Conceptions, Identity, Values and Practices

Didactic Perspectives on Computer Science

M. A., M. A. Gregor Paul Große-Bölting

Dissertation
zur Erlangung des akademischen Grades
Doktor der Ingenieurwissenschaften
(Dr.-Ing.)
der Technischen Fakultät
der Christian-Albrechts-Universität zu Kiel
eingereicht im Jahr 2023

1. Gutachter: Prof. Dr. Andreas Mühling
Christian-Albrechts-Universität zu Kiel
2. Gutachter: Prof. Dr. Carsten Schulte
Universität Paderborn

Datum der mündlichen Prüfung:
24. Januar 2024

Zusammenfassung

Die Auseinandersetzung mit Kernthemen der Informatikdidaktik, z. B. Motivation, Interesse, Retention oder Diversität, erfordert den Einsatz von Theorien oder theoretischen Konstrukten. Dies führte beispielsweise in den letzten Jahren zu einem Trend bei der Untersuchung von Identität als "theoretischer Lupe" für Diversitätsfragen. In der Forschung wird oft nicht ausgeführt, warum ein bestimmter theoretischer Rahmen für ein Problem verwendet wird und warum nicht ein ganz anderer. Darüber hinaus werden Fragen nach der Verbindung zwischen den Theorien selten gestellt. Eine genaue Untersuchung dieser Schnittstellen könnte neue Antworten liefern und neue Perspektiven auf schwierige Problembereiche wie Retention und Diversität eröffnen. Darüber hinaus trägt die genaue Definition der Theoriegrenzen zum inter- und intradisziplinären Verständnis der verschiedenen Theoriekonstrukte bei.

Diese kumulative Arbeit integriert Forschungsergebnisse aus sieben empirischen Einzelstudien, um die theoretischen Perspektiven von Informatikkonzepten, -identität, -werten und -praktiken in Bezug auf Retention und Diversität herauszuarbeiten. Insbesondere wird erörtert, welche theoretischen Möglichkeiten jede Perspektive bietet, um Herausforderungen im Zusammenhang mit Retention und Diversität zu untersuchen, und wie diese Perspektiven zusammenhängen. Da das Ziel dieser Ausarbeitung selbst theoretisch ist, soll die vorgestellte Untersuchung vor allem eine Exploration der Verbindungen und Unterschiede der verschiedenen Perspektiven sein.

Im Ergebnis kommt die Thesis zu dem Schluss, dass die Perspektiven jeweils ihre eigenen Stärken und Schwächen als theoretische Rahmung für Retention und Diversität haben. Außerdem sollten einige Fragen in diesem Zusammenhang nicht gestellt werden, ohne zwei oder mehr der theoretischen Perspektiven zu berücksichtigen. Insbesondere Praktiken sollten stärker in den Fokus der Forschung rücken, da sie die Kluft zwischen den verschiedenen Konzeptionen überbrücken können und auch für die Integration anderer wichtiger Perspektiven, wie der Wissenschaftstheorie, offen sind. Damit verbunden ist die Forderung nach einer stärkeren Ausrichtung der Forschungspraxis auf theoretische Multiperspektivität.

Abstract

Tackling core issues of computer science education, e.g. motivation, interest, retention or diversity, involves the use of theory or theoretical constructs. This led for example to a trend in the investigation of identity as a theoretical lens to questions regarding diversity in the last couple of years. Research often omits to address why a particular theoretical framework is used for a problem and why not a completely different one. Furthermore, questions about the connection between the theories are rarely asked. A close examination of these interfaces could reveal new answers, opening up new perspectives on difficult problem areas such as retention and diversity. In addition, the precise definition of theory boundaries contributes to inter- and intradisciplinary understanding of the various theory constructs.

This cumulative thesis integrates research from seven individual empirical studies to elaborate the theoretical perspectives of computer science conceptions, identity, values and practices in relation to retention and diversity. Specifically, it discusses what theoretical affordances each perspective provides to examine challenges related to retention and diversity and how the perspectives interrelate. As the aim of this elaboration is itself theoretical, the investigation presented is meant to be an exploration of the interconnections and divergences of the different perspectives.

The thesis concludes that the perspectives actually each have their own strengths and weaknesses as theoretical framing for retention and diversity. Moreover, some questions in this context should not be asked without considering two or more of the theoretical perspectives. Practices in particular should become a greater focus of research because they are able to bridge the gap between the different conceptions and are also open to the integration of other important perspectives, such as the theory of science. This is linked to a call for research practice to focus more on theoretical multi-perspectivity.

Für meine Familie

Contents

Chapter 1

Introduction

In the winter semester of 2021/22, there were more female than male students in Germany[1] for the first time, after there had already been a higher number of female first-year students since the winter semester of 2016/17 [Hüs22]. At the same time – according to data from the Federal Statistical Office analyzed by the *Centrum für Hochschulentwicklung* (CHE, Center for Higher Education Development) – student numbers have been falling steadily since a peak in the 2011/12 winter semester. In contrast, enrollment numbers in the computer science field of study (which includes the subjects of bioinformatics, computer and communications technologies, computer science, engineering computer science/technical computer science, media computer science, medical computer science and business informatics) have been rising for years. While they have declined slightly over the past two years, they remain at a high level overall (in the 2020/21 winter semester, there were 34,391, in WS 2021/22 34,560). The number of students in computer science is at an all-time high of over 250,000 [Hac22].

These positive trends – given an acute shortage of skilled workers, especially in IT [SLM+13, pp. 85][HM21] – are offset by two negative observations. The first concerns the number of graduates in computer science: In the 2020 academic year, only 18,047 students graduated from universities or universities of applied sciences in the computer science field of study. The dropout rate for universities is 40% [Hac22]. More detailed information and comparative results are provided by a study conducted by the *Deutsches Zentrum für Hochschul- und Wissenschaftsforschung* (DHZW, German Center for Higher Education and Science Research) in 2022, in which the enrollments in 2014/15 were offset against the number of graduates in 2018. According to this study, the dropout rate at universities in computer science is 44%. This is significantly higher than the average for engineering as a whole (35%), but lower than, for example, in electrical engineering (46%), mathematics (58%), and physics/geosciences (49%). The overall mean for bachelor's degrees at universities is 32% [HHS22]. It can be said that the dropout rate in computer science is at a high level, but comparable to other natural sciences and engineering. Nevertheless, there seems to be a problem with keeping students in the program and successfully leading them to a degree: a problem of retention.

---

[1] Although only figures from Germany are cited in the following, general developments in other Western countries are similar, as the discussion in the literature cited below as well as figures from the US show, e.g. [Sta20].

The second observation concerns the number of female students in computer science: Although the number of female graduates in computer science has been increasing for years, it was just 22.2% in 2021. In the winter semester of 2021/22, the percentage of female students in the field of computer science was 21.8%, and just 18.9% in computer science as a subject [Hac22]. A detailed breakdown of the numbers can be found in Tab. 1.1. There seems to be another problem: a problem with diversity.

Both issues – retention and diversity – are not really new, or surprising, as there has already been a dedicated discussion of both issues in the literature for decades [BC95; FM02; GPJ+17; BBY08]. From a societal perspective, these are major problems because, on the one hand, there is – as already noted – a shortage of skilled workers and, on the other hand, technology created by computer scientists is gaining an increasingly important place in society's life: After all, technology is everywhere, and with the continued rise of AI systems, we need not only more technically capable people overall, but also broad, society-wide representation in particular. Anything less would be undesirable in a democratic society.

The question of why retention and diversity in computer science are low is not easy to answer and is part of a larger complex. One facet is certainly an unclear image of the subject that exists both within the subject, and even more so among beginning students (this is detailled in the next section as well as in Ch. 5). Related to this are various normative notions of how computer scientists see themselves and are seen in society, within which orientation is difficult: Computer scientists are highly valued for their expertise [SLM+13, pp. 85], are seen as intelligent [BDK+23], and are well paid[2]. At the same time, there are various prejudices and stereotypes that go hand in hand with this: Computer scientists are sexist, unhygienic [MCM16], solitary [Won16] or nerds and geeks [Var07]. In this sense, it is not surprising that the subject is not particularly diverse, because it attracts only a certain clientele, while discouraging others.

It is not least a theoretical challenge to get to grips with these complicated issues because it requires looking not only at the individual itself, but also at the individual in relation to the discipline, and finally at the triangular relation of individual and discipline in interactions with other individuals both inside and outside the subject. In the end, it is these interactions that determine whether a person feels included or excluded from the subject.

1.1 Ways to Conceptualize Computer Science

To address this theoretical challenge systematically, I would like to start with an observation that will be substantiated and illustrated with examples in the remainder of this section: There are different ways of how computer science and computer scientists are to be conceptualized.

---

[2] This is reflected in data collected by developer platform Stackoverflow in a recent survey – the largest of its kind in the world, s. `https://survey.stackoverflow.co/2023/#salary-germany`.

Table 1.1. Students in the field of computer science in WS 2021/22 by subject and gender [Hac22].

Subject	**Total students**	**Male students**	**Female students**	**Percentage of women**
Computer Science	138,392	112,178	26,214	18.9%
Business Informatics	67,312	52,754	14,558	21.6%
Media Informatics	18,350	12,000	6,350	34.6%
Engineering/Technical Computer Science	16,127	13,513	2,614	16.2%
Computer and Communications Technologies	6,890	4,232	2,658	38.6%
Medical Computer Science	3,596	1,930	1,666	46.3%
Bioinformatics	2,873	1,592	1,281	44.6%
Sum	**253,540**	**198,199**	**55,341**	**21,8%**

The subject – computer science – and its members – computer scientist – are closely related in the conceptualizations: In part, computer scientists are indeed seen simply as representatives of their subject, in part, the subject is seen as the sum of all people who practice it. Accordingly, both will be considered equally in what follows. I will refer to these conceptualizations as a "perspective" because, as will be explained in detail below, I assume that each captures a particular, *valid* viewpoint on the subject and its representatives.

The theoretical discussion of computer science as a scientific discipline is one such perspective: As a scientific discipline it has been the subject of a meta-scientific discussion about its independence (e.g. from mathematics or engineering) and characteristics since its earliest days. In the context of a discussion of mathematical subdisciplines, Carl Friedrich von Weizsäcker coined the term "structural science" in the early 1970s and regarded the still young field of computer science as such: It is concerned with structures, with the ordering and abstraction of given things, but not with reality itself [Wei71]. This view may seem reasonable from a mathematical point of view, but it has been challenged many times since, because computer science is not only influenced by mathematicians, but also by electrical engineering and increasingly also by other disciplines. In this sense, the former ACM chairman Peter Denning shifted the discussion to the opposites Art vs. Science: Computer science is an art, where it is engaged e.g. with programming and the production of user interfaces, but a science where it acts as a theoretical or as an experimental science [Den05]. This approach has been further developed and elaborated in recent years [TS08], especially because empirical approaches are actually becoming more prevalent as a result of increased work with AI [Ede07].

Another perspective is that of "Weltbilder der Informatik" (i.e. "worldviews of computer science"), which was developed in a DFG study of the same name: According to this, a worldview is, following Berger [Ber13], "ein Gefüge von Wahrnehmungs-, Denk-, Bewertungs- und Handlungsmustern, das sich durch soziale Praxis entwickelt. Individuelle, soziokulturelle und objektiv lebensweltliche Einflussfaktoren bringen ein je spezifisches Weltbild hervor und wirken im Wechsel auf die Kultur." Thus, it is a complex conceptualization that assumes that students begin their studies with an individual and open worldview that, in contact with the subjects culture, receives an imprint that is not limited to knowledge: Students adopt values, habits, and dispositions that are (primarily implicitly) present in the subject. Accordingly, the DFG study uses the term to explain a wide range of phenomena, such as public perceptions of computer science, choice of studies, diversity, and ethical and social responsibility [KGH+13].

Computer Science Conceptions, sometimes also referred to as perceptions or images of computer science, can also be considered a perspective: Various researchers have investigated how especially beginners in the subject understand the subject, what are characteristics for them, and what are points of friction between their personal conception and the normative conception they are taught in the course of their studies [Hew13; KS05; BBY08]. In this regard, role models, another perspective for talking about computer scientists [GD17], serve an important function, especially when it comes to questions about diversity. Role models are exemplary individuals who embody their discipline in a particular way and thus serve as examples and mentors [Gib03]. In this way, they can appeal in particular to people who would otherwise be put off by the subject and provide a projection surface, as for example Grande et al. [GBD17] found out in a study of how far teachers can serve as role models.

Thus, while role models serve as positive identification figures, another perspective is to look at the often negative projection through stereotypes as a "process of ascribing characteristics to people on the basis of their group membership" [OHT94, p. 41]. Wong [Won16], for example, identified such characteristics in an interview study with 32 adolescents: These young people associated computer scientists with many negative aspects – male-dominated, intelligent but antisocial and solitary –, which they derive from existing stereotypes that are reproduced and reinforced by media and society. In a Draw-a-Scientist study of 305 students in grades 6 and 8, Mercier et al. [MBO06] examined perceptions of people who embodied the type of a "computer person". To do this, they had the students draw pictures of typical computer users and coded the drawings with respect to characteristics known from literature. They were able to determine that a stereotype exists: It shows "computer persons" as male and wearing glasses as well as some other negative characteristics. But overall this stereotype does not seem to prevent identification with computer science, as a subsequent interview study showed [MBO06].

Stereotypes in computer science are often discussed in relation to the terms geek, nerd, or hacker [Var07], which in turn can serve as an image of identification or deterrence. This deterrence is discussed under the term stereotype threat [MCM16], which refers to the fear among students that their abilities may be assessed based on a negative stereotype. This fear has negative consequences for girls and women in particular because it can negatively affect their performance overall, but also their sense of belonging, expectations for success, and intentions to choose a STEM career.[3]

Sense of belonging is also cited as its own perspective in the literature, e.g., in Lewis et al.'s [LBR+19] study of the relationship between sense of belonging, communal goals and their affordances, and demographics, particularly with respect to historically marginalized groups. In their empirical study of over 5,000 students from 100 institutions in the US, they found that sense of belonging and communal goals were negatively correlated and sense of belonging and communal goal affordances were positively correlated. Women, Asian, and black students in computer science have significantly lower overall sense of belonging in computer science. The concept *sense of belonging* actually comes from psychology, where it has been extensively discussed [MCM16]. Among other things, it has been shown that sense of belonging is a very good predictor of interest and motivation to pursue a STEM career [MCM16]. On the other

---

[3] This has been extensively researched for mathematics education, e.g. [NBG02; Sch02; KS07].

hand, it is unclear whether it is an independent concept or a subconstruct of identity. For example, Kong and Wang use it as a mere part of their computational identity [KW20].

