These negative weights indicate a necessity of deconstruction of existing conceptualisations for a professional development. In the interviews of the type of Initial Action, teachers acknowledge the relevance of diversity dimension in physics classes to some extent, express a will to prevent discrimination, and report small actions as a teacher in order to address the specific needs of students from differing positionalities on diversity dimensions. These are strong resources found in interviews of the type Initial Action that professional development can build upon. At the same time, teachers claim that the best role diversity dimensions have in their classes is no role at all instead of acknowledging the relevance and actively countering discrimination (“it is gender neutral and I believe that is fine”), do rather formulate solutions that puts the responsibility for work towards diversity on the individual student instead of on the structure and system (“That is not my task”), and are guided in their actions rather by a goal of neutrality and non-discrimination than diversity-inviting and active counter-work.

6.4.3.2 First Critical Action

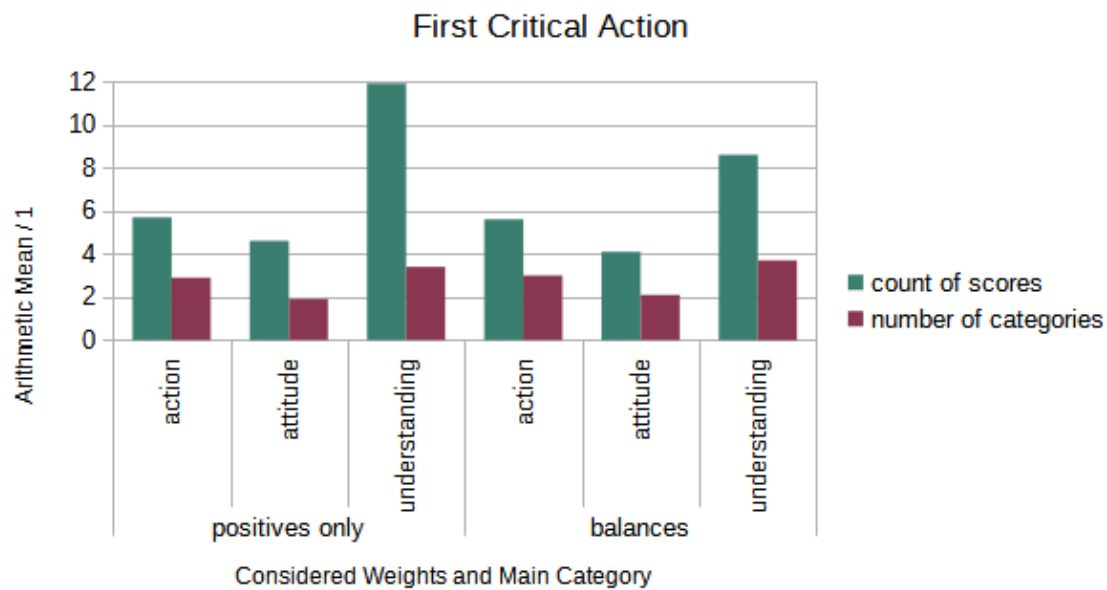


Figure 6-3 - Results for First Critical Action on Critical Consciousness with its Main Categories

The results for the seven interviews of the type First Critical Action are shown in Figure 6-3. There is no interview with no action related to a diversity dimension registered. The arithmetic mean of positive scores is 5.7 actions per interview. The attitudes and understandings both have a higher positive than negative count resulting in the balance score being smaller than the positive count but still positive. No interview in First Critical Action contains positive scores in more than three sub categories of attitudes (1.9 on average) more than four sub categories of understandings (3.4 on average).

From a resource perspective, we find strong positive weights with an arithmetic mean of 1.0 and bigger in the critical understanding of 1) identity development, 2) matrix of oppression, and 3) mechanisms of discrimination, the critical attitude 4) commitment, and the critical actions 5) embodiment and engaged pedagogy, and 6) initial action.

"Also ich habe Schülerinnen die mich fragen, gibt es eigentlich auch weibliche Physikerinnen. Also ganz offensichtlich gibt es da ein Identitäts... problem vielleicht auch."

"Well, I have female students who ask me whether there are female physicists as well. Obviously, there is some kind of... identity problem."

"[Dass] man eben für Unterstützersystem sorgt, die vielleicht zuhause nicht geleistet werden [können]"

"[that] you establish a support system with types of support that cannot be provided at home"

"Wir versuchen irgendwie, also nicht nur ich, eigentlich ist das ein Bestreben über die ganze Schule hinweg, wir wollen diesen Einfluss von sozioökonomischem Status möglichst gering halten."

"We try somehow, not only me, it is a struggle of the entire school, we want to keep the influence of socio-economic status as low as possible."

"Wir haben ja festgestellt, dass wir immer irgendwie relativ wenig Schülerinnen in"

"We noticed that we have relatively little female students in the [physics] high level"

den [Physik-]Profilkursen haben. Und das versuchen wir auch irgendwie... Also das ist nicht mein Bestreben alleine, das versuchen wir auch in der Fachschaftsarbeit irgendwie gezielt anzugehen."

courses in school. And then we try to somehow... well, it is not a struggle of myself only, we try to work together with all of our physics teachers to address the problem specifically."

Differing from Initial Action, identity development is more differentiated for example by naming identity development as such ("some kind of... identity problem") or by analysing supportive relationships next to context ("provided at home", "is a struggle of the entire schools", "with all of our physics teachers"). Also, mechanisms of discrimination are not only without negative weights but instead have positive weights. Positive weights in the critical action of embodiment and engaged pedagogy is characteristic for the interviews of the type First Critical Action.

"Dass ich dann sage, ey, das ist einfach mega respektlos, was du hier machst. Es geht nicht nur darum, dass du hier eine Person irgendwie störst, sondern in dem Sinne auch niedermachst und diese Person sich dann vielleicht nicht mehr meldet."

"Then I say that is very little respect you are showing here. It is not only about annoying another person, it is about you attacking that person in their identity and that the person might not return to raise their voice in classes."

"Also ich versuche, alle Antworten auf jeden Fall zu wertschätzen. Und ich versuche, eine sehr positive Fehlerkultur einzubringen."

"Well, I try to appreciate all answers. And I try to establish a positive culture of failing."

"Da habe ich ehrlich gesagt auch nicht die Kapazitäten für, irgendwie so in jeder Klasse irgendwie komplett anderen Physikunterricht zu machen. Dafür ist das irgendwie dann einfach auch zu sehr irgendwie so schon irgendwie vorgefertigt."

"Honestly speaking, I do not have the capacities for that, for somehow make completely different classes for all classes. For that, somehow, my classes are too much following a standard concept."

Finally, quite some sub categories have positive arithmetic means in weights bigger than 0.0 but smaller than 1.0, indicating first connection points in various sub categories that professional developments can build upon. For example, in some interviews teachers actively reflect the role of shame and self-esteem and report own interventions striving to enable self-esteem development and protection for all students ("it is about attacking that person in their identity"). In other interviews teachers demonstrate a reflective action to establish a culture that invites failing or reflect their own resources and limits explicitly ("I try to establish a positive culture of failing") or to acknowledge the limits of their resources without neglecting the entire responsibility at the same time ("I do not have the capacities").

From an obstacle perspective, we find negative weights for the critical understandings of 1) feminist pedagogical thought, 2) matrix of domination, and 3) transformation narrative, the critical attitudes of 4) education as the practice of freedom, and the critical action of 5) reflection. Differing from the type of Initial Action, in the First Critical Action interviews we do not find negative weights on mechanisms of discrimination and the pluriverse. Missing negative weights on pluriverse indicate a stronger assignment of responsibility with the structure instead of the individual for transformation towards more diversity in STEM fields.

"Naja, also generell, also ich weiß ja erst mal so direkt nicht so sehr über den sozioökonomischen Status."

"Well, in general, I do not actually know anything about the socio-economic status of my students."

"Und mir ist völlig egal, was für ein Geschlecht jemand hat."

"And I do not care at all which gender a student has."

At the same time, we still find violations of critical consciousness sub categories as shown here. For example, teachers indicate that in order to address inequalities due to differences in socio-economic status they would need to know the socio-economic of each student or express that gender does not matter for them at all instead of actively engaging against the reproduction of historically grown inequalities. The negative weights on education as the practice of freedom reflect an attitude that does not understand education as a means to establish diversity in STEM rooted in a justice conceptualisation.

6.5 Discussion

6.5.1 Critical Consciousness as Structure Against Inequalities in Identity Development

We root our work in the analysis of reality that historically grown inequalities along diversity dimensions exist within STEM education – and that this finding is problematic from a social justice perspective that is expressed in, for example, the EU Charter (EU Charter, 2012). The historically grown inequalities can be best understood from a STEM identity development perspective: It is not only about competence development, but for example recognition is very relevant for students' to develop a STEM identity. Recognition makes both, artificial intelligence systems and teachers, highly relevant as both of them can provide such a recognition. Critical consciousness has shown to open up perspectives that 1) connect well to some of the already developed understandings, attitudes, and actions as resources of teachers from Northern Europe, while critical consciousness 2) highlights development potentials with needs for deconstruction as obstacles. As the use of artificial intelligence systems in STEM education cannot build on Well-Established Critical Action of teachers, our findings additionally indicate 3) the need for a multi-faceted approach that on the one hand prepares teachers and on the other hand aims at preventing bias in artificial intelligence systems as well as possible.

6.5.2 Implications for Research Communities

6.5.2.1 Critical Consciousness in Northern Europe

We explored interviews with teachers and identified two types: Initial Action and First Critical Action. The analysis is rooted in transfer theory with a resources and obstacle lens. Building on these findings, professional developments addressing the specific needs of each type can be developed. These professional developments need to 1) deconstruct the conceptualisations that we labelled as negative weights, for example on feminist pedagogical thought, and 2) strengthen and widen the conceptualisations of critical understandings, attitudes, and actions. We would like to highlight three remaining unknowns: 1) In Figure 6-1, we offer the perspective that development of critical consciousness might pass from Initial Action through First Critical Action up to Well-Established Critical Action. Such an interpretation would allow professional developments for answer types of Initial Action to be oriented by the differences to First Critical Action.

Whether such typical development paths can be verified empirically or whether there are other pathways remains unknown. 2) We identified obstacles and resources that can and should be built upon and addressed in professional developments. Which success factors of professional developments guarantee a long-lasting development of critical consciousness in the context of STEM education in northern Europe remains untouched by our findings. 3) Our research is of exploratory nature. More evidence needs to be gathered to show how representative our sample is in terms of critical consciousness for STEM teachers in northern Europe.

6.5.2.2 *Artificial Intelligence Systems in STEM Education in Northern Europe*

Our findings are of exploratory nature, no general conclusions can be drawn. However, we offer an interpretation of what our results would imply for the use of artificial intelligence systems if they are verified by further evidence. We focus on one particular finding that we would like to highlight: The combination of no interview indicating active work against diversity and negative weight scores for reflection offers interesting opportunities and challenges for the use of artificial intelligence systems in STEM education. On the one hand, artificial intelligence systems can precisely offer reflection opportunities that might not have entered classrooms without them. These could be used in teacher-facing dashboards with critical questions for reflection, to give an example. Such interventions could build on critical attitudes of making no harm in terms of working against diversity. On the other hand, strong neutrality beliefs of the teachers à la no impact of diversity dimensions are a challenge for the use of artificial intelligence systems: Following such beliefs, teachers would expect the task of de-biasing artificial intelligence systems to be a task for the computer science community – instead of critically interacting with the outputs offered by the artificial intelligence systems. Even worse, the outputs might be interpreted as more neutral than own judgements. These interpretations lead to a strong demand to de-bias the artificial intelligence systems and/or trigger critical interactions and make the fallibility of artificial intelligence systems a topic.

6.5.3 Limitations

We worked with seven teachers from one region in one country only. Hence, we might have selected a particular sample with our blind spots remaining unknown to us. Additionally, we have all our information from two interview parts with 15 minutes each. Even though we did not find many differences between the two interviews of the same teachers, longer interviews with more space for answers and questions directed specifically on the sub-categories that we were looking for might have revealed more resources and obstacles that professional developments can build upon. Our decision to refer to Northern Europe is based on not extrapolating too much while on the other hand not isolating local contexts too much and thereby aiming at the prevention of knowledge silos. However, we conducted our interviews in one region of one country of Northern Europe only which means that the extrapolation to the entire region will have its failures – for very specific sub-cultures it is even certainly wrong to do so. Finally, we worked with very broad main and even sub-categories in order to have the most holistic perspective possible on critical consciousness. Broad categories made it harder to reach good inter-rater reliabilities and also will make it harder to analyse specific effects from, for example, one sub-category on the STEM identity development of students. Whether the choice of broad categories is helpful in order to address the reproduction of historically grown inequalities is an empirical question that remains unknown from our work.

6.6 Synthesis

In our study we explored the critical consciousness of physics teachers in Northern Europe. We identified resources and obstacles that effect teachers' ability to develop critical consciousness. Within all 14 interviews, we found evidence for a certain concern to not act discriminatory, which we understand as a resource. Additionally, all teachers reported at least initial actions, namely actions aiming to address different needs due to different positionalities on at least one diversity dimension. However, in none of the interviews we found Well-Established Critical Action, but instead found the two types of Initial Action and First Critical Action. The First Critical Action interviews especially differ from the Initial Action interviews by 1) at least some critical actions being reported (even though little in number and not for all sub categories), 2) by assigning responsibility to work against historically grown inequalities less to the individual student and more to the school system, and 3) especially by more instances of positive weighted critical understandings. In all interviews we found obstacles being instances where teachers violated critical consciousness' understandings, attitudes, and/or actions, mostly for the sub categories of feminist pedagogical thought, education as the practice of freedom, and reflection. Our results indicate a need for multi-faceted approaches: On the one hand our analyses indicate a high potential for critical consciousness as an intervention against the reproduction of historically grown inequalities within the context of Northern Europe. There are resources professional developments can draw upon while there are obstacles and more resources to be built. At the same time, we did not find any Well-Established Critical Action which implies a certain need for de-biasing artificial intelligence systems and the preparation of teachers for a critical use of such systems. Critical consciousness seems to be one promising instrument on the way towards a world where many worlds fit in STEM education.

Acknowledgements

Our work lays on the shoulders of various great thinkers who do not yet receive the visibility they deserve in the context of STEM education from our perspective. We want to highlight especially the work on critical consciousness from Paulo Freire (1970) and bell hooks (1994, 2003, 2009). Additionally, we want to express our gratitude to the open software community, highlighting the great project of QualCoder which we used for coding. As a part of scientific transparency, we report that this work was supported by the Federal Ministry of Education and Research (BMBF), grant number 01JD2008.

Supplemental Material

- Interview Guides
- Coding Manual Critical Consciousness
- Research Diary
- Interview Transcripts
- Typology Analyses – Calculations
- Links to software used for qualitative data analyses and coding

Author Contributions

Grimm, A. (A.G.), Navarro Camacho, M. (M.N.C.), Steegh, A. (A.S.), Grosenick, E. (E.G.), Mena León, C. L. (C.M.L.), Holst, V. (V.H.), Hott, J. (J.H.), Karademir, O. (O.K.), Neumann, K. (K.N.)

- Conceptualization: A.G., A.S.

- Project Operational Coordination Work: A.G.
- Data Collection: A.G., J.H., O.K.
- Creation of Coding Manual: A.G., M.N.C., E.G., C.M.L.
- Coding: A.G., E.G.
- Methodology & Results Preparation: A.G.
- Interpretation and Discussion of Results: A.G., M.N.C.
- Original Draft Preparation: A.G., M.N.C.
- Writing-Review and Editing: A.G., M.N.C., A.S., K.N., E.G.
- Documentation: A.G., E.G.
- Mentoring: M.N.C., A.S., M.K.
- Project Management: A.G., M.K., K.N.
- Funding Acquisition: K.N., M.K.

All authors have read and agreed to the submitted version of the manuscript.

Supplemental Material: Coding Manual Critical Consciousness

Diversity dimensions: gender, race, ability, religion & beliefs, sexual identity, age, social class

Critical Consciousness is the combination of Critical Understanding, Critical Action, and Critical Attitude:

- “Critical action is the behavioral element of CC and refers to actions designed to counter or respond to injustice in a liberatory manner (Watts et al., 2011).” (Diemer et al., 2015, p. 810)
- Critical attitude “refers to an individual’s agency and commitment to address perceived injustice(s).” (Diemer et al., 2015, p. 810)
- Critical understanding “refers to the process of people ‘coming to see critically the way they exist in the world with which and in which they find themselves’ (Freire 2000, italics original, p. 83).” (Diemer et al., 2015, p. 810)

Main Category	Sub Category	Explanation	Example
mind: critical understa nding	feminist pedagogi cal thought	<p>Feminist pedagogical thought is an understanding based on theoretical conceptualisations that was mainly developed by feminists. Hence, ‘feminist’ acknowledges these contributions. We refer to ‘pedagogical’ as an intentioned, formative practice aiming at a formation beyond mere technocratic goals, including the shaping of values and identities. Within feminist pedagogical thought, we focus on the contributions of Critical Thinking, Subjectivity, The Role of Shame and Self-Esteem, and Voice.</p> <p>Critical Thinking. Teachers perceive reflection of attitudes and feelings as meaningful for their teaching practice.</p> <p>Subjectivity. Teachers acknowledge that education is never politically neutral – education either reproduces unjust inequalities or it strives towards justice.</p> <p>The Role of Shame and Self-Esteem. Teachers</p>	<p>"So I really encourage the girls that they can do it because there's always this quick response, 'I can't do any of this and I have no idea about it.' It's always quickly dismissed in general, and I make an effort to give more explicit feedback to the girls. But the boys need that too, of course. I think it would be completely wrong to focus only on the girls; boys need encouragement as well. And when I'm asked if there are female physicists, I naturally affirm that and say, for example, 'me.' But I think you have to keep that in mind when choosing the context and the medium for evaluation, for example. If you have a few boys who are good at programming, they usually get along well with Excel, while others, both girls and boys, don't. But I think you have to keep that in mind when planning and conducting lessons." – Interview 23-1</p>

Main Category	Sub Category	Explanation	Example
		<p>acknowledge that students who face discrimination may enter the classroom “wounded or [with] fragile self-esteem[s] [which] leaves the psyche vulnerable – capable of being shamed.” (hooks, 2003, p. 96) Shame refers to a strong feeling, “an inner sense of being completely diminished or insufficient as a person.” (hooks, 2003, p. 94) Teachers acknowledge the need to allow students to “feel their shame, express those feelings, and do the work of healing.” (hooks, 2003, p. 102)</p> <p>Voice. Teachers analyse from a diversity standpoint questions like “Who speaks? Who listens?” (hooks, 1994, p. 40) and “Who remains silent?” - questions of voice. Voice refers not only to telling one’s experience, but rather highlights who uses “that telling strategically” (hooks, 1994, pp. 148–149), both in terms of when listening to others and interrogating their thoughts as well as when speaking and claiming one’s space.</p>	
mind: critical understanding	identity development	<p>Identity development is an understanding of how students develop or not a career aspiration that is anchored in identity theory. We consider both, the identity development in science, technology, engineering, and mathematics (STEM), and the negotiation of this identity with other identities. Identity development consists of Identity Negotiations, and STEM Identity Development.</p> <p>Identity Negotiations. Teachers identify hidden</p>	<p>"Quite generally, yes, certainly. There are, of course, all kinds of studies on this, but I don't think that's what we should be talking about right now. But in reality, I do have female students who ask me if there are female physicists. So, quite obviously, there is perhaps an identity issue. But at the very least, it is perceived as a very male-dominated field. Yes, I think so. And also in the contexts. What context you choose sometimes appeals more to girls and sometimes to boys. I</p>

Main Category	Sub Category	Explanation	Example
		<p>layers behind actions and/or statements that relate to diversity dimensions such as gender, race, and/or class.</p> <p>STEM Identity Development. Teachers know about STEM identity development and its key concepts “recognition, performance, competence, sense of belonging, supportive relationships, agency, and interest & attitudes” (Çolakoğlu et al., 2023, p. 13).</p>	<p>think gender plays a role there as well. I see it in the French textbooks. Somehow, Paris Saint-Germain, the football club, is constantly in there to appeal to the boys. And in the physics books, I find that if there are two student statements, the woman's statement is always correct. This also indicates that there could be a conflict.” – Interview 23-1</p>
mind: critical understanding	intersectionality	Teachers know about the interconnectedness of the diversity dimensions and the unifying character of this interconnectedness into one struggle.	Not coded
mind: critical understanding	matrix of domination	<p>Matrix of domination is an understanding of the existing inequalities on a system level and where their origin is. Matrix of domination consists of Epistemic Injustice, History Matters, Origin of Inequalities, Perceived Inequality, and Social Construction.</p> <p>Epistemic Injustice. Teachers acknowledge that nowadays definition of what is counted as knowledge and who defines what is knowledge reproduces the matrix of domination (Cernei, 2023; Fricker, 2007). The acknowledgement of epistemic injustice is well in line with appreciation of scientific work and knowledge generation. Yet, the acknowledgement of epistemic injustice allows to criticise the exclusion of certain groups of persons and bodies of knowledge, for example the systematic exclusion and depreciation of indigenous knowledges.</p>	<p>"You can't really pinpoint it like that. Of course, what often stands out is that we have quite a few students whose parents are very well-off, and they sometimes have a different demeanor. They are more confident, but sometimes also overconfident." – Interview 21-1</p> <p>"Well, if they are already better off financially, they sometimes have a better tablet or a better phone in the upper grades, and that naturally enhances their position a bit in the group where the others can't keep up. That is, of course, a point where they might also gain more self-confidence." – Interview 21-1</p> <p>"Yes, that would definitely interest me. I generally see very clearly in school that boys and girls function differently. But I also see fundamentally</p>