Identity is another perspective that tries in a special way to focus on the person of the computer scientist and their identification with the social group and the subject. This perspective has become especially topical in recent years and in the context around issues of diversity, and is currently experiencing a trend and intensified debate [GGG+21; KG22; GGG+23]. The computer scientist identity is described under many different terms as engineering [GPH+16], computational [KW20; MHP20], maker [DM17], etc. identity.

The aforementioned examples aim to clarify the concept of perspective in relation to questions of diversity and retention. However, they do not provide an answer to the question of which perspective is the "right" one, and therefore particularly well suited to addressing the pressing issues (e.g. retention and diversity) in computing education research. In this respect, I will follow the philosophy of science considerations of the British philosopher Mary Midgley. Midgley considers various methods of describing the same set of facts across different sciences as valid approaches to understanding reality. Only by considering these diverse perspectives in conjunction can a comprehensive understanding of the subject be achieved:

> Their relation is much like that between seeing a flash of lightning and hearing the subsequent thunder. Lightning and thunder are not really separate items; they are tow partial images of a single electric discharge. Neither is an illusion; they are only both incomplete. Similarly, when we think of an activity in two different ways for different purposes, we fit it into two quite different conceptual schemes. [Mid18, p. 28]

This multiperspectivity, where each perspective is "a set of photos taken from one particular viewpoint" [Mid18, p. 95], is not to be understood in terms of relativism or defense of alternative facts; it is about alternative descriptions that can have justification in their respective contexts [Mid18, p. 44]. The necessity of these different modes of description is rooted in the complexity of the world itself, or as Midgley states elsewhere:

> That is why in this book I keep coming back to a paradox. On the one hand, I want to emphasize that there really is only one world, but also – on the other – that this world is so complex, so various that we need dozens of distinct thought-patterns to understand it. We can't reduce all these ways of thinking to any single model. Instead, we have to use all our philosophical tools to bring these distinct kinds of thought together. [Mid18, p. 193]

She also states emphatically that multiperspectivity is not an academic luxury or an intellectual extravagance, but rather ensures that intellectual life can flourish in the first place:

"Variety is not a luxury, it is not just the spice of life; it is an essential condition of survival, and this is every bit as true of the intellectual life as it is of the biological one." [Mid90] Yet these perspectives cannot simply stand side by side, they must be engaged in conversation and contextualized, otherwise there is no immediate gain in having multiple perspectives to begin with [Mid18, p. 25]. Through comparison, the mapping of an intellectual landscape, strengths and weaknesses become apparent, each depending on the contexts for which the modes of description were created. It is precisely this approach that I would like to pursue in the following, by looking more closely at four selected perspectives and exploring both their relationships to each other and their explanatory power in relation to problems of retention and diversity. My guiding principle is therefore what Midgley says about what she calls an ecological approach:

> [H]ow about considering thought ecologically, as a country to be lived in and cultivated, so that the problem is: which life-forms to encourage and which to control? This is very much the pattern that Aristotle suggested by talking about 'saving the phenomena', by which he meant looking first at the existing forms of thought to see what can be made of them before launching something quite new. You don't start your ecological work with a flame-thrower. [Mid88]

1.2 Research Questions

In the following, four perspectives will be examined in detail. The selection of exactly these four did not happen by chance, but followed an organic development during my dissertation period: Starting from the consideration of conceptions in the context of retention research, the more detailed background of which is presented in Ch. 2, it became clear that more complex questions can hardly be answered with the conceptions perspective alone. This led to a detailed examination of the identity perspective, which, as described in Ch. 3, led to two new perspectives: Values (s. Ch. 4) and practices (s. Ch. 5). However, in addition to this connection, which is grounded in my personal research trajectory, there is also a theoretical connection between the perspectives, which is detailed in Ch. 6. The purpose of addressing the perspectives is to answer the following research questions:

RQ1 What theoretical affordances does each perspective provide to examine challenges related to retention and diversity?

RQ2 How do the perspectives interrelate, and what are the points of convergence and divergence among them?

1.3 Method

The notion of perspective, to which the research questions refer, is a theoretical abstraction derived from the reflections of Mary Midgley (s. Sec. 1.1). As demonstrated in Sec. 1.1 with evidence from interdisciplinary discourse, individual perspectives are used in the research literature to provide a theoretical framework for empirical findings. However, as a theoretical framework, it does not claim to be directly transferable to the real world, and it is not to be expected that certain concepts (such as identity or values) will find immediate and observable counterparts in reality. This raises the question of the methodology with which the previously raised research questions can be answered, because empiricism alone is ruled out by these considerations.

Methodologically, the answers to the questions will be mostly theoretical-argumentative. In doing so, I will draw on the empirical findings underlying this cumulative dissertation, which are elaborated in the papers (s. Sec. 1.5). This allows for a discussion on the conceptual level, which is required by the theoretical starting point – the abstraction of perspective – but references to the real world qua empiricism are made, where possible. In other words, the answer to the research questions is given by the means of theory, because the questions are ones of theory. The qualitative and quantitative results of the research presented in the papers and individual chapters are thus in turn supplemented by the theoretical perspective.

The selection of specific perspectives (conceptions, identity, values, practices) and fields of application (retention and diversity) and their examination from a theoretical point of view is exploratory from the start and makes no claims for completeness. Instead, the aim of the research questions and the expectation to their answers is to clarify the conceptual relationships and thereby provide information about the usefulness of certain theoretical constructs in specific contexts. In contrast to an empirical approach, in which the object of investigation can be expected to be self-contained, this is not the case here. The explicit aim is therefore to examine the usefulness and relationships of different theories and concepts.

For this meta-theoretical discussion, a restriction to the application fields of retention and diversity was deliberately chosen for this thesis. This restriction would not have to be or it could be different and focus instead, for example, on issues of motivation, perception, or interest. The restriction was chosen because the problems mentioned are – as shown – systemic problems of computer science education and formed the starting point of the individual papers. Moreover, they provide a necessary focus for this work in order to anchor the argumentation on two concise aspects. However, the results are expandable and could be broadened to other aspects in the future. I will return to this point in the conclusion.

1.4 Outline

In addition to the information contained in the four published articles, this thesis primarily fulfills three additional objectives: First, a concise summary of the research results of the papers will be presented. Second, a synthesis of the individual papers will be added and enriched. The articles currently stand more or less for themselves and there is no framing discussion. This is made up for by the introduction of the research questions above. Answering these questions serves as a synthesis of the overall findings. Third, this is done by adding previously unpublished material, which is used for supplementation and deepening.

Thus, in large parts a reiteration of the literature and related works, as well as the methods, is dispensed with: These have already been considered sufficiently in the individual publications and can be consulted if necessary. Where necessary, references are given to where in the papers more in-depth information can be found. This results in the following structure for the rest of this thesis:

Chapter 2 summarizes the findings of the first two papers on Computer Science Conceptions and adds insights from previously unpublished interviews.

In **chapter 3**, the Computer Science Identity is taken into consideration: To this end, the results of the third and fourth papers are summarized. Since a synthesis of the extensive literature review on which the papers are based has not yet been done, it is presented here.

Chapter 4 presents previously unpublished (and, in a way still to be discussed, preliminary) material that I have used to explore the values of computer science from the perspective of students who are in the middle of their degree program. Since this study is not yet based on a publication, the methods used will be discussed in more detail.

Chapter 5 presents a previously unpublished, extensive interview study on computer science practices and the comparison with the expectations of first-year students regarding the activities of computer scientists. Again, since this study is not yet based on a publication, the theoretical background and methods used will be discussed in more detail.

Chapter 6 includes an overall discussion of the perspective approach and contains the answers to the research questions, while **chapter 7** discusses the possible limitations of the findings, the specific positionality of the author and its impact on the thesis.

Finally, **chapter 8** ends with a summary of the findings and contributions to the research, as well as a look at outstanding and future research.

1.5 Papers

This thesis is based on the papers listed here and referenced in the rest of the text using the following Roman numerals. The full texts can be found in Appendix A.

I - G. Große-Bölting, Y. Schneider und A. Mühling, „It's like Computers Speak a Different Language: Beginning Students' Conceptions of Computer Science," in *Proceedings of the 19th Koli Calling International Conference on Computing Education Research*, New York, NY, USA, 2019.

II - G. Große-Bölting, Y. Schneider und A. Mühling, „Beginning Students' Conceptions of Computer Science: The Effect of the First Semester," in *2020 International Conference on Learning and Teaching in Computing and Engineering (LaTICE)*, Ho-Chi-Minh-City; Vietnam, 2020.

III - G. Große-Bölting, D. Gerstenberger, L. Gildehaus, A. Mühling und C. Schulte, „Identity in K-12 Computer Education Research: A Systematic Literature Review," in *Proceedings of the 2021 ACM Conference on International Computing Education Research*, New York, NY, USA, 2021.

IV - G. Große-Bölting, D. Gerstenberger, L. Gildehaus, A. Mühling und C. Schulte, „Identity in Higher Education Computer Education Research: A Systematic Literature Review," in *Transactions on Computing Education*, New York, NY, USA, 2023.

Table 1.2 lists my personal contributions to the preparation, as well as the current publication status of the papers. I was first author of all articles used for this dissertation and therefore significantly involved in their planning and implementation.

Table 1.2. Overview of personal contributions to the preparation, as well as the current publication status of the papers.

Paper	**Conceptualization**	**Planning**	**Implementation**	**Manuscript preparation**	**Status**
I	High	Medium	High	High	Published
II	High	High	High	High	In Publication
III	High	High	Medium	High	Published
IV	High	High	Medium	High	Published

1.5.1 Other Publications

During the time of my dissertation, I contributed to the following other publications:

i - A. Mühling, C. Schulte, J. Bennedsen, L. Budde und G. Große-Bölting, „Assessing Students' Understanding of Object Structures," in *Proceedings of the 19th Koli Calling International Conference on Computing Education Research*, New York, NY, USA, 2019.

ii - G. Große-Bölting und A. Mühling, „Students Perception of the Iinner Workings of Learning Machines," in *2020 International Conference on Learning and Teaching in Computing and Engineering (LaTICE)*, Ho-Chi-Minh-City; Vietnam, 2020.

iii - G. Große-Bölting, L. Scheppach und A. Mühling, „The Place of Ethics in Computer Science Education," in *Hochschuldidaktik der Informatik - HDI 2021 - 9. Fachtagung des GI-Fachbereichs Informatik und Ausbildung/Didaktik der Informatik*, Dortmund, Germany, September 12-13, 2018, 2021.

iv - M. Schröder, G. Große-Bölting und A. Mühling, „Deriving Competency-Based Evaluation Criteria for Ethics Assignments in Computer Science," in *Proceedings of the 22nd Koli Calling International Conference on Computing Education Research*, New York, NY, USA, 2022.

v - A. Mühling und G. Große-Bölting, „Novices' conceptions of machine learning," in *Computers and Education: Artificial Intelligence*, Bd. 4, p. 100142, 2023.

Chapter 2

Computer Science Conceptions

This chapter summarizes the results from I and II (see list of publications in 1.5), supplements them with previously unpublished material, and links the topic of Computer Science Conceptions with that of Computer Science Identity.

Since the 1980s – in the tradition of Piaget's theory of genetic epistemology – there has been an increased interest in the study of conceptions held by students and how these change as the result of learning processes [Con90]. For STEM this is exemplified by the research of Smith, diSessa and Roschelle [SdR94]: The authors build on the constructivist assumption that students do not come into class as a "blank slate", but are already shaped by everyday and pre-scientific experiences. As a result, they have existing ideas about how things work and these ideas – even though often times inaccurate from a professional point of view – are usually useful tools for students. Therefore, they should be addressed in teaching not by dismissing them as mere mistakes, but by showing ways in which another conception – one that is more accurate from a professional point of view – can be applied more productively in a specific context.

In computer science there has traditionally been a particular interest in *mis*conceptions regarding programming. One example for this is Pea's description of students' "bugs" [Pea86]. The extensive research in this area has been compiled in a literature review by Qian and Lehmann [QL17]. With the advent of ubiquitous digital technology, however, research has also focused on students conceptions of digital technology in their everyday life, such as the internet [Pap05; DZ12], or the digital artifacts at the very core of computer science: computers. Based on an extensive literature review, Rücker & Pinkwart presented different conceptions that children hold about computers in a general sense [RP16].

It is therefore not surprising that conceptions of computer science have received the attention of various researchers in recent years, especially considering the importance of the first study semester(s). As a result, there is a wide range of existing empirical work in this area, which has been directly drawn upon in papers I and II, and which has in part been reproduced and extended. In this context, the work of Hewner [Hew13], Knobelsdorf and Schulte [KS05], as well as Peters [Pet18] (s. Ch. 3), has been of particular importance.

Hewner [Hew13] conducted 37 interviews with students and advisors and analyzed them with Grounded Theory. He interviewed students from three different colleges in the US throughout their studies and was able to identify three different views of the students on computer science, which he calls *Theory View*, *Programming View* and *Broad View*. The Theory View is guided by a mathematical or theoretical understanding, while the Programming View focuses on the activity of programming. The Broad View understands computer science as a science composed of different disciplines (e.g. theory, robotics, programming).

Knobelsdorf and Schulte [KS05] used a biographical method to assess perspectives on the use of computers and attitudes towards computer science. In this way, they collected self-images and world-images in addition to ideas on the field [SK07]. Their result is that, initially, there is a variety of different ideas [KS05], but these can be divided into three broad areas: those of use, professional use and design. The authors express the assumption that a transition must happen from use to design during the course of study, but this step must be taken by the learners themselves: It does not help students to give an explanation of what computer science is, students have to experience it for themselves; world-image, self-image and habits are too closely interwoven [SK07].

Furthermore, the work of Biggers et al. [BBY08] and of Kinnunen et al. [KMP13] is of interest in the context of CS coneptions and is extensively reviewed in II.

In our context, looking at and systematically analyzing beginning students' conceptions of what computer science is is only one part of a larger research project, the 2018-2021 Computer Science Cohort Survey (Kohortenbefragung Informatik, **KOI**). KOI was initiated and conducted at Kiel University in the Institute of Computer Science to find reasons for the high dropout rates of students. The long-term goal was to develop curricular measures that promote student retention. For this purpose, the project's plan was to accompany a cohort of computer science Bachelor students from their first semester in winter 2018 until the end of their studies (the ideal, according to the standard course of studies) in summer 2021; the study was discontinued in 2020 due to the Covid19 pandemic. The project included several surveys: Questionnaires were given to students, interviews were conducted as well as other research activities. An overview of the surveys in the first semester, which is particularly relevant for this study, can be found in Figure 2.1. Of all the measures depicted therein, the first questionnaire and the interviews at the beginning and end of the semester are considered for the analysis presented in I and II.