Main Category	Sub Category	Explanation	Example
		<p>History Matters. Teachers acknowledge the historically grown and today self-reproducing difference lines along diversity dimensions. Teachers start from the standpoint that we are all part of the reproduction and need to actively unlearn in order to prevent reproduction.</p> <p>Origin of Inequalities. Teachers locate responsibility for diversity mainstreaming and/or existing inequalities, for example with parents, politics, school, and/or society.</p> <p>Perceived Inequality. Teachers know about existing inequalities for the seven diversity dimensions in the society in general and physics education in specific. Teacher know about the legal basis that is relevant in the local context (UN/EU/DE/SH).</p> <p>Social Construction. Teachers perceive the world as socially constructed, especially the diversity dimensions.</p>	very different types of learners." – Interview 25-3
mind: critical understanding	mechanisms of discrimination	<p>In contrast to matrix of domination, mechanisms of discrimination is an understanding of how discrimination and the reproduction of inequalities become manifest in concrete situations. Mechanisms of discrimination consists of (White) Fragility, Habits of Dominance, and Interpersonal Discrimination.</p> <p>(White) Fragility. Teachers know about typical reactions of persons who have not yet built critical consciousness towards being confronted with topics such as for example</p>	"Where it makes a difference, well, it obviously makes a difference in terms of equipment. We have a lot of students who have their own mobile devices that they use in class. They are allowed to do so starting in high school. They simply have better opportunities to use such devices if they receive them as gifts. You can always see that some just bring more with them. Students from [place] usually bring more from home than those from [place]-Downtown. That's just how it is. You also notice the influence of the parents. There

Main Category	Sub Category	Explanation	Example
		<p>racism, hetero-sexism, or classism. A typical reaction can for example be whataboutism or gaslighting.</p> <p>Habits of Dominance. Teachers know about common strategies that are used to dominate and bring oppression into effectiveness, which is most often done unconsciously. Such strategies include for example loud speech, cutting of the speech of other persons, speaking as first person right away without taking a break to think after a question, or using very fast and room-taking gestures.</p> <p>Habitus. Teachers know about cultural reproduction through cultural capital and other forms of capital.</p> <p>Interpersonal Discrimination. Teachers know about stereotypes, prejudices, othering, as well as structural & individual discrimination.</p>	<p>are just some parents who have a very big influence on the school. It's not easy to just give everyone an F. Some have so much influence that you know if someone gets a bad grade, the father might complain, and when he complains, it carries weight. You notice that in school as well, that some people just have a lot of influence due to their status, due to positions in local politics or whatever, that some people simply have influence. And that, of course, also has an impact." – Interview 27-3</p>
mind: critical understanding	transformation narrative	<p>Transformation narrative is an understanding of how society can be transformed. Transformation narrative consists of Allyship, Mass-Based Movements, and Sociopolitical Control.</p> <p>Allyship. Teachers acknowledge both the necessity of centering voices of those most effected, from a diversity perspective the persons who are discriminated against, as well as the "duty and the right" (hooks, 1994, p. 57) of privileged persons to engage in intersectional feminist movements.</p> <p>Mass-Based Movements.</p>	<p>"We are responsible. As a class teacher, I had a high school class for two years. There were two students in it who did not want to be assigned to male or female or felt the opposite. This, of course, needs to be addressed. One has to keep an eye on it, especially in relation to bullying, to ensure it doesn't occur. Or it may need to be discussed if necessary. There is a responsibility, of course. But generally, we try to give boys and girls the same tasks. A math problem or a physics problem is not assigned based on whether it's given to a boy or a girl. That doesn't matter. But as a</p>

Main Category	Sub Category	Explanation	Example
		<p>Teachers know about contemporary mass-based (or aiming-at-being-mass-based) movements of intersectional feminism and their demands for transformation, both on a local and on a global level.</p> <p>Sociopolitical Control. Teachers perceive their actions as having an impact or potentially having an effect of social and/or political change.</p>	teacher, you have to be aware of it." – Interview 21-3
body: critical action	building groups of action	<p>Building groups of action consists of Building Justice Bonds, Critical Behaviour, and Sociopolitical Participation.</p> <p>Building Justice Bonds. Teachers engage with other humans in a way that is respectful of their culture and curious to build bonds, groups of action, and/or social movements towards a practice of freedom.</p> <p>Critical Behaviour. Teachers are involved in activities and/or groups that strive towards justice – and describe this involvement on an abstract level.</p> <p>Sociopolitical Participation. Teachers are involved in activities and/or groups that strive towards justice – and describe this involvement on a concrete level of one particular activity, group, or diversity dimension.</p>	Not coded
body: critical action	dialogic conversation	Teachers promote communication and expression through didactical instruments that provoke critical thinking and justifying argumentations.	Not coded

Main Category	Sub Category	Explanation	Example
body: critical action	embodiment and engaged pedagogy	Teachers actions are guided by compassion focused at students' well-being and feelings. Teachers' actions are based on a reflection of their own needs and feelings. Teachers self-actualise frequently whether their attitudes & actions are in line so that their need of integrity is fulfilled.	"I make every effort to treat everyone equally and not to make any distinctions between genders. I am not biased in that regard and believe that everyone can excel or has the potential to excel. I actually see a small problem in societal and public influences or upbringing. I often hear from parents who are much more tolerant when their daughters perform poorly in physics compared to their sons. I find this completely unnecessary, and there is absolutely no reason to be more lenient in this regard." – Interview 22-1
body: critical action	initial action	Teachers perform small actions that address diversity mainstreaming. These actions are not yet big enough to fulfill the criteria of other sub categories in critical action. However, the sub category initial actions serves to capture actions by teachers that mark first intentioned moves based on understanding and attitude.	"So I try to value all answers. I strive to foster a very positive error culture, and sometimes maybe a bit too positive. I definitely get feedback on that. It often feels like I always say, 'Yes, great answer, but...' instead of just saying, 'That's wrong.' But I try to maintain a very positive error culture. When something comes from the boys, especially from the one more difficult class, the seventh graders, or when some girls speak very softly when answering, which happens quite often, and some boys are much too loud, making it hard to understand them, I get really angry and upset and take it almost to a personal level. I say, 'Hey, that's really disrespectful what you're doing here. It's not just about disturbing one person; it's also about putting them down, and that person might not speak up again.' So I try to address it as directly and actively as possible." – Interview 24-1

Main Category	Sub Category	Explanation	Example
body: critical action	natural science culture	Teachers actively work towards transforming natural science culture so that it becomes effectively inviting for students from all positionalities on diversity dimensions.	"So I try to value all answers. I strive to foster a very positive error culture, and sometimes maybe a bit too positive. I definitely get feedback on that. It often feels like I always say, 'Yes, great answer, but...' instead of just saying, 'That's wrong.' But I try to maintain a very positive error culture. When something comes from the boys, especially from the one more difficult class, the seventh graders, or when some girls speak very softly when answering, which happens quite often, and some boys are much too loud, making it hard to understand them, I get really angry and upset and take it almost to a personal level. I say, 'Hey, that's really disrespectful what you're doing here. It's not just about disturbing one person; it's also about putting them down, and that person might not speak up again.' So I try to address it as directly and actively as possible." – Interview 24-1
body: critical action	passion of experience	Passion of experience refers to the emotionality students can develop when connecting their experiences to the topic at hand, in our case from natural sciences. Placed in critical action, we use the sub category passion of experience to capture teachers' actions that aim at unfolding that passion of experiences in students. To be more precise, teachers invite students to tell their experiences in the classroom in order to nurture a culture of appreciation of each unique voice, of listening to each other, and of highlighting the subjectivity of standpoints as well as the need for diversity mainstreaming. "This doesn't	Not coded

Main Category	Sub Category	Explanation	Example
		mean [... to] listen uncritically or that classrooms can be open so that anything someone else says is taken as true, but it means really taking seriously what someone else says." (hooks, 1994, p. 150) Still, it can be necessary to interrupt students and to ask how that relates to the issue at hand.	
body: critical action	reflection	Teachers frequently reflect in a structured and systematic way in order to self-actualise when necessary.	"So, for something to have changed, I would have had to actively think about it in these weeks. I'd say the first interview might have given me the idea that I should do something. But specifically for this class, I already had a complete lesson plan. So, I couldn't have taken any action, so to speak. Therefore, I didn't delve further into it. I believe I didn't have any learning opportunity, so to speak. Except for the one instance where it was made clear to me that one should perhaps pay attention and listen to it. But in these three weeks, I didn't really have the chance to be pointed out or reminded about it. I think if I had seen examples, like, this is how you could do it, I would probably be more articulate about it now." – Interview 23-3
body: critical action	resilience	Teachers can navigate situations where full engagement would over-demand their own resources. Resilience includes the ability to prioritise in these situations on the one hand up to a level that allows reproduce their resources and at the other hand that aims for highest possible impact. Resilience includes as well to not withdraw completely in the face of difficulty to reach change but to stay engaged,	"I try to provide worksheets as much as possible, where they can write down their answers. This means they don't need extra sheets of paper. They know that they can ask me for pens at any time. Sometimes it starts right there. However, I can't supply materials for 100 students; that would exceed the limits eventually. But I try, within the scope of what we have available at the school, to step in a bit and provide support." – Interview 26-1

Main Category	Sub Category	Explanation	Example
		even though if no immediate impact is visible. In the context of education, resilience includes the ability to navigate between 1) preparing students for exams in line the natural science education standards as an answer to the selection function of the educational system and 2) aiming for pluriversal structures and social justice along diversity dimensions.	
heart: critical attitude	commitment	Teachers have a “a moral concern with inequity, motivation to address it, and perceived ability to make a difference”. (Diemer et al., 2015, p. 815) The attitudes manifests in a way beyond mere understanding that inequities play a role in natural science education. Concretely, teachers do not only acknowledge that role but make the moral concern their own, showing commitment to address that inequities.	"I make every effort to treat everyone equally and not to make distinctions between genders. I am not biased in that regard and believe that, in principle, everyone can excel or has the potential to excel. I actually see a small problem in societal and public influences or upbringing. I frequently hear from parents who are much more tolerant of their daughters performing poorly in physics compared to their sons. I find this completely unnecessary, and there is absolutely no reason to be more lenient in this regard." – Interview 22-1
heart: critical attitude	education as the practice of freedom	Teachers view education as a “counter-hegemonic act” (hooks, 1994, p. 2), with this act being rooted in diversity mainstreaming. Teachers actively “challenge the ‘banking system of education’, that approach to learning that is rooted in the notion that all students need to do is consume information fed to them by a professor” (hooks, 1994, p. 14). Finally, teachers, understand themselves as learning from their students as well. Education as the practice of freedom is an attitude, because 1) we do not focus on capturing actions, and 2) education as the practice of	"So, in this unit, the topic was predefined as solar cells. Therefore, I think the main area I would work on is the context. I wouldn't know what to change there because certain contexts were already chosen. Regarding the dashboard, I would clearly say no. I don't want to create an overview that shows that the girls are performing worse. I actually don't want that. It almost reinforces the problem. If some teachers see, 'Look, I've always known that girls perform worse,' and then they get an overview that confirms this, and it happens to be true for this class for whatever