The systematic investigation of the students' conceptions was based on the observation that most of the answers to a question included in the first survey about what computer science is were only extremely vague. This impression was confirmed by a systematic analysis: On the basis of a categorization in three levels (0, 1 and 2), which was undertaken on the basis of a literature review of different definitions of computer science [SM17], it could be shown that of 310 free text answers 219 received a 0 rating, 83 a 1 rating and only 8 a 2 rating. From this it is evident that students at the beginning of their studies seem to have only a vague idea of their subject of computer science or at least one that is more or less far from the

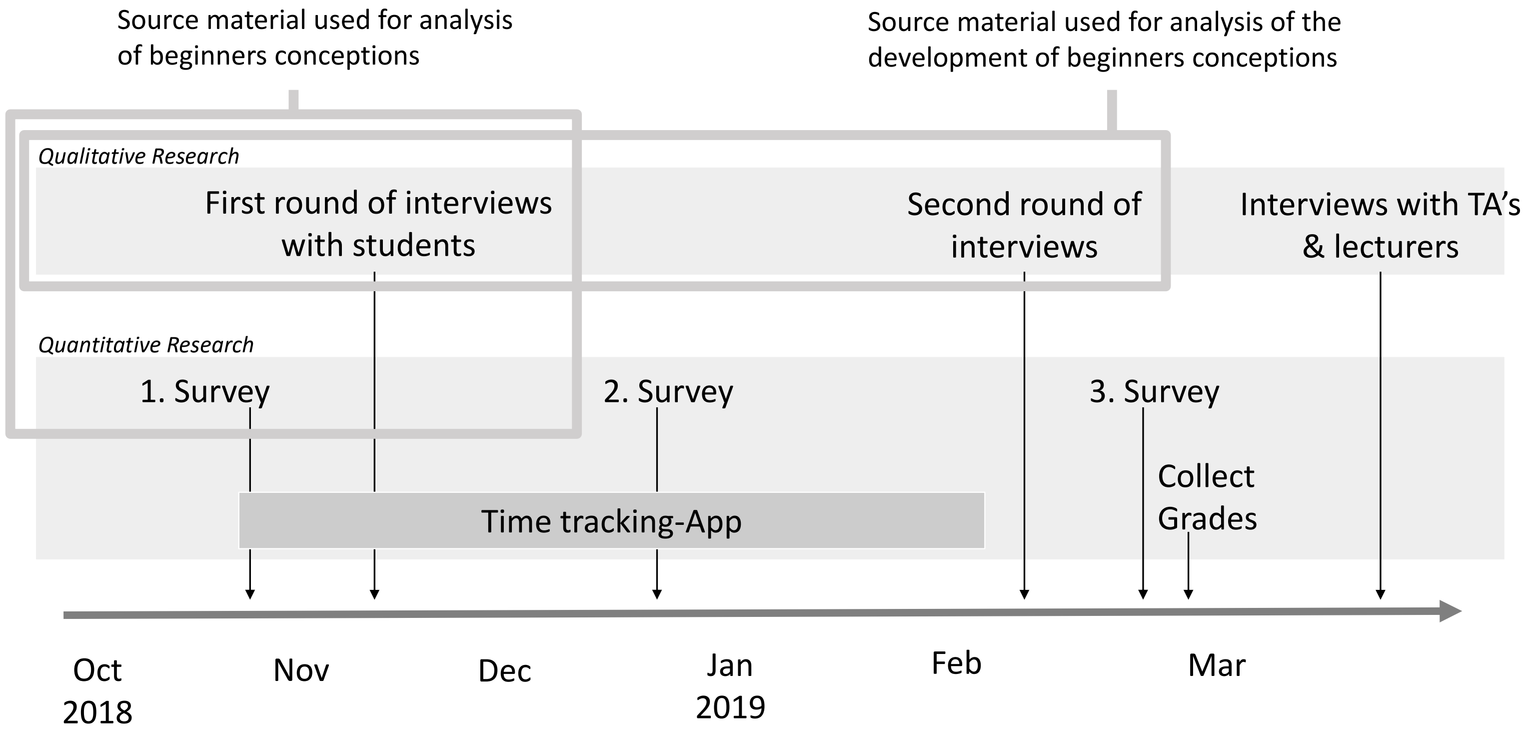


Figure 2.1. Overview of research activities in the KOI project and use for the following study.

professional definition. But what exactly do these ideas look like?

To get to the bottom of this question, interviews were conducted with 14 students. Based on the answers, a typification was created, which led to five different types: Creator, Mathematician, Differentiated picture, Interpreter, No clear picture. These are presented in the following:

The **Creator** sees computer science primarily as a means to influence the world around and to create new things. Computer science is suitable for this because the subject empowers to solve certain problems particularly well. Accordingly, programming represents the most important part of the discipline for the creator.

> Q: What is computer science for you?
>
> A: Well, definitely programming, creating programs, writing programs. Learning of one, if not several programming languages or at least the system behind a programming language. (26-15)

Algorithms and efficiency are also considered important. This conception of computer science is less scientific, CS is rather seen as engineering or as a craft. It is also recognizable that the interviewees who embody this type tend to view computer science more like a school subject (and thus also less scientific): You have to learn a certain amount of things, write exams, and in general check things off to end up leaving university as a computer scientist and finding a job in business. The openness and opportunities offered by scientific work do not play a role for the respondents.

In addition, however, we also identified a subtype, which we call **Idealists**. The idealist is a creator who pays particular attention to the way the world can be changed: They want to program and create things to make a positive difference in the world. Computer science is seen as a means to address the big problems facing humanity, as the quote shows:

> So the question of how mobility can be regulated, how to get a grip on it, how to get environmental pollution under control. I think these are all algorithmic questions that you have to solve somehow, if you want to have a reasonably good life here on earth. (15-12)

The **Mathematician** has a theoretical view of computer science that includes a special focus on mathematics. Abstractions and theoretical systems are more central to the mathematician than actual and concrete applications. Accordingly, programming and the technical, that is, the engineering side of computers, are less important. In contrast, interviewees in this group particularly often expressed an interest in solving puzzles. Thus, although the abstract aspects are the main focus, all interviewees assigned to this type expressed that what they particularly value about computer science is trying out abstract and mathematical constructs and seeing immediate results.

> At the beginning I was a bit afraid because I had never programmed anything and I don't know anything about computer games, that I can't keep up. But now that I've noticed that even there it's actually just mathematics, to think about how to get there, to write mathematically, yes, a manual like that. (22-11)

Another interviewee summarized their view of computer science by saying that it "is just applied mathematics" (22-11).

The **Interpreter** views computer science as a translation activity between human and machine:

> It's like computers speak a different language. That's how I always imagined it. Because I never understood exactly what was happening. I only saw what was happening. It's like, for example, two people talking and suddenly one of them makes a somersault and the other doesn't know why. And then I just learn the language to understand why he did the somersault. And so it was with the computers. (19-14)

For this to be possible, the computer is seen as almost intelligible, able to understand and be addressed, albeit in a different way than humans. This particular nature makes it necessary for the computer scientist to intervene; they assume a mediating and hermeneutic position.[1]

At this point it should also be pointed out that the types represent idealizations: Although individual interviewees were clearly assigned to one type, echoes of other types were also found in most of them. This is especially true for the interpreter, whose echoes can be seen through mostly two aspects also in interviewees with a different type: first, by emphasizing the interface and mediation function of computer scientists, and second, by humanizing machine behavior.

We labeled conceptions as **No clear picture**, where interviewees were not able to articulate a clear picture of their conception of computer science:

> Q: You can also ask the other way around and positively: What constitutes computer science for you?
>
> A: (thinks and laughs) Yeah. Phew, that's hard to explain, isn't it? (21-15)

Furthermore, some of the interviewees had cited a valid definition of computer science according to the criteria outlined above, but were unable to elaborate on it when asked. This is due to the fact that shortly before the questionnaire was issued, a definition was presented in lecture and it was simply replicated:

---

[1] As Michel Serres [Ser87] explains the ethymology of hermeneutics, it traces back to Hermes, the messenger of the gods in greek mythology: Hermes, too, is a mediator between – in his case – gods and humans.

> Q: Can you tell us a little bit about how you came to this definition or answer?
>
> A: Mostly because it was asked in the middle of the lesson and maybe twenty minutes before we had just done that. (14-14)

While the simple replication was easy for the students to whom this applied, they were not able to elaborate and explain the content-related aspects. Thus, they were not able to give their own concept or idea of computer science when asked.

The **Differentiated Picture** is characterized by the fact that not only one isolated aspect of computer science is regarded as central, but different ones and their interplay. Computer science is thus seen as a complex of different sub-disciplines, so that, for example, theory and practice or the view of the mathematician and the creator are understood as equally important. But even beyond that, further differentiations and sub-disciplines are possible:

> So I think it's a mixture of... well, the hardware and the software. On the one hand, we're programming now during our studies and at Computer Systems we get to know how a computer works at all. And maybe that's the connection between both sides. (26-14)

This view of computer science is particularly relevant because, it corresponds to the organization of the study program (especially in the first semester, s. Fig. 2.2) and the (normative) ideas of the lecturers, as will be discussed further below. It is also close to what Hewner [Hew13] identified as the Broad View.

At the time of the interviews, the first bachelor semester included three lectures for the students according to the curriculum: Mathematics A, Object Oriented Programming (Prog-OO), and Computer Systems (CompSys). These courses can be partially assigned to types directly found by us. In addition, there is the "intersection" between the three courses that would include the differentiated type. This means that the students in the first semester would all be confronted with different perspectives on the subject of computer science and already existing, own images would be challenged. Whether this assessment corresponds to the actual circumstances, this question was explored in a second round of interviews with the same students at the end of their first semester. Since some students had already left the program at this point, only nine people were available for this purpose, and their perceptions at the end of the semester were compared with their perceptions at the beginning.

Of the nine students for whom data were available, a change in ideas about computer science was noted in 7. A, who started out as a Creator/Idealist, has become oriented toward mathematics, their idealism playing only a minor role by the end of the semester. C, initially Mathematician, has oriented themself towards programming. Interpreter D, even at the end of the semester, does not show a completed conception of computer science, but has moved toward computer systems as the semester has progressed. E, with a more sophisticated picture,

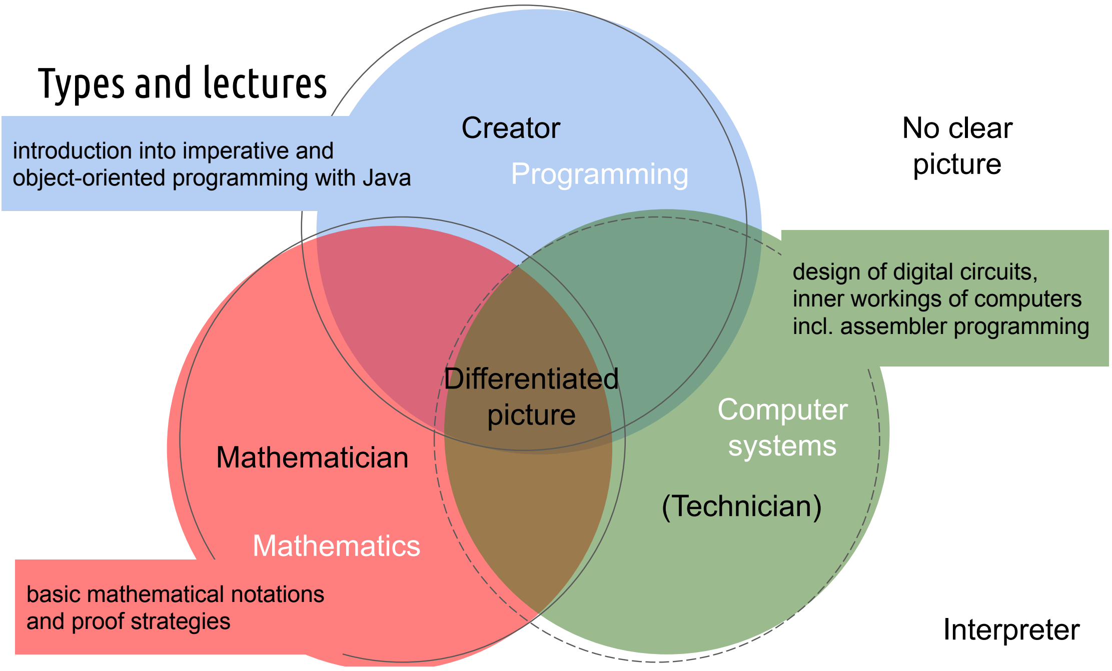


Figure 2.2. Schematic representation of the five identified types of conceptions of first semester bachelor computer science students, along with corresponding first semester lectures. The **Technician** is only a considered type and could not be shown by the interviews with students. Therefore, the type is shown with a dashed line. **Interpreter** and **No clear picture** each are placed outside the circles because they cannot be assigned to any of the usual (normative) views of computer science and are furthest from the assumed core (**Differentiated picture**) of the discipline.

retains it but better grasps the importance of mathematics to computer science. Creator F remains with their conception; it has become richer in that the information-processing nature of the discipline has been recognized and integrated into their own image of computer science by the end. G, as another representative of the interpreter type, shows a development towards computer systems, as does H, who has also integrated mathematics into their conception of computer science. For I, too, computer systems represents an important landmark for the change in their own conception of computer science at the end of the semester. A visual summary of these changes can be found in Fig. 2.3.