Main Category	Sub Category	Explanation	Example
		freedom goes beyond mere understanding but involves an own positionality as educator for freedom.	reasons, I wouldn't want to know it reinforced. I would find it odd, even to the point of being uncomfortable, if the dashboard were split by boys and girls." – Interview 23-3
heart: critical attitude	pluriverse	Teachers acknowledge the importance of standpoints, the matrix of oppression, and intersectionality. Teachers understand a space that invites students from all positionalities on diversity dimensions as a vision, allocate responsibility to create that space within the school system & themselves (rather than to the students who face discrimination by, for example, developing coping strategies), and know some important re-configuration screws within the school systems. A pluriversal attitude includes the critique of deficit-oriented framings (Cheuk, 2021; Kayumova & Dou, 2022) – instead of deficit-orientation, heterogeneity that diverse students bring is understood as richness, and the responsibility to unfold diversity is located with the institutions, not with the students who face discrimination.	"No, gender is a category of division. But there are generally different types of learners. Some prefer everything to be presented to them in a straightforward manner, with key points written on the board, and they manage well with that. Both boys and girls can fit this type. Then there are others who prefer to discover things on their own, to experiment and try things out themselves. And, as I said, there are those who don't want to experiment at all; they prefer to be told. At the end of the day, you can never be entirely sure which learning group each student belongs to. My theory is that you need to offer different types of instruction so that there's something for everyone. And they don't even need to be put into boxes. I also function differently from time to time. Sometimes I want to be engaged in one way, and other times in another, in terms of learning opportunities." – Interview 25-3
heart: critical attitude	positionality on culture	Teachers can critically position with regard to culture. This includes a) a fundamental rights perspective on what is not justifiable with culture, namely anything that clashes with human rights, b) a critical listening to when culture is	Not coded

Main Category	Sub Category	Explanation	Example
		used to replace for example race in order to keep racism working, and c) a non-hierarchical perception of the plurality of existing cultures combined with the awareness of the discriminating hierarchical readings of multiple cultural performances that exist in societies. Teachers understand culture as dynamic, meaning that it evolves and changes over time, and fuzzy, meaning that borders of cultural communities may be hard to specify.	
heart: critical attitude	positionality on identity	Teachers are able to explicitly & critically reflected position themselves on the diversity dimensions in terms of their own identities.	Not coded
heart: critical attitude	willingness to change	Teachers are willing to change and to be changed in their attitudes.	"I think I know how I answered that. No, I would actually like to do it more. I'd like to enhance it, as I mentioned earlier, with different contexts. The problem is that I might venture into contexts where I'm not very familiar myself. In high school, I always talk about sports and amusement parks; they might want to hear about other contexts now. I need to get a bit more creative with that." – Interview 24-3
factors of change		Concrete issues facilitate the development of critical consciousness.-	Not coded
formalia and facts	experience in teaching	We report the years that teachers have practiced their profession.	Not coded
formalia and facts	gender (read)	We report the read gender as we did not ask the teachers to name their own gender directly. We are aware of the risk to read gender in binary categories only but believe a reported read gender to potentially add meaning to the analyses that might provide	Not coded

Main Category	Sub Category	Explanation	Example
		insights beyond not reporting it.	
formalia and facts	positional ities on diversity dimensions	We report on the dimensions gender, race, ability, religion & beliefs, sexual identity, age, and class – in case no information is available, we report the missing information.	"With my personal history, perhaps in that it has always been relatively unimportant to me whether someone is male or female. I need to like the person if I want to maintain a closer relationship with them." – Interview 26-3
formalia and facts	school form	"Gemeinschaftsschule" / "Gymnasium"	"When it comes to content differentiation, absolutely. Especially in the community school, where we have both comprehensive and secondary schools, the teaching is differentiated anyway. This means that students can work either at the B-level or at a remedial level. Since they are all in the same class, we always have differentiated instruction and materials, and we meet them where they are." – Interview 22-1

References of the Piece of Scholarship

- Archer, L., Calabrese Barton, A., Dawson, E., Godec, S., Mau, A., & Patel, U. (2022). Fun moments or consequential experiences? A model for conceptualising and researching equitable youth outcomes from informal STEM learning. *Cultural Studies of Science Education*, 17, 405–438. <https://doi.org/10.1007/s11422-021-10065-5>
- Archer, L., Dawson, E., DeWitt, J., Seakins, A., & Wong, B. (2015). “Science Capital”: A Conceptual, Methodological, and Empirical Argument for Extending Bourdieusian Notions of Capital Beyond the Arts. *Journal of Research in Science Teaching*, 52(7), 992–948. <https://doi.org/10.1002/tea.21227>
- 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>
- Baker, R., & Hawn, A. (2021). Algorithmic Bias in Education. <https://doi.org/10.1007/s40593-021-00285-9>
- Brickhouse, N. W. (2001). Embodying Science: A Feminist Perspective on Learning. *Journal of Research in Science Teaching*, 38(3), 282–295. [https://doi.org/10.1002/1098-2736\(200103\)38:3%3C282::AID-TEA1006%3E3.0.CO;2-0](https://doi.org/10.1002/1098-2736(200103)38:3%3C282::AID-TEA1006%3E3.0.CO;2-0)
- Brown, B. A. (2004). Discursive Identity: Assimilation into the Culture of Science and Its Implications for Minority Students. *Journal of Research in Science Teaching*, 41(8), 810–834. <https://doi.org/10.1002/tea.20228>
- Carlone, H. B., & Johnson, A. (2007). Understanding the Science Experiences of Successful 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>
- 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>
- Çolakoğlu, J., Steegh, A., & Parchmann, I. (2023). Reimagining informal STEM learning opportunities to foster STEM identity development in underserved learners. *Frontiers in Education*, 8, 1–16. <https://doi.org/10.3389/feduc.2023.1082747>
- Costanza-Chock, S. (2020). Design justice: Community-led practices to build the worlds we need. The MIT Press.
- Diemer, M. A., McWhirter, E. H., Ozer, E. J., & Rapa, L. J. (2015). Advances in the Conceptualization and Measurement of Critical Consciousness. *The Urban Review*, 47, 809–823. <https://doi.org/10.1007/s11256-015-0336-7>
- 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 STEM identity: The importance of talking science. *Science Education*, 103, 623–637. <https://doi.org/10.1002/sce.21499>

- Düchs, G., & Ingold, G.-L. (2018). Frauenanteil bleibt stabil. *Physik Journal*, 17(8/9), 32–37.
- 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>
- Escobar, A. (2017). *Designs for the Pluriverse: Radical Interdependence, Autonomy, and the Making of Worlds*. Duke University Press.
<http://www.jstor.org/stable/j.ctv11smgs6>
- EU Charter of Fundamental Rights. (2012). EU. <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:12012P/TXT>
- Freire, P. (1970). *Pedagogy of the Oppressed*. Penguin Random House UK.
- Gago, V. (2019). La potencia feminista. O el deseo de cambiarlo todo. *Traficantes de Sueños*.
https://traficantes.net/sites/default/files/pdfs/TDS_map55_La%20potencia%20feminista_web.pdf
- Götschel, H. (2015). *Queere Physik?! In Sexuelle Vielfalt im Handlungsfeld Schule*. transcript verlag.
- hooks, bell. (1994). *Teaching to Transgress—Education as the Practice of Freedom*. Routledge. <https://doi.org/10.4324/9780203700280>
- hooks, bell. (2003). *Teaching Community—A Pedagogy of Hope*. Routledge.
<https://doi.org/10.4324/9780203957769>
- hooks, bell. (2009). *Teaching Critical Thinking—Practical Wisdom*. Routledge.
<https://doi.org/10.4324/9780203869192>
- Hott, J. (2024). *Resource Model*.
- Jemal, A. (2017). Critical Consciousness: A Critique and Critical Analysis of the Literature. *The Urban Review*, 49, 602–626. <https://doi.org/10.1007/s11256-017-0411-3>
- Kayumova, S., & Dou, R. (2022). Equity and justice in science education: Toward a pluriverse of multiple identities and onto-epistemologies. *Science Education*, 106, 1097–1117. <https://doi.org/10.1002/sce.21750>
- KMK. (2004). *Bildungsstandards im Fach Physik für den Mittleren Schulabschluss*. Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland.
- Kuckartz, U. (2018). *Qualitative Inhaltsanalyse: Methoden, Praxis, Computerunterstützung* (4. Auflage). Beltz Juventa.
- 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>
- MBWK SH. (2019). *Fachanforderungen Physik*. Ministerium für Bildung, Wissenschaft und Kultur des Landes Schleswig-Holstein.

- McCausland, J., & McDonald, S. (2024). White shame and white ambivalence in learning to be a well-started White anti-racist science teacher. *Journal of Research in Science Teaching*, 1–29. <https://doi.org/10.1002/tea.21946>
- Mecheril, P., Olalde, O. T., Melter, C., Arens, S., & Romaner, E. (2020). *Migrationsforschung als Kritik?: Konturen einer Forschungsperspektive*. <https://dx.doi.org/10.1007/978-3-531-19145-4>
- MNC. (2021). Marco Nacional de Cualificaciones. Marco Nacional de Cualificaciones. <https://www.cualificaciones.cr/mnc/>
- Muñoz Izquierdo, C. (2012). Tres problemas fundamentales del sistema educativo. *Perfiles educativos*, 34(SPE), 154–159.
- O'Connor, C. (1997). Dispositions Toward (Collective) Struggle and Educational Resilience in the Inner City: A Case Analysis of Six African-American High School Students. *American Educational Research Journal*, 34(4), 593–629. <https://doi.org/10.3102/00028312034004593>
- 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.
- Rahm, J., & Moore, J. C. (2016). A case study of long-term engagement and identity-in-practice: Insights into the STEM pathways of four underrepresented youths. *Journal of Research in Science Teaching*, 53(5), 768–801. <https://doi.org/10.1002/tea.21268>
- Tiðberger, M. (2017). *Critical Whiteness—Zur Psychologie hegemonialer Selbstreflexion an der Intersektion von Rassismus und Gender*. Springer.
- Watts, R. J., Diemer, M. A., & Voight, A. M. (2011). Critical consciousness: Current status and future directions. *Youth Civic Development: Work at the Cutting Edge*, 2011(134), 43–57. <https://doi.org/10.1002/cd.310>
- 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>

Arte de Mirar Atrás

*Holer la belleza de una veranera,
Aspirar en profundidad su atmósfera –*

*Ya no soy la persona quien era:
¡Qué gratitud pensando en nuestras experiencias!,
¡Auténtica!, vos irritándome,
Invitándome
A conocer otros mundos con sus vigencias;
Ya no estarás en mi vida diaria,
Yo ya no viviré en Moravia –*

*Pero, ¡raro!, lo más que lo intento,
Es que no lo logro, no encuentro
Ni un pensamiento pesado hoy,
Está esa veranera solamente,
Ella escribió quien soy
Lo que me hace feliz, ya que es fuente
De agua, ¡dulce y salada!, desde nuestros recuerdos –*

*Más que tristeza siento una profunda necesidad por conocernos:
¡Inclinarme ante el arte de como transformaste mis acuerdos!*

7 General Discussion

7.1 Construct-Specific Discussions over Findings from all Pieces of Scholarship

7.1.1 De-Biasing: Addressing Bias – Mitigation Possible, Elimination Out of Sight

How can De-Biasing contribute to structurally address existing inequalities in physics education in the context of the rising use of artificial intelligence systems in Northern Europe?

In order to address the first research question, I analysed the four pieces of scholarship with regard to what can be learnt from them for De-Biasing as shown in Figure 7-1. Each column represents one piece of scholarship. From this analysis, four themes emerged that can be condensed into the following questions: Which implications do the existing inequalities in physics education have for De-Biasing? How can the gap be bridged between existing principles for De-Biasing and little impact on practice? In how far can De-Biasing address the problem of biased artificial intelligence system? Which implications do the findings on Critical Consciousness have for De-Biasing?

Study 1 Theoretical Model	Study 2 Concrete Case	Study 3 Empirical Qualitative Study	Study 4 Empirical Quantitative Study
a) Special Focus Needed: STEM Identity and Under-Served Students	a) Principles Without Practice	a) Certain Concern and Initial Actions	a) Uncertainty about (Most) Effective De-Biasing
b) Vulnerability and Iterability in STEM Identity Work	b) Domain-Specific Standards Needed	b) Major Obstacles: Feminist Pedagogical Thought, Education as the Practice of Freedom, & Reflection	b) Regulation of Training Datasets is One Promising Piece for De-Biasing
c) Obligation to Act and Accountability	c) Domain-Specific: Historically Grown Inequalities, Evaluation Categories, and Normative Standpoints	c) No Well-Established Critical Action	c) First Indication and Evidence: Slicing and Training Datasets Point in the Same Direction
d) Suppositions for Bias and Equity each – Recognition, Performance, Competence	d) Additional Counter-Measures Needed for De-Biasing		d) Needed: Methodological Innovation and Evaluation Criteria Beyond A/B-Testing

Figure 7-1 - Results from all pieces of scholarship with relevance to De-Biasing – I refer to the fields by indicating the field, for example 'DB-1-a' refers to the results on De-Biasing from this figure, piece of scholarship 1 the theoretical model, and the field a) special focus needed: STEM identity and under-served students

The existing inequalities have relevant implications for De-Biasing. First of all, a special focus on STEM identities (not only competence!) of under-served students (not all students, but students who are under-served from a diversity perspective) is needed

(DB-1-a)²². The implication for De-Biasing is highly relevant as it moves implications of recognition in the focus (DB-1-b). Without this analysis, one could argue that focusing on the correct identification of the students who did not yet achieve a competence yet is most important because these are the students who need targeted support the most. From an identity perspective however, it becomes increasingly important to also focus on the correct identification of the students who achieved a competence in order to provide recognition to them, be it through direct feedback from the system or through a teacher facing dashboard which invites the teacher to provide feedback. Hence, a careful adjustment of what to optimise De-Biasing on is needed.

Artificial intelligence systems' potentials shall be harvested while threats of bias shall be prevented effectively. The approach of Responsible Learning Analytics allows to navigate between these two goals by acknowledging that learning analytics come with (1) potentials that lead to an Obligation to Act, and (2) threats that are addressed through Accountability (DB-1-c). For physics education in Germany, we found that the current problem with the existing principles is that they are not concrete and mandatory enough to address the threats of biased algorithms when using learning analytics (DB-2-a). What is needed to make principles more concrete are domain-specific and context-sensitive standards (DB-2-b). Context-sensitivity is necessary because a summative assessment could demand to balance for the "benefit of doubt" (Jeong et al., 2021), while a summative setting as ours may rather demand for a correct detection. Domain-specificity is needed because the task of De-Biasing is so broad that addressing all possible threats in all domains would put so much stress on the design of artificial intelligence systems that the potential could not be harvested anymore. Such a potential under-use of artificial intelligence systems is widely discussed already (Floridi et al., 2018, p. 690). Hence, a focus is necessary – on the relevant historically grown inequalities, evaluation categories, and concrete normative standpoints (DB-2-c). Based on these analyses, we developed six suppositions for two normative focal points and different standpoints that can guide next steps with a focus on the three STEM identity dimensions of recognition, performance, and competence (DB-1-d). Our empirical evidence indicates that methodological innovation and evaluation criteria beyond A/B-testing of more or less competence on a whole-class-average are necessary (DB-4-d). The methodological innovation is necessary in order to produce scientific evidence that can inform policy making through actionable results. The A/B-testing refers to focusing not on simple evaluation categories such as "students have learnt more on average with the artificial intelligence system in place". Instead, a focus on recognition and under-served students is necessary in order to effectively reduce the existing inequalities.

Our empirical evidence indicates promising potentials for De-Biasing on the one hand. On the other hand, our findings indicate limits of De-Biasing and known remaining biases even in De-Biased systems. For instance, our findings are not clear that any technique alone is De-Biasing completely or even the most effective way for De-Biasing in general (DB-4-a). We already knew from literature on De-Biasing that there is no single strategy that is most effective for all settings (L. Li et al., 2023, p. 505). Nonetheless, De-Biasing through regulation of training datasets seems to be one promising piece of a regulatory framework for physics education that most effectively prevents threats in terms of discrimination and enables both, potentials in terms of learning and diversity mainstreaming (DB-4-b) – which is well in line with findings from more general literature as well (L. Li et al., 2023, p. 506).

²² I refer to the fields by indicating the field, for example 'DB-1-a' refers to the results on De-Biasing from this figure, piece of scholarship 1 the theoretical model, and the field a) special focus needed: STEM identity and under-served students.

However, our results indicate that existing biases that cannot be addressed by the regulation of training datasets alone. Comparing our results of training dataset analyses with our slicing analyses, we find first indications that biases in the slicing analyses seem to be stronger where patterns in the student answers can be found in the training dataset analyses, in our case for example along educational background (DB-4-c). Little reliability of prediction in the training dataset analyses combined with little impact of slicing configurations. That indicates that training dataset analyses could serve as an effective instrument for risk prediction of bias in a regulatory framework. Finally, always analysing all biases neither seems feasible nor doable, especially with a look at intersectionality and De-Biasing (DB-2-d). Instead, we propose working towards domain-specific non-De-Biasing counter-measures against discrimination as most effective way to address some biases, especially multi-faceted intersectionalities. Whether addressing threats by De-Biasing or implementing counter-measures is more efficient is strongly context dependent. Both can be viable tools to reduce historically grown inequalities.

De-Biasing can build upon a certain Critical Consciousness. However, no well-established critical action but instead initial actions and first critical actions indicate a potential as well as a need for De-Biasing to fill the gap of missing recognition through teachers. We found evidence for a certain concern to not act discriminatory which we understand as a resource of physics teachers in Northern Europe (DB-3-a). Additionally, all teachers reported at least initial actions, namely actions aiming to address different needs due to different positionalities on at least one diversity dimension. The major obstacles that can be addressed in professional developments in our piece of scholarship are found in the following sub categories of Critical Consciousness: the critical understanding of feminist pedagogical thought, the critical action of education as the practice of freedom, and the critical action of reflection (DB-3-b). Finally, we did not find any well-established critical action (DB-3-c). No well-established critical action while having inequalities that operate especially at the level of recognition creates an increased demand for De-Biasing of artificial intelligence systems.

De-Biasing can mitigate negative effects of systems of automated decision making, but it cannot eliminate all threats and some of the threats remain even in De-Biased systems.

7.1.2 Critical Consciousness: A Promising Piece for a Social Justice Architecture

How can Critical Consciousness contribute to structurally address existing inequalities in physics education in the context of the rising use of artificial intelligence systems in Northern Europe?

In order to address the second research question and analogue to the analyses on De-Biasing, I analysed the four pieces of scholarship with regard to what can be learnt from them for Critical Consciousness as shown in Figure 7-2. From this analysis, four themes emerged that can be condensed into the following questions: What can be learnt with regard to the existing inequalities in physics education? What role does identity work play in addressing these inequalities? Which implications does the growing use of artificial intelligence systems have for the Critical Consciousness of teachers? And finally, which

types of Critical Consciousness can be found in the cultural context of Northern Europe and what can be learnt from that for professional developments for Critical Consciousness?