Even from this brief summary, it can be seen that the curriculum of the first semester of the bachelor's program in computer science has had a great impact, as could already be assumed from the literature. In this respect, the course Computer Systems seems to have been the most influential. It contains the knowledge that was least familiar to the students surveyed, and therefore had the greatest "surprise value". We had already noticed the absence of this technical perspective when working on paper I, and noted the lack of a *Technician* type that would represent systems and hardware-related computer science. This "designated gap"

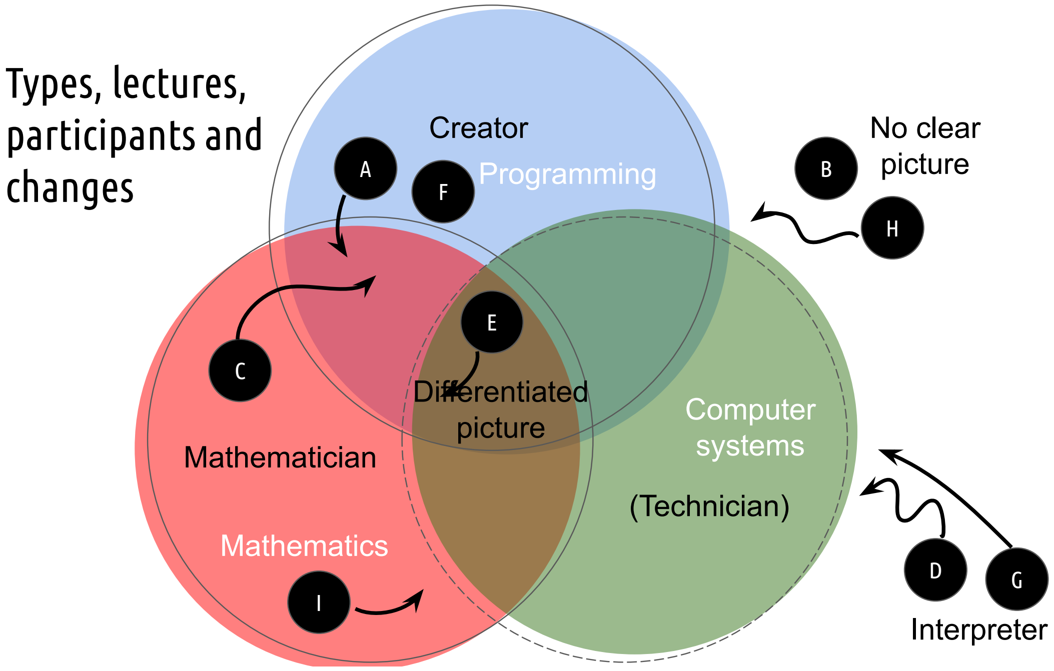


Figure 2.3. Visualization of the changed conceptions of bachelor computer science students after their first semester. Capital letters (A-I) represent the interview participants and their positioning at the beginning of their first semester with respect to the five types of conceptions we identified. The arrows indicate how the students' conceptions of CS changed during the semester. Only for two persons (B, F) no such shift could be found.

now emerges, after the first semester, as the most important point of orientation for students in changing their ideas about computer science. It is also striking that students were strongly influenced by what their own conception did not include at the beginning of the semester. For example, A starts out as a creator, but the math lecture of the first semester has a great influence on the change of his conception, while exactly the opposite is true for C: they starts out as a mathematician, but is especially influenced by the programming lecture.

In addition to the curricular influences revealed in this way, students were also asked during the interviews about "aha" moments or formative experiences, i.e., events that, in their own subjective assessment, shaped their image of computer science and of computer science studies the most during the first semester. Four groups of events emerged: (1) working with an exemplary processor in the Computer Systems lecture; (2) working on a private programming project or on the programming project in the Prog-OO lecture; (3) working on additional, voluntary tasks that were also provided in the Prog-OO lecture; and (4) upgrading one's own computer or the experience of being able to fully understand the technical terminology used in the process. An illustration of the formative experience can be seen in Fig. 2.4.

Switching perspectives from students to teachers, the third round of interviews was

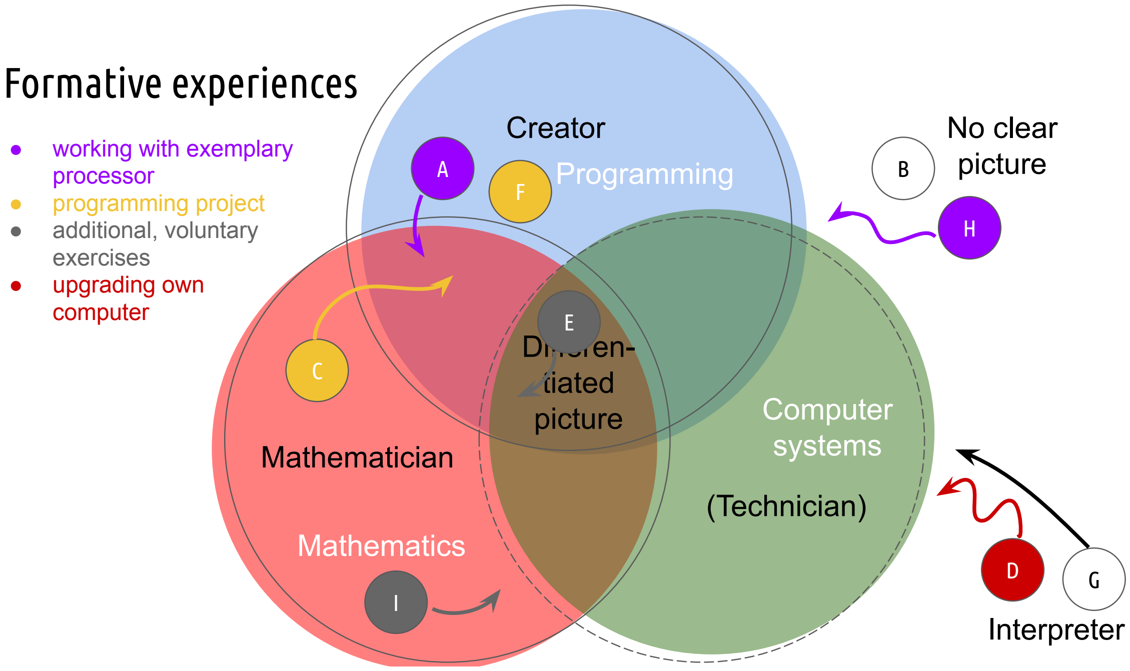


Figure 2.4. Visualization of formative experiences that have shaped students' conceptions of computer science, particularly during the first bachelor semester.

conducted after the semester with all instructors involved in the introductory courses (with the exception of mathematics, where only the professor agreed to take part in an interview but did not agree to a recording).[2] This means an internal division of the interview group into three subgroups: the professors who are responsible for the lectures and the overall conception of the course, the research assistants who are responsible for the exercise operation and teaching, and employed students who also give exercises and make corrections.

Within these three groups, the answers to the question of what computer science is vary greatly. The professors tend toward a view that corresponds to the Differentiated view. According to this view, students of computer science should have a variety of skills that they acquire in their studies and that enable them to participate in business and science. Research assistants often take a more limited view, tending to focus on their own research and work:

> Q: Okay. A completely different topic, so to speak: What do you have to be able to do when you successfully complete computer science here? What skills do you leave the university with?
>
> A: Should or actually?
>
> Q: Yes, should. And actually like to as well (laughs). I'm also interested in what

---

[2] The evaluation of this material was not previously published.

> you say then, if you're already mentioning it like this.
>
> A: (laughs) Well, it's just that the hardware is not taught here like that anymore. So it's very. . .goes more in the direction of software, I think. (. . .) And that's why I think that at least in this area you should be able to write. . .yes. . .somewhat more demanding programs. (28-10-22)

In the case of students, on the other hand, no uniform picture could be discerned: They have a Differentiated view, a view that favors a subdiscipline (like Mathematician, Technician or Creator), or no idea at all of what computer science actually is as a discipline.

In relation to the students, this seems to be mainly due to the very different backgrounds and stages within their own studies: While some students are still relatively at the beginning of their studies, others are already advanced. With regard to the employees, too, the focus on one's own main area of work and research is hardly surprising, since one's own qualifications and interests revolve around this, which must therefore have a special value in one's personal view. Among the professors, on the other hand, there is a broad picture, in line with the communities of practice theory [LW91] (s. Ch. 3 for a description), according to which the professors, as long-established teachers, constitute the center of the discipline and thus pass on its contents to the newcomers in a special way.

The interview study thus shows that the literature-based [SM17] (s. paper I) approach to determining a normative image of computer science is also justified by the fact that it can be found "in reality". The interviewed teachers of computer science, especially those who have a high degree of professionalization, hold the same image of the subject that can be found in the literature. Through the authority associated with this image, it is handed down; further development is only possible as a protracted, iterative process and only if the coming generations critically engage with the existing image, question it and develop it further. Without such a critical examination, change can hardly be expected. Again, this is understandable based on how Lave and Wenger [LW91] describe the changes in knowledge in a community of practice (CoP) (s. Ch. 3).

To situate these thoughts in a larger context, I would like to return to the interviews with the students: The assignment to a conception type (like Creator, Mathematician, etc.) was made with respect to the surveys on the basis of specific keywords. For the interviews the typification was done based on the view expressed by the interviewee as to what constitutes the subject of computer science (or the request to explain their original, short answer in more detail). The answers to this question included content, as well as specific practices associated with computer science. In addition, however, depending on the interview participant, there was also a more or less strong reflection on how computer science and computer scientists are seen by others. An example of this is interview participant F:

> [I] talked with my mother and it was about the fact that on the radio, on N-Joy[3] somehow only three different songs were playing. [. . .] And then my mother said: Yes, how about writing a program that looks at what songs are playing every day. And I was like: Hey, that's kind of a good idea. And then I just sat down and wrote a program that pulls all the songs from the N-Joy website that are on during the day. [. . .] And if I were to go a bit further and look at what, which song now runs how often per month and so on, I would have noticed: Yes, others might have written it all out by hand. Maybe how often this and that song plays per day, and I just wrote a program that does that, and for me that's what distinguishes computer science.

This interview excerpt allows for several observations: The interview participant views themself from the perspective of another person, in this case their mother. She associates certain expectations with the subject of computer science, which F – although still a beginner in the discipline – tries to fulfill. This fulfillment of expectations is linked in this case to the realization of practices associated with computer science, specifically *programming*. F's narration has the form of what Sfard and Prusak [SP05] call an *identity shaping narrative*: The narrative renders the actual identity of F at the time of the interview. However, it can also be interpreted with regard to Lave and Wenger [LW91] as a step on the way from the periphery to the center of a computer science community of practice (s. Ch. 3). What is important in all cases is that the conception of the subject alone is not enough to capture what is taking place in this interview excerpt. The conceptions describe the subject as a field of knowledge that has a greater or lesser focus on specific subfields, but social interaction and "appearing as someone" are lost over this, as are the dynamics.

Thus, it is difficult to impossible to explain the change between the two interview dates on the basis of conception theory alone. An identity theory like the CoP provides the means for this: In many cases we see a movement towards the center, the normatively conceived Differentiated Picture or Broad View, which has been discussed before.

The interviews were analyzed against the background of conceptions and without considering identity theory. Nevertheless, it seemed useful to the coders to name the types with personifications: It is the Mathematican, not just Math, the Programmer as opposed to Programming, etc. The only exceptions to this are No clear picture and the Differentiated picture as the opposite poles of a normative scale of being a computer scientist. Accordingly and on the basis of the literature reviewed by Schneider and Mühling [SM17] as well as the interviews conducted with teachers, one could say that Differentiated picture is synonymous with *computer scientist*. By this personification it is already indicated that that computer science is not just a neutral field of knowledge, but that besides knowledge also the application of

---

[3] A local broadcasting service based in northern Germany.

practices, the self-image and the world-image play a role. For this, the literature offers a further hint, as Peters [Pet17], too, draws the connection from conceptions to identity.

In the introduction (s. Sec. 1.1), it was discussed how the British philosopher Mary Midgley ascribes value to different approaches to knowledge, depending on the context in which they are used. The question to be discussed in conclusion, then, is what the context of the conceptions approach is: What are its strengths and weaknesses and how do they depend on specific usage scenarios?

By focusing on students' conceptions in I and II, it was possible to establish that students do indeed begin their studies with different ideas: There is no homogeneous picture and above all there are different perceptions, some of which deviate significantly from the normative, the Differentiated picture. Students should therefore be given the opportunity and active support, especially in the first semester, to further develop their computer science conception. This happens, as has been shown, above all through challenges[4], since they represent for example so far not considered or neglected fields of knowledge. Accordingly, of the first semester courses, CompSys was particularly influential, as hardly any of the students had any previous contact with this subject area. In addition, the lectures that were relevant for the formation and further development of one's self-image were those that did not correspond to one's own type of computer science conceptions. The design of the curriculum in the first semester thus plays a very important role in influencing the conceptions. At the same time, it was also evident that none of the students surveyed reached the Differentiated Picture in the course of the semester who did not have it from the beginning: One semester is therefore not enough to develop an elaborated conception. This means that the curriculum of the following semesters should also be designed in such a way that there is exposure to the various sub-disciplines of computer science in order to provide a broad and multi-layered picture of the subject. It is difficult to integrate the various topics into an overarching picture, and individual content can be more or less connected: Here, too, it is necessary to provide curricular and extracurricular support if one wants to promote a comprehensive integration of all knowledge into a unified picture.

As demonstrated, the conceptions approach is well suited for a quick assessment and survey. The straightforward category system of the five types of computer science conceptions allows for the easy development of an instrument to assess students and provide interventions based on their specific needs. A similar approach is taken by the persona research of Gerstenberger et al. [GWS21] and others [HMM+21; BC22]. This is especially useful for the high-throughput application at universities, where all students are expected to receive the best possible education. The retention problem can thus be countered to a limited extent: Students who in principle get along in their studies, but need specific support in one area,

---

[4] Again, this is in accordance to one kind of identity theory, that connects the individual identity development with experiences of crisis [Eri68] or critical incidents [Fla54].

can thereby experience targeted support, which may subsequently result in them staying in their studies and not be discouraged by disappointed expectations or poor performance (expectation management).

This approach cannot be expected to address or even consider larger, systemic problems. As mentioned above, it is hardly possible to explain how a change takes place between two points in time on the basis of conceptions alone. Moreover, the approach itself makes the normative foundations of the discipline the presupposition of its typification. This makes it difficult or even impossible to deal with the complex problems related to diversity. In addition, as expected, not all phenomena can be explained with regard to retention either: For example, in the interview material we identified the idealist subtype in the first round (s. paper I), a result that is similar to Peters' [Pet18] previous research. However, this subtype was not present in the second round (also similar to Peters). This is presumably due to the fact that the study of computer science provides few means to bring about the kind of social or economic change desired by idealists: in contrast, logic, programming, etc. are at the forefront of teaching. The *Sense of Belonging* literature [LBR+19; SBL+18] shows that societal and social issues are of particular interest to marginalized groups when they decide to study computer science. The fact that there is no awareness of the importance of these issues within the discipline, presumably has a chilling effect. Further research is therefore indicated.