Study 1 Theoretical Model	Study 2 Concrete Case	Study 3 Empirical Qualitative Study	Study 4 Empirical Quantitative Study
a) Rather Stable Historically Grown Inequalities	a) Principles Without Practice	a) Certain Concern and Initial Actions	a) Uncertainty about (Most) Effective De-Biasing
b) Special Focus Needed: STEM Identity and Under-Served Students	b) Domain-Specific: Historically Grown Inequalities, Evaluation Categories, and Normative Standpoints	b) Major Obstacles: Feminist Pedagogical Thought, Education as the Practice of Freedom, & Reflection	b) Strong Ethical Basis Needed for Creation of Meaningful Evidence
c) Vulnerability and Iterability in STEM Identity Work	c) Additional Counter-Measures Needed for De-Biasing	c) No Well-Established Critical Action	
		d) Little Competence in Dealing with Potential Biases	
		e) First Critical Actions' Characteristics: Structural Responsibility and More Instances	
		f) Need for Both, Critical Consciousness and De-Biasing	

Figure 7-2 - Results from all pieces of scholarship with relevance to Critical Consciousness – I refer to the fields by indicating the field, for example 'CC-1-a' refers to the results on Critical Consciousness from this figure, piece of scholarship 1 the theoretical model, and the field a) rather stable historically grown inequalities

In physics education, persistent historically grown inequalities exist. These inequalities exist along multiple dimensions of diversity, for example gender, race, and class (CC-2-b). The inequalities are domain-specific not only in terms of, for example, differing levels of inequality, but also in terms of which evaluation categories are most relevant. Given seven protected diversity dimensions in the European Union and domain-specific most relevant evaluation categories, a strong ethical basis is needed when addressing historically grown inequalities in order to reach actionable results and thereby allow for evidence-informed political decision making (CC-2-b). The generation of evidence, for example, needs to happen within clearly defined boundaries of which diversity dimensions to analyse for based on an analysis of historically grown inequalities in the specific domain. It also needs to be clear which justice understanding is trying to be reached – should physics classes only not add on the existing inequalities by discrimination or should physics classes even actively strive towards less inequalities and representation of all students in physics as in the entire society? Having a well-defined normative framework is increasingly important given the complex task of De-Biasing algorithms (CC-4-b). The role of science is to offer well-defined theoretical normative frameworks and for each of that framework the evidence of how to best implement the given normative framework. Which framework is the best is then a political task that science cannot answer. These theoretical normative frameworks are particularly relevant for Critical Consciousness, as they define the normative direction and the problems that shall be addressed by Critical Consciousness. In our pieces of scholarship, we have developed explicit theoretical normative frameworks of responsible learning analytics in physics education that are rooted in the pluriverse as well as in the physics- or STEM-specific evaluation categories. The evaluation categories are particularly interesting because inequalities in physics education do not mainly operate at

the level of more or less competence development and a process of selection based on that. Instead, the inequalities operate at the level of identity development (CC-1-a).

The core supposition derived from the first piece of scholarship on the theoretical model is that in order to effectively address historically grown inequalities in physics education, all dimensions of STEM identity development need to be addressed. Identity development consists of, next to the dimension of competence, the dimensions of recognition and performance (Carlone & Johnson, 2007). For recognition and performance, we deduced that a special focus on under-served students is needed due to the mechanisms of vulnerability and iterability (CC-1-b). Vulnerability is the particular dependency of under-served students on recognition due to the lower level of recognition that has been offered to them so far (CC-1-c). Iterability is, from a theoretical perspective, the norm-setting and potentially exclusive process of, for example, presenting mainly male physicists over and over again and thereby establishing physics as something for men only (CC-1-c). As in the past many physicists used to be men, for example, careful counter-action of critically conscious teachers is needed if historically grown inequalities are not to be reproduced through difference in STEM identity developments. An additional and more general example that underlines the necessity to address historically grown inequalities at the level of STEM identity development due to the relevance of these inequalities are the findings on identity negotiations in racialised power structures which we encounter in the context of Germany and Northern Europe (M. Eggers, 2005). Another potential process of reproduction of historically grown inequalities that critically conscious teachers need to deal with has its origin in the growing use of artificial intelligence systems: the threats of biases.

Prior to our own investigations, various threats of biases have been established in a range of research fields. At the same time, established principles and codes of conduct for practitioners had very little impact on the practice of actually De-Biasing algorithms. We therefore started with a case study asking: Where exactly do the principles fail to provide clear guidance? We found that the existing principles were not concrete and mandatory enough to address the threats of biases – leaving open questions such as which diversity dimensions to analyse for or which justice definition to take for programming code (CC-2-a). A solid and very concrete core of guidance is needed in order to bring ethical considerations into praxis by making principles actionable. At the same time, always analysing all biases neither seems feasible nor doable, especially with respect to intersectionality and De-Biasing (CC-2-c). Adding on that, our quantitative evidence about most effective De-Biasing is far from clear in terms of what is the most effective technique for De-Biasing. We could gather further evidence that indicates that carefully choosing training datasets can successfully prevent biases (CC-4-a). However, our evidence indicates that artificial intelligence systems in physics education cannot be De-Biased by training dataset regulation alone and it remains unknown whether training dataset regulation is the most efficient way of De-Biasing. From our normative standpoint of departure discrimination through biases (also intersectional discrimination) should not exist. As a logical consequence of the currently missing effective De-Biasing, a need for additional counter-measures against the threats of biases that cannot be meaningfully addressed with a reasonable effort can be derived – another foundation for a call for the development of Critical Consciousness of teachers. Even so, we found little self-confidence and competence in estimating the threats of bias in artificial intelligence systems in the interviews that we conducted with teachers (CC-3-d).

The findings from our interviews with teachers indicate that Critical Consciousness holds a potential as effective and justice-goal-rooted counter-measure in the cultural context of

Norther Europe. In all interviews, we found a certain will to make physics more diverse as well as initial actions directed towards addressing, for example, gender-specific student needs – both of which can be powerful resources that professional development of Critical Consciousness can built upon (CC-3-a). However, we did not learn about well-established critical action – teachers mainly reported rather superficial interventions with little potential to effectively reach transformation (CC-3-c). The major obstacles that we identified in the interviews are little feminist pedagogical thought, no attitude of education as the practice of freedom, and little established practice of reflection (CC-3-b). Feminist pedagogical thought contains, for example, acknowledging the important role of who does (not) speak in the classroom and thereby receives recognition as well as the role of shame and self-esteem in the STEM identity development of students. In the interviews, we could distinguish two types: Teachers who reported only initial actions on the one hand and teachers who reported first critical actions as well (CC-3-e). Those teachers who reported first critical actions especially had an attitude that allocated the responsibility to counter historically grown inequalities rather with the system, for example the school or themselves, instead of allocating it with the students themselves. In addition to that, the teachers who reported first critical actions named way more examples of critical understandings and attitudes as well – without having a well-established Critical Consciousness for all categories. From our interviews, we derived a need for both: De-Biasing artificial intelligence systems wherever possible and establishing Critical Consciousness of teachers at the same time (CC-3-f). De-Biased artificial intelligence systems and critically conscious teachers are crucial as they can provide meaningful recognition which especially under-served students need in order to develop STEM identities which is a foundation to reduce historically grown inequalities in physics education. I summarise our findings speaking from our theoretical normative framework as follows:

Structural interventions against the vicious cycle of reproduction of historically grown inequalities in physics education are needed – in the face of biased artificial intelligence systems even more than ever. Critical Consciousness seems to be one promising part of an effective and efficient structural intervention.

7.1.3 STEM Identity Development in a Pluriverse: Action Needed

The aim of this section is zooming out and asking the big questions in order to discuss the findings on the two concrete constructs in front of the bigger goal of STEM identity development in a pluriverse. Reviewing promises and potential downfalls is only possible from a given normative standpoint. Hence, I recall the key justice characteristics of the pluriverse. Next, I discuss the potentials for interactions between De-Biased artificial intelligence systems and Critically Conscious teachers that go beyond the potentials both of them would have alone. Finally, I discuss the potentials and limitations of the presented pieces of scholarship in contributing to the major challenge on the way towards a pluriverse.

What are the promises and potential downfalls of addressing historically grown inequalities in physics education in Northern Europe

through Critical Consciousness and De-Biasing for STEM identity development for all students in a pluriverse?

The pluriverse in our definition focuses on justice along diversity dimensions. This focus is well suited within a Northern European context where the German constitution protects six and the EU-Charter seven explicitly named diversity dimensions – which makes findings relevant and actionable for political decision making. Next to the “Justice for whom?”, the “Justice of what?” is defined by analysing for who does (not) develop a STEM identity. STEM identity development with a focus on access to careers and thereby power is well in line with our analysis of inequalities in physics education in Northern Europe where differences often are not visible in competence achievement but instead in career choice. Additionally, the pluriverse comes with a clear goal of reaching a representation in physics of the actual shares of sub-populations in the society as a whole. Finally, the pluriverse assigns the responsibility for transformation with the system instead of the individual. These decisions can but do not have to be in conflict with other relevant normative standpoints in physics education. When designing interventions in order to transform physics education towards a world where many worlds fit, the complex character of identity development needs to be considered.

Identity work is a complex process with many actors involved (Brickhouse, 2001; Brown, 2004). The complex character makes it hard to predict effects on identity development due to possible consequences of one single intervention beyond the effect of the intervention alone. For example, a student that is recognised by an artificial intelligence system can proudly walk up to the teacher and report the success. The report to the teacher can make the teacher change the perception of the student as more competent than assumed, thereby triggering more recognition in the future. Additionally, the student could also report the success to parents outside of the school. It is also possible that the student considers an offering for an out-of-school activity such as a student laboratory because the student now has a stronger awareness of the own competence. What I want to say is that even if the De-Biased artificial intelligence system alone does not lead to a direct increase of the STEM identity of the student, it may nonetheless lead to an increased willingness to participate in activities that then can lead to an increase in STEM identity development. Also, De-Biased artificial intelligence systems can lead to more instances of recognition, even beyond the increase in recognition through the artificial intelligence system itself. Identity work is a complex process that takes place in an environment with manifold interaction effects which holds threats but also potentials for impacts beyond the single intervention alone.

Our four pieces of scholarship can provide some guidance but obviously cannot provide evidences for all questions where evidence would be needed. We focused a lot on the how, contributing rather theoretical and qualitative findings. The quantitative evidence we have is too small to allow for final conclusions and allows for conclusions on the level of indications only. However, a solid qualitative and theoretical framework is needed in order to address questions of justice that come with a high level of complexity. In our findings, we could show how Critical Consciousness and De-Biasing can play a key role in reaching a pluriverse. These theoretical frameworks, qualitative descriptions of typologies with their resources and obstacles, and first quantitative evidences on De-Biasing from a pluriversal stance lay a powerful foundation to continue research on reducing inequalities in physics education in Northern Europe. I summarise promises and potential downfalls as:

Critical Consciousness and De-Biasing have the potential to play a key role in reaching a pluriverse in physics education in Northern Europe. A pluriversal standpoint with a focus on STEM identity development provides an actionable framework for research on questions of justice in Northern Europe.