To put it in a nutshell: Conceptions view the individual from a cognitivist or at least solitary perspective, whereas identity approaches also take into account the social, cultural and historical conditions and thus do justice to the complexity of certain questions. Which questions these are and how this is specific to computer science will be discussed in the following chapter.

Chapter 3

Computer Science Identity

This chapter summarizes the results from III and IV (see list of publications in 1.5), additionally provides an overarching view of the use of the concept of identity in computer science, and explains the transition to the following consideration of computer science values and practices.

Conceptions try to capture which internal model or image a person has of an object. The relation is mostly limited to the person and the object (world view) and does not include, for example, a level of self-reflection (“I have this image because”) or of the self (self-image). As we have seen in Ch. 2, this makes the elicitation of conceptions relatively easy; in particular, by typifying conceptions, it is possible to operationalize this view and to provide tools or interventions, such as suggestions for redesigning a curriculum. However, these constraints also make many observations impossible and significantly limit the scope of explainable phenomena: Complex problems, such as the question of retention and lack of diversity, cannot be answered without additional relations and levels of reflection. At this point, more complex explanatory models become necessary, and in this context identity has become a go-to theoretical model in recent years.

The concept of identity itself has a long history, s. Fig. 3.1. Fukuyama [Fuk18] suggests that this history begins with Plato’s Politea and the concept of *thymos*, while other authors consider Aristotle’s *phrónesis* as a kind of origin [CCS16]. While these origins continue to be referenced, there is little evidence of continuous development in educational studies. In this field, the foundations of the discussion were laid by Vygotsky [PW95] and Mead [Mea13], and to a lesser extent Piaget, whose developmental model was greatly expanded in an identity theory direction by Erikson. Erikson’s “Identity: Youth and Crisis” [Eri68], in turn, is something of a *big bang* for the psychological discussion of identity.

In the early 1990s, Lave & Wenger’s “Situated Learning” [LW91] was published, which led to an increased, socio-cultural examination of the concept of identity and subsequently inspired numerous authors to further explore the topic, but also to delimit and criticize it (e.g., by Sfard and Prusak [SP05]). Lave & Wenger state their theoretical roots in Engeströms [Eng14] activity theory and the social constructivism of Vygotsky, respectively. At the center of their theory is the *community of practice* (CoP):

> The community of practice of midwifery or tailoring involves much more than the

> technical knowledgeable skill involved in delivering babies or producing clothes. A community of practice is a set of relations among persons, activity, and world, over time and in relation with other tangential and overlapping communities of practice. A community of practice is an intrinsic condition for the existence of knowledge, not least because it provides the interpretive support necessary for making sense of its heritage. Thus, participation in the cultural practice in which any knowledge exists is an epistemological principle of learning. The social structure of this practice, its power relations, and its conditions for legitimacy define possibilities for learning (i. e., for legitimate peripheral participation). [LW91, p. 98]

Learning is accordingly a process of participation in a CoP, which for beginners starts in the periphery and leads to the center with increasing knowledge and acceptance in the community. Learning is thus not merely the acquisition of certain skills, but at the same time a social process, which in turn builds on a historically shaped, social context. For Lave & Wenger, this process is synonymous with the development of an identity:

> We have claimed that the development of identity is central to the careers of newcomers in communities of practice, and thus fundamental to the concept of legitimate peripheral participation. (...) In fact, we have argued that, from the perspective we have developed here, learning and a sense of identity are inseparable: They are aspects of the same phenomenon. [LW91, p. 115]

With respect to a complex community, such as can be assumed for a discipline such as computer science, it should also be emphasized that Lave & Wenger do not assume the necessity of a singular center and likewise, that a designated periphery could be found [LW91, pp. 36]. This makes it possible to understand disciplines as CoPs that unify certain aspects, while in others there may well be differences in the existing forms of knowledge and applied practices.

In general, identity theories can be divided into three major strands, as also shown in Figure 3.1: the psychological, the socio-cultural, and the socio-political theories, whose main representatives (in the context of educational science) are shown in the figure. Psychological theories are characterized by the fact that they conceive of identity primarily as individual attributes of a person, which often has the advantage of being observable and measurable. In contrast, socio-cultural theories view identity as shaped by a social community and culture. Accordingly, identity is based on the interrelationship between the individual and his or her social context (e.g. community), which consists not only of a set of individuals, but also of the traditions, behaviors, etc. lived by the individuals. Socio-political theories, in turn, prefer to look at power and dependency relations to describe the formation of identity. Thus, they are less common for describing educational settings.

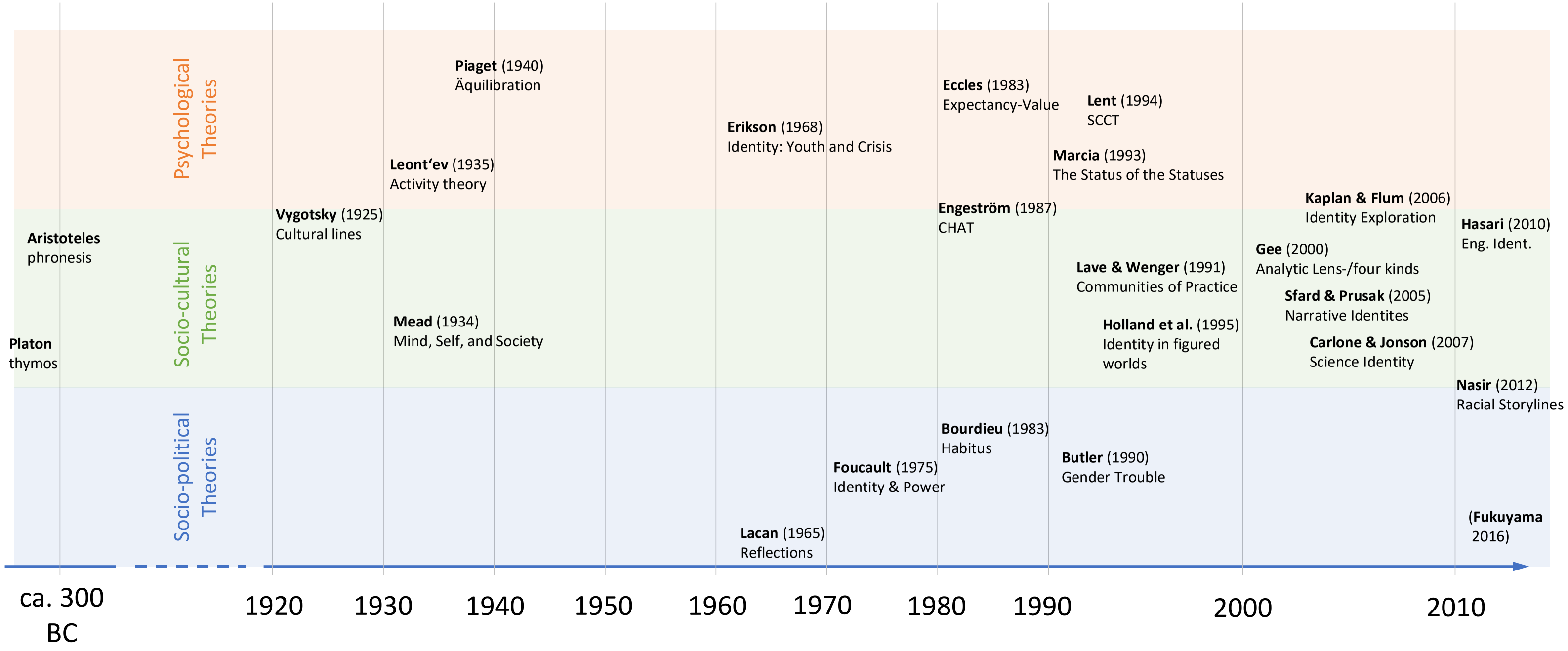


Figure 3.1. Historical overview of the development of identity theories.

Obviously, different divisions and placements could be discussed and other reviews have chosen to make other divisions. For example, Verhoeven et al. [VPV19] investigate the role of school in adolescents' identity development with a comprehensive literature review. The authors explore the theoretical perspectives by distinguishing between socio-cultural, psychosocial, social pychological, and sociological theories and combinations thereof. Darragh [Dar16] discusses various theoretical frameworks and identity definitions in the field of mathematics learners' identity. She identifies mainly socio-cultural frameworks derived from Wenger [Wen99], Holland and colleagues [HLS+01] as well as Gee [Gee00] as the largest influences on identity research from outside of the discipline. Within these theoretical frameworks, she provides the summarized definition of identity as participative, narrative, discursive, psychoanalytic or performative, bringing socio-cultural and psychological frameworks together.

Identity – unlike the conceptions discussed earlier – involves relations other than the individual and the object. First of all, this is indeed the self-reflection or the self-image and how this is shaped by the relationship to the object. Beyond that, however, depending on the theory, there are further levels: the social, cultural and historical conditionality of the object is considered in social theories; questions about power relations and political location in the framework of political theories. Further questions and emphases arise depending on the theory, i.e., depending on what is seen as the main factor influencing the shaping of an individual's identity. This results in a high degree of internal theoretical differentiation, which makes orientation within identity theories difficult [FK12].

At the same time, Computing Education Research (CER) is not known for developing theories itself [NK18], but mainly for importing them, which makes sense and is understandable due to the subject-specific circumstances. However, this import of theories can cause difficulties if theories are imported but not located and reflected upon, or if theories are seen as "drop-in" that can simply be used without being questioned (as discussed in III and IV). This creates a certain proliferation that usually requires taking a step back and questioning, sorting, and ordering the use of theories.

In recent years, identity theory has been used with increasing frequency in CER [KG22] (s. III and IV). One example is the work by Anne-Kathrin Peters [Pet17; Pet18], who investigated in a longitudinal survey how students experience participation in computer science and how they identify as computer scientists. First, 120 essays written by students enrolled in an information technology and computer science program were collected. Then, over the following three years, 61 interviews were conducted with these students regarding their choice of study, future career, study experiences, and perception of computer science as a discipline. Peters discovered seven different ways of student participation in computer science: using, learning about technology, creating, problem solving, problem solving for others, creating new knowledge and contributing to social endeavours [Pet18]. Peters situates her work in social identity theory, drawing primarily on Lave and Wenger, but also reflects extensively and exemplarily on the broader theoretical background [Pet17].

The need to create a systematic overview of identity research in CER has been addressed by Kapoor and Gardner-Mccune [KG22]. They examine 55 articles on identity in undergraduate computing and consider not only identity itself but also related concepts as inclusion criteria for their corpus. The authors note a trend in identity research over the past five years. Their study primarily investigates the thematic orientation of the articles considered and finds two main themes: identity-centered studies and non-identity-centered studies, each of which break down into subgroups. The largest subgroup (36) represents a number of studies within the identity-centered studies that attempt to explore the connection of identity to other factors. Another large group (23) of articles looks at different descriptions and conceptualizations of identity in computing. The authors conclude that more methodological and conceptual uniformity, reuse of developed tools and further development of a common vocabulary for talking about identity would be desirable.

However, as the authors themselves note:

> [T]here is a lack of work that unifies what we know about identity in CER at an undergraduate level. Our future work will describe these identity definitions and terminologies, methodologies for studying identity and synthesize the factors that influence identity formation. [KG22]

Thus, this synthesis has not yet been done by the review, but only preliminary work. This desideratum of research was addressed by III and IV (s. Sec. 1.5), which appeared in 2021 and 2023, respectively, and attempt to encompass a much larger area than just the undergraduate level, but both the K-12 field (III), as well as the entire field of higher education (IV). Moreover, unlike the study by Kapoor and Gardner-Mccune [KG22], this study does not undertake a descriptive overview, but rather a qualitative analysis of the articles on which the study is based, which is also placed within the larger context of identity theory. To this end, both a historical and a conceptual framework is developed (especially in IV) that serves as the basis for the analysis. Before providing an overview of the results, a brief methodological overview will first be provided, which will highlight the distinctive nature of systematic reviews per se and, in particular, justify the choices made for the method in this case.

A systematic literature review differs from a literature review, as found in the *Related Work* section of research papers, primarily in that it follows a clear, comprehensible, and reproducible process [Kit04]. Instead of an arbitrary and subjective selection of articles, a comprehensive review of all literature available on a topic is undertaken. The aim is completeness and the best possible fit to the topic. In addition, however, the resulting corpus is often evaluated by means that are otherwise only used in connection with empirical studies, such as qualitative content analysis according to Mayring [May14].

The basis of our literature review was a research interest in the concept of identity. This is, as explained above, multifaceted, since identity appears under different theoretical conceptualizations and labels. Accordingly, the goal of the review was to clarify the conceptualizations that can exist on different dimensions: historical, conceptual, empirical, and so on. This picture should be as complete as possible with respect to the research literature in computer science education and should take into account all relevant publications.

A number of different types of systematic literature reviews can be distinguished [XW17]: *Scoping reviews* explore a topic and attempt to capture its scope and the manifestations of its characteristics [AO05; MRP01]. *Database-driven reviews* use specific search phrases to identify the most appropriate papers in one or more databases. The quality of the results depends on the initial search phrase, but also on the selection of databases and the time resources of the reviewers. This type of review has a high recall, usually many of the relevant papers are identified. *Snowballing reviews* on the other hand start from an initial corpus of papers (for example, particularly relevant or classic papers) and track references forward and/or backward: forward by looking at papers that cite the paper at hand, backward by searching a paper's references and adding papers that also match the topic to the corpus. This process can be done in different ways, parallel, iterative, additive, etc. The success depends strongly on the quality of the initial corpus, but the papers have a high precision, i.e. they are usually very accurate [JW12; Woh14]. Another possibility is a combination of database and snowballing review, thus achieving a good balance of precision and recall. These *hybrid approaches* first form a corpus of relevant papers via a database search and extend it with the additional papers from a snowballing process. The result is a comprehensive and relevant overview of a topic [MPM+20].

According to the previously formulated goal, the hybrid approach was chosen to conduct the review. Although this is the most time-consuming, it also promises the best and most comprehensive results. In order to obtain the broadest possible overview and also to take into account the variety of different manifestations of computer science, which also appears internationally as Informatics, Computer Engineering, Information and Technology Studies, etc., *some* relation to computer science was set as an inclusion criterion: Engineering and Science papers should also be included, provided there is a disciplinary connection. Accordingly, a certain methodological complexity follows, which is reflected in a multitude of procedures for systematic literature analyses [TGA+18]. In our case, the process was as follows:

1. Setup: Formulation of research interest, selection of the process, search phrases, databases, inclusion and exclusion criteria, and arrangements for special situations, such as dealing with conflicts between coders.
2. Conducting the database search.
3. Deciding on inclusion and exclusion of papers by two coders and resolution of conflicts.

Table 3.1. Number of codings of reasons for using identity theories in the K-12 and Higher Education subcorpora, and total.

Why?	# K-12	# Higher Ed.	Total
Diversity	19	13	32
Retention	7	17	24
Recruitment	7	5	12
Better Performance	7	3	10
Motivation	5	4	9
Without, i.e. not specified	3	5	8

4. Selection of papers and formation of an initial corpus.
5. Iterative execution of the forward and backward snowballing process.
6. Division into subcorpora: K-12 and Higher Education.
7. Coding and analysis according to the research questions of interest for the subcorpora.
8. Articulation and presentation of results.

For both corpora, more or less the same research questions were investigated – with different emphases: Why is identity theory used, that is, what do the researchers hope to achieve or for which problem does this theoretical approach represent the solution? How is the theory applied? Which theories are used most frequently?

In terms of "why", a clear picture emerges, as summarized in Tab. 3.1: identity theory is seen as particularly suited to address two of CER's most pressing social problems, diversity and retention. Reasons that focus on the individual or individual learning success, such as motivation, learning success, or recruitment, take a back seat to these reasons.

Regarding the "how" of using identity theory, different category systems emerged in the coding of the different subcorpora, which are shown in Fig. 3.2. The K-12 category system somehow maps to the one in the Higher Education study: "Deepening understanding" corresponds more or less to "Theory development", while "Developing instruments" and "Evaluating interventions" each represent "Theory use". By categorizing in this way and looking at the studies from this perspective, some further observations can be made. It is, for example, noticeable that so far only a small number of instruments have been developed that make use of identity theory. Interventions and studies of interventions are available, but are often presented in a way that makes transfer to other contexts, i.e., adaptation, difficult; moreover, they mostly examine only a limited temporal and spatial frame. In contrast, there is a lack of other studies, such as Peters' [Pet18], that examine identity developments over a

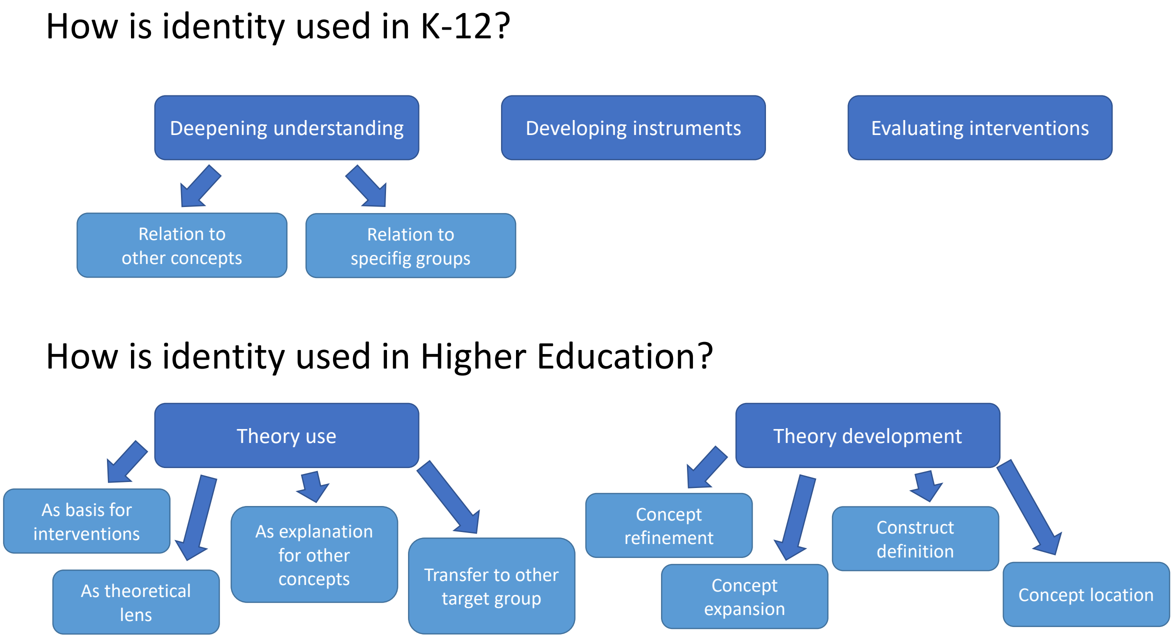
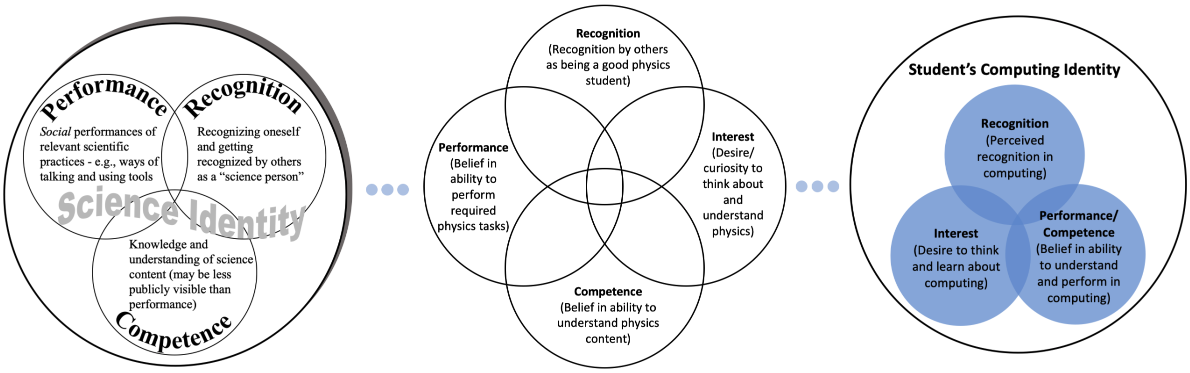


Figure 3.2. The figure shows the category systems for theory use in comparison for the two sub-studies.

longer period of time with a large number of participants. According to the codings on "why", identity is most often used as a theoretical lens to illuminate the connections of particular groups to computing and to address questions of stereotypes, role models, intersectionality, and belonging. In this regard, there seems to be a great deal of consensus regarding the usefulness of the concept of identity.

Consistent with this is the type and number of theories used: Socio-political theories play only a small role (used 3 times in the total corpus), psychological theories are used already more frequently (19). By far the most frequently used theories, however, are socio-cultural theories (43). Of these 43 uses, 22 alone are accounted for by Lave and Wenger's [LW91] Communities of Practice-theory. Other theories that are frequently used are the Carlone and Johnson [CJ07] line of theory (including Hazari et al. [HSS+10], Godwin [GPH+16], and Mahadeo et al. [MHP20]), which could be found 8 times, the psychological theory of Eccles [Ecc83] (7 times), and the Four Kinds of Identity by Gee [Gee00] (5 times). What is noticeable is the rather low diversity of theories, as well as the lack of self-developments and elaborate theory imports, which will be discussed in more detail in IV.

At this point, however, two other results will be discussed in more detail: For the Higher Education study, a closer look was taken at which (psychological) attributes are most often associated with identity and which (socio-cultural) practices. These questions arise naturally since the most commonly used theory holds that identity is held together by communities of



(Carlone & Johnson, 2007) (Hazari et al., 2010) (Mahadeo et al., 2020)

Figure 3.3. The identity theory consisting of three (to four) subconstructs, originally developed by Carlone and Johnson [CJ07], has established a lively tradition that has been extended by various researchers in engineering education, e.g. by Hazari et al. [HSS+10]. The Computing Identity model and instrument by Mahadeo et al. [MHP20] is in this line of tradition.

practice within which certain practices are pursued, can be learned, and are handed down. In terms of attributes, an overview should be derivable from the available tools and constructs.

In fact, the analysis showed interesting results in this regard: While a clear picture of the attributes could be derived from the psychological articles – which, as shown above, were much less frequent in the subcorpus – this was not possible for the practices. For the attributes, a picture emerged that can be directly connected to the Carlone and Johnson-theory line [CJ07], as seen in Fig. 3.3. This assumes – in the most recent and computer science-specific iteration of Mahadeo et al. [MHP20] – that identity breaks down into three subconstructs: performance/competence, interest in, and recognition (i.e. perception as a computer scientist by others). The systematic analysis brought to light a fourth subconstruct that is mentioned in the literature: professional responsibility as and values of computer scientists. This results in a psychological construct of identity as shown in Fig. 3.4, supplemented by the new, fourth subconstruct (Values/Responsibility).

With regard to practices, a systematic discussion could only be found in one case, in the research by Anne-Kathrin Peters [PP13; PR14; Pet18], which has already been mentioned several times. Apart from that, there are only vague references to very general practices, such as project work, problem solving or coding, which in their vagueness and superficiality leave many questions unanswered. In other words, based on our research, the picture of a computer science community of practice is offered, for which the exact practices are unknown to CER. This is more than surprising given that Lave and Wenger's [LW91] theory is – again – the most widely used.

As with the conceptions above, with respect to identity the question of the specific context of the application of identity theories naturally arises: What questions can be considered and

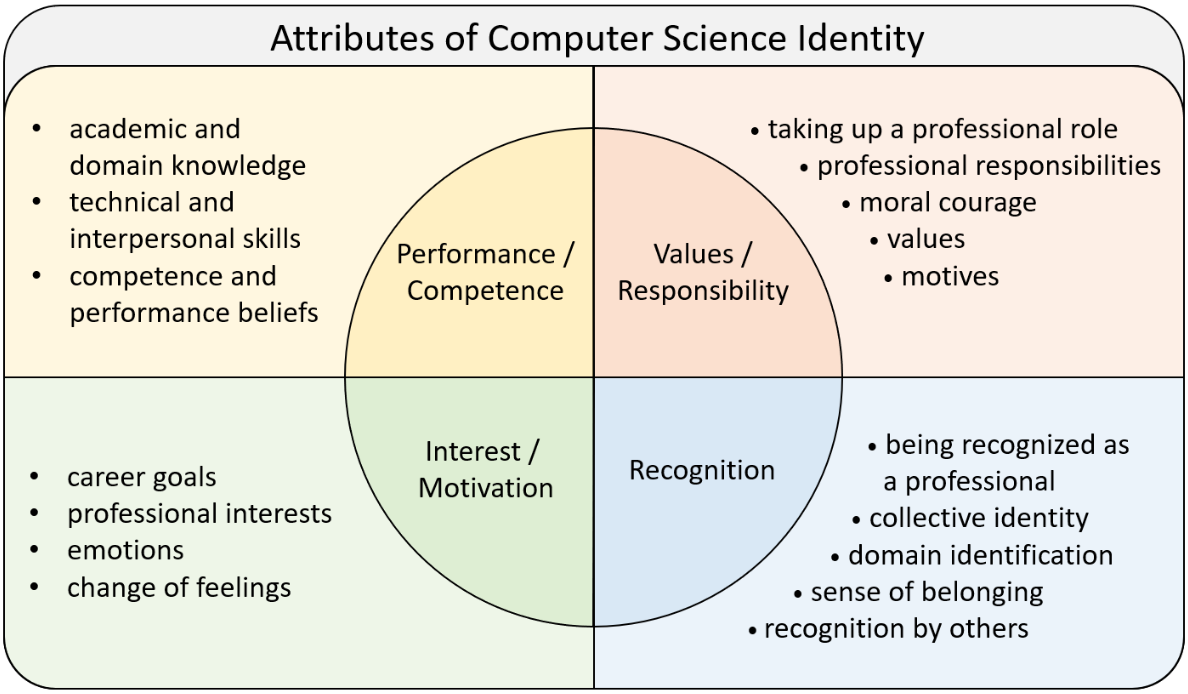


Figure 3.4. The result of the analysis of the different attributes of a computer science identity, extended by the newly found, fourth dimension Values/Responsibility.

answered particularly well through this kind of theorizing? This question has been asked by other researchers in the past. Flum and Kaplan [FK12] summarize the strength of the approach as follows:

> Developmentally, identity is an integrative concept. It may capture the objective and subjective; it commonly connects between the self and aspects of the world-out-there; it synthesizes past, present and future experiences. The process of identity formation is also anchored in a sense of 'being part of' — a web of relationships, group solidarity, and communal culture. [FK12]

However, they also note with Brubaker and Cooper [BC00] that "[i]dentity is described as a 'heavily burdened', 'elusive' and 'deeply ambiguous' term which is, nevertheless, viewed as being 'indispensable' " and cite Lichtwarck-Aschoff et al.'s [LGB+08] observation that "(v)agueness and fuzziness . . . might be important inherent qualities of the concept of identity itself". In her systematic literature review of identity in mathematics education, Darragh [Dar16] interprets this fact as a strength of the theory, which is thus highly adaptive to different circumstances and can be adjusted like a microscope to different zoom levels for observing an issue:

> Identity is a lens that is adjustable; one can zoom in to the level of interactions between individuals or zoom out to look at the wider socio-political context. We can look at the big picture, that is, at issues of mathematics learning in general. We can look at the experiences of specific groups of people and issues of equity. Or we can look at the individual level and try to understand learners' relationships with mathematics. [Dar16]

Whether this is a strength or a weakness seems to depend on the respective understanding or location, but it is certain that there is a multitude of different theories on identity development, with a high degree of internal differentiation of the concept, which places huge demands on researchers, but at the same time also promises tremendous analytical potential. This is especially true for complex problems of developmental psychology and socio-politics, such as retention and diversity, which therefore – as the literature review has shown – are particularly often the subject of these theories in CER.

There is the presumable disadvantage that research results based on (at least socio-cultural or socio-political) identity theories are much more difficult to operationalize. In addition to the observed, limited repertoire of developed instruments, this is also due to a theoretical consideration: Socio-political and socio-cultural theories are usually anti-essentialist, i.e., they assume that learning success and identity development are not determined by natural, innate factors, but are essentially shaped by social and historical circumstances. Moreover, individuals are not considered to be endowed with only one identity; rather, many different identities come together in each person and must be considered intersectionally. Both circumstances make identity a highly personal, individual phenomenon that lends itself poorly or not at all to categorization or typologization: Such a categorization would run counter to the theoretical foundations of these theories and is sometimes actively rejected, because such typologies – as discussed and presented in detail in the previous chapter – often come with normative presuppositions.