7.2 Implications

7.2.1 Research

For De-Biasing, our evidence indicates some but limited potential for De-Biasing of artificial intelligence systems in physics education in Northern Europe. We found biases for almost all slicing configurations. First of all, more evidence is needed in order to see whether our findings on slicing analyses and training dataset analyses can be further validated with other datasets. Further investigation on the origin of these biases for improved understanding is necessary. One well-established approach is analysing through instruments of explainability. However, we see it necessary to not only use models that operate at the level of words but also consider more complex linguistic features such as grammatical structures. Otherwise the limitations of an analysis can obscure causal effects by not analysing for them and come with the risk of pointing in wrong directions or drawing wrong conclusions. Secondly, we coded the student answers with multiple researchers. Analysing whether the biases entered through one particular coder seems promising – especially because we validated inter-coding only on the level of the entire group instead of multiple sub-groups level, for example for under-served students. Finally, our findings indicate relevant limits for both, De-Biasing and Critical Consciousness. Acknowledging that while sticking to the goal of a pluriverse in physics education in Northern Europe has relevant implications for De-Biasing: We need structural counter-measures that against the threats artificial intelligence systems introduce. For the artificial intelligence systems, it gets more important to investigate what their active and positive contribution could be for a pluriverse. For STEM identity development, we know that context and interest for the context are highly relevant. Supporting teachers in adjusting their topic quickly to a context that is interesting especially for the under-served students in their class through generative artificial intelligence systems holds promising potentials. However, generative artificial intelligence systems are known to come with the risk of hallucinating (Ho et al., 2024). In education, systematically wrong instructions cannot be tolerated. One upcoming approach to address these risks of generative artificial intelligence systems is retrieval augmented generation: Prompting the system with a given knowledge base that it should exclusively use for generation (K. Li & Zhang, 2024). We believe that enabling the use of retrieval augmented generation systems holds a promising potential for pluriverse-directed interventions because of two characteristics: 1) it can be directed towards relevant categories for STEM identity development such as interest and context, and 2) it connects well with teacher-addressing approaches such as the development of Critical Consciousness and the time constraints teachers face in their practice. The combination of all these approaches, De-Biasing as well as active and positive contributions, could be summarised under the umbrella goal of social justice in artificial intelligence systems in education.

For Critical Consciousness, we have well-described historically grown inequalities in physics education. In addition to that, we have a culturally sensitive instrument for qualitative assessment of Critical Consciousness as well as first qualitative empirical

evidences that indicate promising potentials for under-served students' STEM identity development. However, it remains unknown how exactly teachers' Critical Consciousness effects under-served students' STEM identity development. Exploring the effects in qualitative depth can help to 1) understand not only contributions on a macro-level but also the under-lying processes on a micro-level, and 2) provide a valuable source of instances for professional developments of Critical Consciousness. Next to these effects, exploring the culture-specific aspects of the development of Critical Consciousness seems urgent in order to create an evidence basis to inform the most effective way of raising teachers' Critical Consciousness. As Critical Consciousness is a construct that includes attitudes and actions, special care needs to be taken to investigate the long-lasting and sustainable effects of such a development. Last but not least, for up-scaling quantitative instruments are necessary in order to have a solid foundation for a study on the effect of teachers' Critical Consciousness on the STEM identity development of under-served students. Such a study is needed as evidence in order to decide evidence-informed whether Critical Consciousness does not only seem to be but really is (or not) a central tool to reduce historically grown inequalities in physics education in Northern Europe.

Looking beyond the two concrete constructs of De-Biasing and Critical Consciousness, it is of high relevance to keep the bigger picture in mind when addressing questions of social justice. As we have shown, De-Biasing and Critical Consciousness are interdependent topics with potentials beyond their own effects. In De-Biasing, for example, a narrow focus on F1-scores of full groups is not enough. Even a sub-group evaluation for gender would not be enough. It needs to be defined whether the artificial intelligence system is optimised for the identification of the correct or the incorrect answers. Equal F1-scores for sub-groups can have very different impacts: If one sub-group has a high precision while the other one has a high recall; one group will get very accurate feedback but not receive feedback in all cases while it is the other way around for the other group. From a STEM identity perspective, the implications are quite different: In one group, all students receive recognition – in the other group, only some students receive recognition. However, the recognition is one crucial element in STEM identity development, especially for under-served students. Hence, an important implication for research on issued of social justice in physics education is that it is crucial to consider as many details as possible as the picture may drastically change when more details are known. The level of detail needs to fit to the level where the reproduction of inequalities operates – for social justice in STEM education I consider that level to be STEM identity development. Eventually, research with explicit normative standpoints can lead to “a world where many worlds fit” (Escobar, 2017; Kayumova & Dou, 2022).

7.2.2 Practice

Interventions for social justice highly depend on the objective. In De-Biasing for example, a clear definition of the objective is a necessary point of departure. Do we focus on the students who did not yet understand a concept such as energy in order to provide meaningful support to them? Do we want to mitigate the risk of a bias in who receives most recognition, the benefit of doubt (Jeong et al., 2021)? Or do we conceptualise a miss-classification not as a benefit of doubt but as a problematic lack of support? Our objectives will strongly depend on whether we have a summative setting that serves as an evaluation for future directions or other settings. In a formative setting, it depends whether the artificial intelligence system is rather used in order to provide recognition or to provide individual support where needed. Hence, a bias definition cannot be universal over an entire domain but needs to be adjusted to the context. At the same time, bias definitions are far from irrelevant and can therefore not be arbitrarily defined. Instead, bias definitions need to be

rooted in a bigger objective and normative framework, for example the pluriverse. Addressing the need for context-specificity and a rootedness in bigger normative frameworks at the same time is a highly complex task.

Designers of artificial intelligence systems face the difficult challenge of highly complex tasks combined with under-specified guidance and no clear De-Biasing technique that scientific evidence could recommend. At the same time, not addressing matters of social justice in physics education means reproducing historically grown inequalities. Currently, evidence indicates that De-Biasing can contribute to reduce historically grown inequalities but cannot eliminate the problems of bias in their entirety. Hence, designers of artificial intelligence systems with a will to contribute to the transformation towards a pluriverse should do three things of high priority from my perspective: 1) Start De-Biasing practices by specifying whatever is needed even if no full guidance is available – some start is better than ignoring the full issue. 2) Raise awareness for the imperfections of artificial intelligence systems – users need to be aware of the risks of biases, especially when systematic De-Biasing is not even in place. 3) Demand their customers and organisations to contribute to the work on evidence-informed De-Biasing – be it in form of active work by themselves or demands towards policy-makers in order to put De-Biasing on political agendas.

7.2.3 Policy

As we have shown, inequalities in physics education often do not operate at the level of competence. Hence, under-representation of for example women in physics demands for change from a pluriversal standpoint. Such pluriversal (re-)distribution objectives in terms of equal representation might be in conflict with other objectives. In the context of artificial intelligence systems, a conflict with the goal of as much learning as possible for all students can be thought of: If a biased artificial intelligence system reaches higher competence achievements for all students but increases the differences between well-served and under-served students, what shall be done? A strict analysis of the system alone from both points of view would lead to 1) the analysis that more learning is what is wanted and that the system is good, and 2) that more inequality is not wanted and that the system is bad. The conflict can be resolved if, for example, the system is not simply put in place without further action but additional counter-measures are taken to effectively reduce inequalities. Again, these decisions are of political nature. What I want to highlight are two characteristics of these potential conflicts: 1) They can (but not necessarily do) exist on the level of a defined set of measures, and 2) Scientific evidence will come with a certain level of uncertainty as the inequalities operate at a long-term time scale in a very complex system, embedded in even bigger structures of inequalities in entire societies. Uncertainty does not mean that scientific evidence should not be used or is unable to help. However, scientific evidence for political decision making on questions of justice rather informs tendencies and indications than clear measures or options such as black or white. Finally, uncertainty does not necessarily need to lead to measures of little ambition. Instead, frequent monitoring and re-adjustments as well as a failure friendly environment can be used as well.

A decision for a normative framework always is a political one and can never be a merely scientific one. From a scientific perspective, we need to work with concrete normative frameworks in order to provide meaningful evidence. However, which normative framework is the best needs to be decided by our democratic institutions. Policy-makers are the ones who decide on the importance of a pluriverse and STEM identity development for all students in our democratic societies. If policy-makers decide for the normative

7 General Discussion

framework of the pluriverse, we have shown that systematic and structural interventions are needed in order to prevent the reproduction of historically grown inequalities. Our evidences cannot inform a political programme yet as our evidences are of small quantitative or purely qualitative nature. Hence, policy-makers with the pluriverse in STEM education as a priority should acknowledge the relevance of further research as well as the need for structural interventions to reach a world where many worlds fit.

Brise sein

*Beschwingt so durch die Gegend hüpfen,
Der Blick schweift unbesorgt umher,
Mit Flausen im Kopf über die Schulter grinsen,
Aus aller Tiefe herzlich lachen*

*So möchte ich durch's Leben zieh'n,
Dieses Gefühl in anderen entfachen,
Gemeinsam für einen Moment die Schwere überwinden –*

*Ach, welch' Kunst abzutauchen, kurz entflieh'n,
Unter den Mantel der Geborgenheit zu schlüpfen
Und von dort munter in die Welt hinaus zu linsen!*

8 References of the Dissertation Frame

- Archer, L., Calabrese Barton, A., Dawson, E., Godec, S., Mau, A., & Patel, U. (2022). Fun moments or consequential experiences? A model for conceptualising and researching equitable youth outcomes from informal STEM learning. *Cultural Studies of Science Education*, 17, 405–438. <https://doi.org/10.1007/s11422-021-10065-5>
- Archer, L., Dawson, E., DeWitt, J., Seakins, A., & Wong, B. (2015). “Science Capital”: A Conceptual, Methodological, and Empirical Argument for Extending Bourdieusian Notions of Capital Beyond the Arts. *Journal of Research in Science Teaching*, 52(7), 992–948. <https://doi.org/10.1002/tea.21227>
- 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>
- Baker, R., & Hawn, A. (2021). *Algorithmic Bias in Education*. <https://doi.org/10.1007/s40593-021-00285-9>
- Bolukbasi, T., Chang, K.-W., Zou, J. Y., Saligrama, V., & Kalai, A. T. (2016). Man is to Computer Programmer as Woman is to Homemaker? Debiasing Word Embeddings. *Advances in Neural Information Processing Systems*, 29. https://papers.nips.cc/paper_files/paper/2016/hash/a486cd07e4ac3d270571622f4f316ec5-Abstract.html
- Brickhouse, N. W. (2001). Embodying Science: A Feminist Perspective on Learning. *Journal of Research in Science Teaching*, 38(3), 282–295. [https://doi.org/10.1002/1098-2736\(200103\)38:3%3C282::AID-TEA1006%3E3.0.CO;2-0](https://doi.org/10.1002/1098-2736(200103)38:3%3C282::AID-TEA1006%3E3.0.CO;2-0)
- Brown, B. A. (2004). Discursive Identity: Assimilation into the Culture of Science and Its Implications for Minority Students. *Journal of Research in Science Teaching*, 41(8), 810–834. <https://doi.org/10.1002/tea.20228>
- Butler, J. (2005). *Giving An Account Of Oneself*. Fordham University Press.
- Carlone, H. B., & Johnson, A. (2007). Understanding the Science Experiences of Successful 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>
- 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>
- Çolakoğlu, J., Steegh, A., & Parchmann, I. (2023). Reimagining informal STEM learning opportunities to foster STEM identity development in underserved learners. *Frontiers in Education*, 8, 1–16. <https://doi.org/10.3389/feduc.2023.1082747>