For this reason, identities are also more difficult to bring together with more abstract or normative perspectives on computer science, such as a Nature of Science or Science Theory view, because these draw on more abstract means of description that are detached from the individual and contain certain normative presuppositions qua categorization. Moreover, these perspectives generally attempt to take an "objective" view of the discipline of computer science, whereas identity theory takes the subjective experience as its starting point. We have argued – and seen other researchers agree – that identity theory is therefore well suited to dealing with particular problems. Without going into detail, we should also assume with regard to Nature of Science/Science Theory that they are valid perspectives with a specific epistemic interest that provide a vocabulary to solve their own respective problems. If we allow this assumption to hold, it presents us with a problem: How is it possible to bring identity theory and Nature of Science/Science Theory into conversation with each other?

What would be a unifying element to bring both strands of theory together?

This question is particularly important in a didactic context, where normativity cannot just be dispensed with: A discipline must communicate what its subject matter is and what its methods are. This is done through curricula and the practical activities of teachers in schools, universities, and extracurricular contexts. These practitioners of teaching must be provided with something to create a consistent picture of computer science that at the same time allows for broad identification without giving up definitional sharpness. I will argue further that this unifying principle can be the concept of practices. I will return to this in the chapter after next (s. Ch. 5) and present a study that, on the one hand, satisfies the desideratum of an overview of practices identified in the systematic literature review and, on the other hand, designs a theoretical framework of the concept of practices that is at once pragmatic and bridging between different streams of theory.

Before that, however, in the following Ch. 4 I would like to discuss another perspective that was identified in the review as a fourth subconstruct and that is influential for the constitution of communities and identities: the perspective and notion of values. In this context, values are seen partly as part of identity, and partly as a point and principle of orientation within communities (specifically, the community of computer scientists) in their own right, making a systematic discussion of values seem not only sensible, but imperative.

Chapter 4

Computer Science Values

This chapter presents new material to establish the importance of values in the identity of computer scientists and to distinguish between these values. To this end, the results of a quantitative questionnaire study and a qualitative group survey are presented.

In the following, the concept of values will be discussed in two ways: firstly, as a facet of identity and secondly, as a distinct perspective on computer science. This duality requires explanation and theoretical grounding, which I will address first in this chapter. I will then focus on the consideration of values as part of identity, which follows immediately from the results of IV's systematic review: Values are examined as a fourth subconstruct of Computing Identity according to Mahadeo et al. [MHP20], their instrument is extended and the extension is substantiated with a quantitative survey. This is followed by a qualitative investigation that takes values as an independent perspective in its own right and attempts to flesh out the concept of values through concrete computing or computer science values. Since this is a presentation of previously unpublished research data, sections 4.1 and 4.2 will more strictly follow the conventions of a research paper and will have an explicit division into a methods section, results section, and discussion section. Both studies were conducted with a relatively small number of participants (134 and 48) and in a limited context (studies at the Universities of Kiel and Paderborn, and the University of Kiel only, respectively), so that the preliminary nature of the results should be explicitly mentioned. Nevertheless, the results point to an interesting direction for further research and, together with the theoretical discussion here, can serve as an additional perspective on computer science. But let us first turn to the question of the dual nature of values mentioned above and how it relates to the concepts discussed so far.

In his politically motivated theory of identity, Appiah [App18] further defines identity in terms of three characteristics. The first is that identities come with a label [App18, p. 8]. Labels are used by others to address a member of a community or identity and are, at the same time, a means of self-assurance through self-designation. The second characteristic is that identity and the attribution of identities are meaningful to people:

> That's because of the second important thing identities share: they matter to

> people. And they matter, first because having an identity can give you a sense of how you fit into the social world. Every identity makes it possible, that is, for you to speak as one 'I' among some 'us': to belong to some 'we'. [App18, p. 9]

With the feeling of social belonging, identities also open up spaces for action: "But a further crucial aspect of what identities offer is that they give you reasons for doing things." [App18, p. 9] There are practices that are connected with a certain identity and that one lives or performs because one feels that one belongs to it. Thus, identities also have a "normative significance" [App18, p. 10], because as a member of a group, one is expected in turn to behave and act in a certain way: "It creates what you would call norms of identification: rules about how you should behave, given your identity." [App18, p. 10] What this means exactly is rarely fixed, but is involved in a continuous process of negotiation, just as belonging to an identity is subject to changing criteria.

The third characteristic is that identity not only gives the bearer of the identity reasons to act, but also, conversely, makes the bearer a target of actions: "[N]ot only does your identity give *you* reasons to do things, it can give others reasons to do things to you." [App18, p. 10] Thus, identities are also always used to justify structures of power, status, and respect.[1]

At this point, however, we are primarily interested in the second aspect of normative significance or the linking of an identity to the expectation of certain behaviors, which can be accompanied in part by sanctioning in the case of violation or gratification in the case of special fulfillment. As we have seen, this is not only meant negatively as a restriction to certain behaviors, but also opens up spaces for action and gives reasons to do things. Wenger discusses this aspect in detail in *Communities of Practice* [Wen99], his extension and systematization of the joint theory of him and Jean Lave [LW91] (which was presented in more detail at the beginning of the previous chapter 3) under the term *Meaning*.

Meaning, especially the negotiation of meaning, is for Wenger a central aspect of human activity: "Human engagement in the world is first and foremost a process of negotiating meaning." [Wen99, p. 53] How meaning is created or negotiated is linked to specific communities of practice through a dual, ongoing process of participation and reification. Participation is defined as "a process of taking part and also to the relations with others that reflect this process. [...] I will use the term participation to describe the social experience of living in the world in terms of membership in social communities and active involvement in social enterprises." [Wen99, p. 55] In contrast, reification refers to the objectification of an abstract or idea, such as the idea that democracy, economics, or justice can act as actual agents and become the object of harm or attack. Reification thus provides crystallization points for participation and negotiation of meaning by transforming abstract concepts into tangible objects

---

[1] Furthermore, Appiah notes that once identities are formed, stereotypes will follow to describe the typical members of a group: "Once identities exist, people tend to form a picture of a typical member of the group." [App18, p. 10]

that can be revisited and referenced in the world: “Whereas in participation we recognize ourselves in each other, in reification we project ourselves onto the world, and not having to recognize ourselves in those projections, we attribute to our meanings an independent existence.” [Wen99, p. 59] This objecthood is thereby not bound to an individual, but can, as a consequence of its tradification, pass over into the shared repertoire or inventory [Wen99, pp. 82] of a community. Reification is defined broadly by Wenger:

> With the term reification I mean to cover a wide range of processes that include making, designing, representing, naming, encoding, and describing, as well as perceiving, interpreting, using, reusing, decoding, and recasting. Reification occupies much of our collective energy: from entries in a journal to historical records, from poems to encyclopedias, from names to classification systems, from dolmens to space probes, from the Constitution to a signature on a credit card slip, from gourmet recipes to medical procedures, [...]. In all these cases, aspects of human experiences and practices are congealed into fixed forms and given the status of object. [Wen99, p. 59]

Participation generates reification, but at the same time it depends on already existing reifications. These can be both relatively ephemeral or a fixed part of a community in the form of the shared repertoire already mentioned. Importantly, both concepts – participation and reification – are complementary and together provide the basis for negotiating meaningfulness. A community’s shared repertoire consists of “routines, words, tools, ways of doing things, stories, gestures, symbols, genres, actions, or concepts that the community has produced or adopted in the course of its existence, and which have become part of its practice.” [Wen99, p. 83] Such a repertoire is thereby full of ambiguities that are resolved in the context of actualization through participation and are part of the negotiation of meaning. At the same time, the repertoire also provides means of enabling and constraining: enabling, because there are fixed terms, concepts and procedures that can serve as shortcuts and quick orientation; constraining, because traditional limits are set by the repertoire, which are connected to the authority of historicity (“we have always done it that way”) and therefore cannot be changed easily.

Equipped with this conceptual framework from Wenger, it is now possible to substantiate the dual nature of values introduced earlier: Values are part of the shared repertoire of a community, thus more ambiguous and often implicit; only through deliberate reflection do they become visible. In this sense, values represent a distinct perspective on computing because they are one of the most important vehicles for generating and negotiating meaningfulness. As part of the community’s repertoire, values exist independently of individuals. However, individuals must acknowledge and integrate these values into their behavior by aligning themselves with them, at least implicitly. Therein lies the identity-creating character of values, that one must identify with the values of a community or at least have an implicit sense of the existence of such values. To put it in a nutshell: Values as a reified part of a shared repertoire are a perspective in their own right; values actualized in action as part of participation within a community are part of identification.

The relationship between practices, the negotiation of meaning, and the importance of values can be illustrated by the following two examples. On the one hand, code testing is an important practice for many software teams. What should be tested, to what extent, and why is the subject of negotiation processes. The *why* is thereby frequently linked to quality requirements and/or to conceptions of thoroughness and orderliness, i.e. to shared values of the programmers. These values, in turn, influence the scope and nature of testing because they conflict with or balance other values such as efficiency and cost-effectiveness. On the other hand, value conflicts also play a role in other contexts and express their negotiation and ambiguity character. For example, teams often have to make a decision about the software to be used for a project, such as programming environment and programming languages. In the process, different values clash: While one programmer may want to try out a new, innovative programming language because they value creativity or learning, another programmer may rely on an established language and defend its use because they particularly value efficiency and reliability. Yet another programmer may advocate the use of a strongly typed language because they value its safety and efficiency (in execution), while the next may propose a dynamically typed language because they see its efficiency (in producing code) and freedom of expression as important. These different values must be balanced. Such processes are difficult, but lead to the emergence – at best – of a shared understanding of values that arise from the community, but also finds its way back again (e.g., by documenting and recording the decision). The discussion itself has an identity-forming character, because the programmers experience themselves as competent in recourse to the values and as part of a community that shares certain ideas – even if they arise from conflict.

Since values also play a role in Eccles' influential psychological identity theory, the Expectancy-Value-Theory (EVT) [Ecc83], a demarcation is necessary, because what Eccles means and what is referred to here by values has more differences than similarities. Eccles and Wigfield [EW20] describe their understanding of values within the framework of the *Task Value Construct*, in which task values are considered subjective, i.e. different tasks may be valued differently by individuals. Apart from this, there are at least four further characteristics: *intrinsic value* as the pleasure to be expected from performing a task; *attainment value* as the relative importance, based on personality or identity, ascribed to a task or action; *utitlity value* as the usefulness a task has for one's current or future plans; and *cost*, divided into effort, oppurtunity, and emotional cost, describing the effort of implementing a task [EW20]. This list alone illustrates the differences: Eccles' values is a framework for evaluating individual motivation that does not draw on a shared canon of values, but attempts to answer the question on the basis of an almost economic cost/benefit calculation: "Do I Want to do This Activity and Why?" [WC10]

In contrast, the understanding of values aimed at here is more open, broader, also has a subjective component, but in addition – in the form of reification – an intersubjective one. Values are conscious or unconscious orientations for human action. They are the normative basis for people to shape their lives. As creative principles, they give meaning and direction to human existence. They are closely related to norms, i.e. rules of action which express that a certain action is required, permitted or forbidden: For every norm there is a value which is to be concretized, i.e. realized [Pfe09]. Conversely, each value can be associated with a particular norm as a kind of rule of action. Thus, when Appiah [App18, pp. 8] speaks of the "normative significance" of actions within a community and refers to rules that standardize how individuals of a community are to behave, there are always values in the background that serve as patterns of orientation.

4.1 Identity's Fourth Subconstruct

Values, understood as part of identity, should be measurable. Since they are realized and possibly actively reflected upon by members of a community in the course of participation, there should be an awareness of the importance of values among individuals. This awareness can be raised by a self-report via a questionnaire. An idea that suggests itself is the extension of the identity construct according to Mahadeo et al. [MHP20], since the other dimensions identified in the literature are congruent and only the dimension of values and responsibility is newly added (s. Ch. 3). This idea was explored in a quantitative pilot study in which four additional questions were added to the Mahadeo et al. [MHP20] questionnaire to represent the fourth subconstruct of identity. The research context, method, and results are presented below and discussed afterwards.

Research context Data was collected in two samplings as an online questionnaire at two German universities, Kiel and Paderborn University[2] using LimeSurvey. The first round took place in the winter semester 2021/22 as part of introductory courses, the second round in the summer semester 2023 in an "Ethics in Computer Science" course that bachelor students have to take in the middle of their studies as part of their degree program. The second round took place only at the University of Kiel and immediately before the collection of the qualitative data described in the next section 4.2.

In any case, participation in the survey was voluntary, anonymous and in accordance with German and European data processing regulations. A corresponding privacy statement

---

[2] Attempts were made to get other universities to cooperate and the questionnaires were also forwarded to contacts, but no feedback or participation was received.

was provided at the beginning of the survey and had to be accepted in order to continue. The questionnaire was slightly modified between the two rounds to address the respective target groups more effectively and to implement the findings of the initial round. The original questionnaire contained the instrument of Mahadeo et al. [MHP20] as well as an adapted version of the instrument developed by Kong and Wang [KW20] (originally for use in primary education). Both questionnaires can be found with all items in Appendix B.

The questions from Mahadeo et al. [MHP20] were automatically translated into German using DeepL[3], then minor linguistic adjustments were made to make the questions sound less "wooden". The new items were developed based on the attributes identified in Ch. 3. Since there were uncertainties regarding the wording of an item related to professional responsibility, two different variants were devised for it (IDM_VAL3 and IDM_VAL4). For the evaluation by the study participants, as in Mahadeo et al. [MHP20], a 5-point Likert scale was used (do not agree at all = 1, fully agree = 5); in addition, there was the option of giving no answer.

Methods For the analysis, only questionnaires were considered that were actually completed, i.e. processed to the end. The analysis was performed with R Studio (version 2023.03.0+ for Windows) and lavaan (version 0.6.15).

Prior to the CFA presented here, an exploratory factor analysis was conducted to validate, on a test basis, the meaningfulness of the data. For this purpose, Bartlett's test for sphericity was applied to prove that the correlation matrix was not random and a threshold greater than 0.5 was set for the KMO statistic. The results of Bartlett's test for sphericity show that the correlation matrix is not random (χ^2 = 672.9551, p < 0.0001) and the KMO statistic is 0.82, which is well above the minimum standard established above for conducting an EFA and indicates a very good fit of the data. As the EFA shows promising results, several possible models were identified a priori for the CFA to review:

1. Baseline model: This model contains all items within one factor. Since this violates the assumed 4-factor structure, it can be assumed that this model does not perform well.

2. The original Mahadeo et al. [MHP20] model, that is, all items except the new IDM_VAL1 through IDM_VAL4, which represent the fourth subconstruct. This model is expected to perform very well because it represents the original construct that has already been empirically studied and validated by Mahadeo et al [MHP20].