- Collins, P. H. (1990). *Black Feminist Thought: Knowledge, Consciousness and the Politics of Empowerment*. <https://doi.org/10.4324/9780203900055>
- Costanza-Chock, S. (2020). *Design justice: Community-led practices to build the worlds we need*. The MIT Press.
- Crenshaw, K. (1989). Demarginalizing the Intersection of Race and Sex: A Black Feminist Critique of Antidiscrimination Doctrine, Feminist Theory and Antiracist Politics. *University of Chicago Legal Forum*, 1989(8), 139–167.
- Dennis, M., Masthoff, J., & Mellish, C. (2016). Adapting Progress Feedback and Emotional Support to Learner Personality. *International Journal of Artificial Intelligence in Education*, 26(3), 877–931. <https://doi.org/10.1007/s40593-015-0059-7>
- 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>
- Dressel, J., & Farid, H. (2018). The accuracy, fairness, and limits of predicting recidivism. *Science Advances*, 4(1), eaao5580. <https://doi.org/10.1126/sciadv.aao5580>
- Düchs, G., & Ingold, G.-L. (2018). Frauenanteil bleibt stabil. *Physik Journal*, 17(8/9), 32–37.
- Eggers, M. (2005). *Rassifizierung und kindliches Machtempfinden: Wie schwarze und weiße Kinder rassifizierte Machtdifferenz verhandeln auf der Ebene von Identität* [CAU Kiel]. https://macau.uni-kiel.de/receive/diss_mods_00002627
- Eggers, M. M., Kilomba, G., Piesche, P., & Arndt, S. (Eds.). (2023). *Mythen, Maske und Subjekte: Kritische Weißseinsforschung in Deutschland* (5., aktualisierte Auflage). Unrast.
- El-Mafaalani, A. (2021). *Mythos Bildung* (2nd ed.). Kiepenheuer & Witsch (KiWi).
- Escobar, A. (2017). *Designs for the Pluriverse: Radical Interdependence, Autonomy, and the Making of Worlds*. Duke University Press. <http://www.jstor.org/stable/j.ctv11smgs6>
- EU Charter of Fundamental Rights. (2012). EU. <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:12012P/TXT>
- Eurostat. (2022). *Tertiary education statistics*. https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Tertiary_education_statistics
- Eurostat. (2023). *Demography in Europe 2024: Take a guess!* <https://ec.europa.eu/eurostat/cache/interactive-publications/demography/2024/00/index.html>
- Fischer, J. A. (2022). *Basiskonzepte konkretisieren—Entwicklung und Evaluation einer Interventionsmaßnahme zur Förderung kumulativen Lernens durch den Einsatz von Basiskonzepten*. https://macau.uni-kiel.de/servlets/MCRFileNodeServlet/macau_derivate_00004549/Dissertation_jafischer.pdf

8 References of the Dissertation Frame

- Fletcher, R. R., Nakeshimana, A., & Olubeko, O. (2021). Addressing Fairness, Bias, and Appropriate Use of Artificial Intelligence and Machine Learning in Global Health. *Frontiers in Artificial Intelligence*, 3, 561802. <https://doi.org/10.3389/frai.2020.561802>
- Floridi, L., Cows, 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>
- Freire, P. (1970). *Pedagogy of the Oppressed*. Penguin Random House UK.
- Gago, V. (2019). *La potencia feminista. O el deseo de cambiarlo todo*. Traficantes de Sueños. https://traficantes.net/sites/default/files/pdfs/TDS_map55_La%20potencia%20feminista_web.pdf
- 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>
- Gebru, T., Morgenstern, J., Vecchione, B., Wortman Vaughan, J., Wallach, H., Daumé III, H., & Crawford, K. (2018). Datasheets for Datasets. *Proceedings of the 5 Th Workshop on Fairness, Accountability, and Transparency in Machine Learning*, 2018(80). https://www.fatml.org/media/documents/datasheets_for_datasets.pdf
- GG - Grundgesetz für die Bundesrepublik Deutschland. (2022). <https://www.gesetze-im-internet.de/gg/BJNR000010949.html>
- 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>
- Gombert, S., Di Mitri, D., Karademir, O., Kubsch, M., Kolbe, H., Tautz, S., Grimm, A., Bohm, I., Neumann, K., & Drachsler, H. (2022). Coding energy knowledge in constructed responses with explainable NLP models. *Journal of Computer Assisted Learning*, jcal.12767. <https://doi.org/10.1111/jcal.12767>
- Götschel, H. (2015). Queere Physik?! In *Sexuelle Vielfalt im Handlungsfeld Schule*. transcript verlag.
- Gunda-Werner-Institut & Center for Intersectional Justice (Eds.). (2019). *'Reach everyone on the planet...': Kimberlé Crenshaw und die Intersektionalität* (1. Auflage). gwi-boell. <https://doi.org/10.25530/03552.11>
- Ho, H.-T., Ly, D.-T., & Nguyen, L. V. (2024). Mitigating Hallucinations in Large Language Models for Educational Application. *2024 IEEE International Conference on Consumer Electronics-Asia (ICCE-Asia)*, 1–4. <https://doi.org/10.1109/ICCE-Asia63397.2024.10773965>
- hooks, bell. (1994). *Teaching to Transgress—Education as the Practice of Freedom*. Routledge. <https://doi.org/10.4324/9780203700280>

- hooks, bell. (2009). *Teaching Critical Thinking—Practical Wisdom*. Routledge. <https://doi.org/10.4324/9780203869192>
- Jemal, A. (2017). Critical Consciousness: A Critique and Critical Analysis of the Literature. *The Urban Review*, 49, 602–626. <https://doi.org/10.1007/s11256-017-0411-3>
- Jeong, H., Wu, M. D., Dasgupta, N., Médard, M., & Calmon, F. (2021). Who Gets the Benefit of the Doubt? Racial Bias in Machine Learning Algorithms Applied to Secondary School Math Education. *35th Conference on Neural Information Processing Systems*. NeurIPS 2021.
- Karademir, O., Borgards, L., Di Mitri, D., Strauß, S., Kubsch, M., Brobeil, M., Grimm, A., Gombert, S., Rummel, N., Neumann, K., & Drachsler, H. (2024). Following the Impact Chain of the LA Cockpit: An Intervention Study Investigating a Teacher Dashboard's Effect on Student Learning. *Journal of Learning Analytics*, 1–14. <https://doi.org/10.18608/jla.2024.8399>
- Kayumova, S., & Dou, R. (2022). Equity and justice in science education: Toward a pluriverse of multiple identities and onto-epistemologies. *Science Education*, 106, 1097–1117. <https://doi.org/10.1002/sce.21750>
- Kitto, K., & Knight, S. (2019). Practical ethics for building learning analytics. *British Journal of Educational Technology*, 50(6), 2855–2870. <https://doi.org/10.1111/bjet.12868>
- KMK. (2020). *Bildungsstandards im Fach Physik für die Allgemeine Hochschulreife*. Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland.
- 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
- Krajcik, J. S., & Blumenfeld, P. C. (2005). Project-Based Learning. In R. K. Sawyer (Ed.), *The Cambridge Handbook of the Learning Sciences* (1st ed., pp. 317–334). Cambridge University Press. <https://doi.org/10.1017/CBO9780511816833.020>
- Latif, E., Zhai, X., & Liu, L. (2023). AI Gender Bias, Disparities, and Fairness: Does Training Data Matter? *arXiv*. <https://doi.org/10.48550/arXiv.2312.10833>
- Li, K., & Zhang, Y. (2024). Planning First, Question Second: An LLM-Guided Method for Controllable Question Generation. *Findings of the Association for Computational Linguistics ACL 2024*, 4715–4729. <https://doi.org/10.18653/v1/2024.findings-acl.280>
- Li, L., Sha, L., Li, Y., Raković, M., Rong, J., Joksimovic, S., Neil, S., Gašević, D., & Chen, G. (2023). Moral Machines or Tyranny of the Majority? A Systematic Review on Predictive Bias in Education. *LAK23: 13th International Learning Analytics and Knowledge Conference*, 499–508. <https://doi.org/10.1145/3576050.3576119>
- Matzat, L., Zielinski, L., Cocco, M., Penner, K., Spielkamp, M., Gießler, S., Lang, S., & Thiel, V. (2019). *Atlas of Automisation—Automated decision-making and participation in Germany*. AW AlgorithmWatch gGmbH.

8 References of the Dissertation Frame

- Mecheril, P., Olalde, O. T., Melter, C., Arens, S., & Romaner, E. (2020). *Migrationsforschung als Kritik?: Konturen einer Forschungsperspektive*. <https://dx.doi.org/10.1007/978-3-531-19145-4>
- Mignolo, W. D. (2007). Delinking. The rhetoric of modernity, the logic of coloniality and the grammar of de-colonialityFootnote. *Taylor & Francis Online*, 21(2–3), 449–514. <https://doi.org/10.1080/09502380601162647>
- MNC. (2021). *Marco Nacional de Cualificaciones*. Marco Nacional de Cualificaciones. <https://www.cualificaciones.cr/mnc/>
- 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>
- Muñoz Izquierdo, C. (2012). Tres problemas fundamentales del sistema educativo. *Perfiles educativos*, 34(SPE), 154–159.
- OECD. (2016). *Excellence and equity in education* (Volume I; PISA 2015 Results). OECD.
- OECD. (2018). *Equity in Education*. OECD Publishing. <https://doi.org/10.1787/9789264073234-en>
- Ogette, T. (2019). *Exit racism* (5th ed.). unrast-Verlag.
- Paloma, P. (2023). *Labour: Vol. Cacophony* [Audio recording].
- Pardo, A., Jovanovic, J., Dawson, S., Gašević, D., & Mirriahi, N. (2019). Using learning analytics to scale the provision of personalised feedback. *British Journal of Educational Technology*, 50(1), 128–138. <https://doi.org/10.1111/bjet.12592>
- Pardo, A., & Siemens, G. (2014). Ethical and privacy principles for learning analytics. *British Journal of Educational Technology*, 45(3), 438–450. <https://doi.org/10.1111/bjet.12152>
- Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts*. (2021). European Commission. <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52021PC0206>
- Saini, A. (2017). *Inferior: How science got women wrong and the new research that's rewriting the story*. Beacon press.
- Saini, A. (2019). *Superior: The return of race science*. Beacon Press.
- Suresh, H., & Gutttag, 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>
- Tiðberger, M. (2017). *Critical Whiteness—Zur Psychologie hegemonialer Selbstreflexion an der Intersektion von Rassismus und Gender*. Springer.

- Uttamchandani, S., & Quick, J. (2022). An Introduction to Fairness, Absence of Bias, and Equity in Learning Analytics. In C. Lang, G. Siemens, & A. F. Wise (Eds.), *The Handbook of Learning Analytics* (2nd ed., pp. 205–212). SOLAR.
<https://doi.org/10.18608/hla22.020>
- 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>