3. A two-factor model in which the first factor includes all of the items from Mahadeo et al. [MHP20] and the second factor includes only the new value items. This model should perform poorly because good performance would imply that the subconstructs are too

---

[3] `https://www.deepl.com/`

independent of each other, i.e., that the new Values subconstruct is a foreign body to the old subconstructs.

4. A model with four factors for each of the adopted subconstructs. Since all Value items are included here, the performance should be within an acceptable range due to the intended redundancy of IDM_VAL3 and IDM_VAL4, but not optimal.
5. Models each corresponding to the aforementioned, but each alternately omitting one of the IDM_VAL items in order to test the influence of each item on the latent factor. Since IDM_VAL1 and IDM_VAL2 are not redundant, the respective absence should lead to a significant deterioration of the performance, while the models that do not contain IDM_VAL3 or IDM_VAL4 provide the best performance.

Missing data was handled using the "ml" method of the lavaan package. This is the so-called full likelihood approach, where the likelihood is calculated case by case, using all data available for that observation. For the evaluation of the CFA, the metrics mentioned by [Bro15] are used. The following thresholds were considered as evaluation parameters:

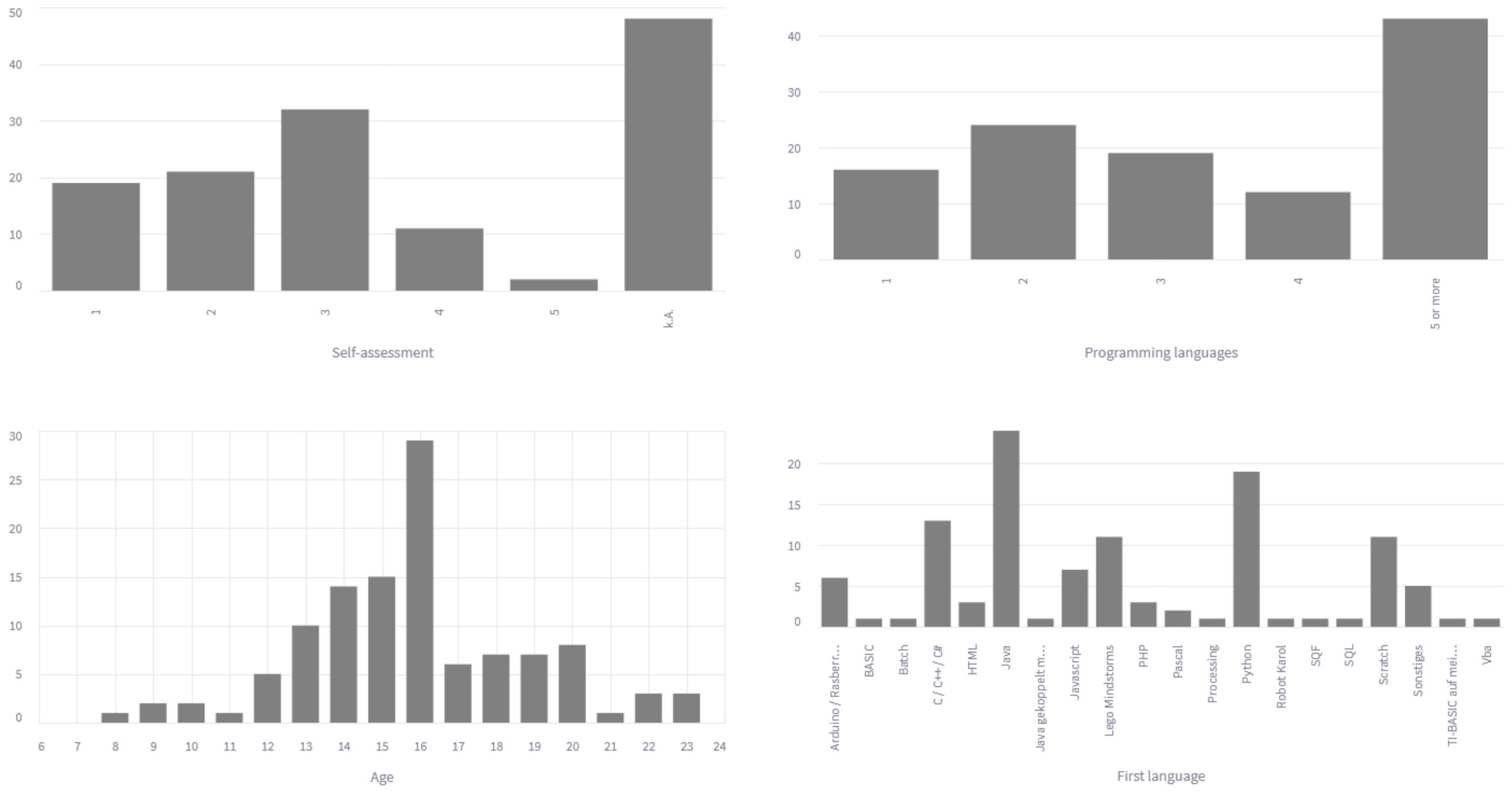
> Support for contentions of reasonably good fit between the target model and the observed data (assuming ML estimation) is obtained in instances where (1) SRMR values are close to .08 or below; (2) RMSEA values are close to .06 or below; and (3) CFI and TLI values are close to .95 or greater. [Bro15, pp. 85]

Besides, the result should be significant and the 90% confidence interval for RMSEA should be below 0.1. The Bayesian information criterion (BIC) was used as a relational metric to compare the models with each other: The model that has the lowest BIC in comparison and thus the highest goodness of fit should be selected as the final model if all other metrics are within the above parameters.

Results In total, 134 students participated in the survey, 90 in the first round, 44 in the second round. 17 of the respondents study at the University of Paderborn, the rest come from the University of Kiel. 95 of the respondents stated to be male, 33 female, the rest (2) identified themselves as diverse. Regarding the study situation, it can be noted that a large part of the students (49) were in an early stage of their studies at the time of the survey. Information about prior programming knowledge, including lines of code written and the number of programming languages known, is presented in Fig. 4.1.

Of 134 data sets, 58 have missing data, 148 of 1742 (8.5%) data points are missing in total. It is noticealbe that data is missing particularly frequently (in 57 cases) for the characteristic IDM_REC3: "My instructors/lecturers see me as a computer-savvy person". This suggests that students lack sufficient contact with lecturers to make such an assessment for themselves.

Table 4.1. Top left: Self-assessment of students programming knowledge, where 1 is low and 5 is high. **Top right:** Number of programming languages with which the respondents have already programmed themselves. **Bottom left:** Age of respondents at first programming experience. **Bottom right:** First programming language respondents had contact with.



Due to the Likert scale, which cannot be changed by the participants, there are no outliers in the data. Likewise, there are no obviously incorrectly completed entries in which, for example, all answers are the same.

Summarized results of the CFA for all models are given in Tab. 4.2. A full breakdown of model performances including factor loadings, intercepts and variances can be found in Appendix C.

Discussion In line with the expectations formulated above, the original Mahadeo model shows very good results in almost all metrics. Also as expected, the model with four factors and all value items performs moderately-acceptably. Further, removing IDM_VAL1 causes the factor analysis to not work well because there is not sufficient significance. Opposed to expectation, this is not the case for a model with IDM_VAL2 removed; on the contrary, this model performs best within the scales given above. According to the EFA, which did not result in a reliable assignment of IDM_VAL4, it would also have been expected that a model that did not include this item would perform better. This is obviously not the case; on the contrary, it is the model with the worst properties (apart from significance). Therefore, given the question development and the fact that IDM_VAL3 and IDM_VAL4 are different versions of the same question, and given the balanced or very good results, the model without IDM_VAL3 seems to be the best with the information available: It is within the parameters defined above and has the lowest BIC value. Collecting more data is necessary to make a definitive statement.

However, since this model is within the limits of the metrics under consideration, the development of the new items for the fourth subconstruct can be considered successful overall: Apparently it has been possible to demonstrate this fourth subconstruct as a latent factor in the data, at the same time independent enough to make the respective items appear as a factor of their own and dependent enough (see third model) to not give the impression of measuring something different from the first three factors. Validity appears to be given by the development against the theoretical background described in detail in paper IV and summarized in Ch. 3. For the proof of reliability the reproduction of the study with a larger sample is recommended.

Furthermore, as mentioned above, the frequent non-response to IDM_REC3 is remarkable in the data. To address this issue, the study should be repeated in a context where students are expected to have more contact with their instructors, e.g., in a master's program. At the same time, it is to be expected here that the number of study participants may not be large enough to obtain meaningful results. The result is also interesting in that it shows that students do not have enough contact with lecturers to be able to evaluate their performance from the lecturers' point of view. While this is understandable in the context of the "mass higher education", it deprives students of an important dimension in the context of their identity

Table 4.2. The table shows the main metrics used to evaluate the quality of the CFA for different model configurations tested: The *Baseline* model contains only one latent factor with all variables, the *Mahadeo* model the model originally presented by Mahadeo et al. [MHP20] with three factors and their associated items. *Mahadeo vs. Values* contains two factors, with the first modeling the original, the second modeling the new items. Finally, *VAL complete* models the new, fourth subconstruct with all four items alongside the original three subconstructs. The *without* models each leave one of the items aside.

Model	**P-value (Chi-square)**	**CFI**	**TLI**	**RMSEA**	**90 % conf. interval**	**SRMR**	**BIC**
Baseline	0.000	0.726	0.671	0.142	0.123 – 0.162	0.100	4436.725
Mahadeo Items	0.042	0.974	0.961	0.065	0.013 – 0.103	0.064	2856.592
Mahadeo vs. Values	0.000	0.799	0.755	0.123	0.103 – 0.143	0.087	4394.238
VAL complete	0.001	0.939	0.920	0.070	0.044 – 0.094	0.073	4324.606
without VAL1	0.076	0.974	0.965	0.048	0.000 – 0.079	0.065	3968.281
without VAL2	0.024	0.963	0.949	0.058	0.022 – 0.087	0.068	3984.379
without VAL3	0.014	0.959	0.943	0.062	0.029 – 0.090	0.076	3939.890
without VAL4	0.001	0.940	0.918	0.077	0.049 – 0.103	0.074	3972.439

development: feeling *perceived as someone* leads to their own identification, as described by several researchers [SP05; MHP20; HSS+10; CJ07; App18], and is an important factor. In this context, it would be interesting and a possible goal of future research to develop interventions that allow – within the framework of university operations – the possibility of closer contact between students and lecturers, that enable the basis for self-assessment through the lens of the teacher, and that investigate the influence on learning success, motivation, and identity development.

4.2 Which values?

Values, understood as conscious or unconscious orientation for human action, which as a reification form the shared repertoire of a community, can be collected as an object of qualitative research through the reflection of active members of the community. This approach was explored in a qualitative study based on a guided group discussion. The target group in this case were students at the end of their second year of study and thus advanced beginners in computer science. The starting point was the question of values that are conveyed to the students in their studies, i.e. not the values that they necessary themselves have, but those that are exemplified to them in their studies and brought to them as particularly important. The research context, method, and results are presented and subsequently discussed in the following.

Research context The study was conducted in the context of a practice session for the lecture "Ethics in Computer Science" at Kiel University in the summer term 2023 and in five different groups led by four different teachers. Participation in the practice sessions is basically voluntary, but students were specifically invited to this particular session and sweets were distributed to those present. For participants, the "Ethics in Computer Science" lecture is a mandatory part of their study program. The largest fraction of students in the summer semester are in their fourth semester of a computer science 1-subject program and thus can be considered advanced beginners within the discipline.

First, a questionnaire (s. Sec. 4.1 and Appendix B) was presented to those present to answer. The questionnaire includes general demographics, as well as questions about study situation, previous programming experience, and a computer science identity instrument (s. Appendix B). Students were given the option of answering the questionnaire. A total of 48 students answered the questionnaire, 39 of them completely.

Subsequently, the collection of values was done through a guided process in which students were asked to collect computer science values first individually, then in small groups. The exact question for this was: "What values are there (not: should there be!) in computer science? In particular, what values are you taught in your studies? Start from everyday observations (e.g. from your lectures)."

After a processing time of 10 minutes, the students were asked to enter their results on a Conceptboard[4]. In one case, further processing took place on a whiteboard; the collection was carried out by the exercise leader. In a final plenary discussion the collected values were sorted and discussed. The sorting was intended to provide an overview of related or similar values and, for example, to identify the same values identified by different smaller groups.

The concept of *value* was previously introduced in the fourth week of the lecture as conscious or unconscious orientations for human action that form the normative basis for people to shape their lives, and as design principles that give meaning and direction to human existence. To further sharpen the concept, value judgments and the form of value judgments were discussed, as well as the distinction and relationship of values to norms and interests. Also discussed were value conflicts and different types of values, such as value idealism and value naturalism. The survey took place in the eighth week of the lecture and thus at a time when the term had already been used repeatedly.

Methods The results were evaluated by qualitative content analysis according to Mayring [May14] in MAXQDA 2022. For this purpose, the concept boards were exported as an image (or a photograph in case of the whiteboard), imported into MAXQDA and coded there. Coding was initially done more or less literally, and the terms were carefully merged in a subsequent step if they were obviously similar or just spelled differently. For further categorization, the terms were classified in a way that was developed independently in two of the exercise groups, according to personal values of the computer scientist, values of the system, values of the process of creation, and social values. In addition, there were some coded terms that do not represent values or were "fun" answers (e.g. Float, Int, String as values). These were not considered for further categorization.

Results Of the 39 students who completed the questionnaire in full, 32 identify themselves as male, 6 as female; one person did not specify gender. The majority (28) of the students are – as expected – 1-subject computer science students, a smaller number (8) study two subjects or business informatics (3). The age distribution is also quite homogeneous: 29 students state their age as 21-25, nine people as 20, one person as 19.

The students were asked various questions about their programming skills to enable an assessment of their standing within the discipline. It is notable that a large proportion of 19 students did not self-assess their own programming skills. Of those who do self-assess, the distribution is such that the vast majority (10) give themselves a three (on a scale from one to

---

[4] Conceptboard is an interactive online whiteboard application, s. `https://conceptboard.com/`.

Table 4.3. Frequency of mention and distribution of values among categories developed by students in group work.

Personal Values	**Count**	**System Values**	**Count**	**Social Values**	**Count**	**Process Values**	**Count**
Responsibility	7	Privacy	7	Equality	5	Transparency	7
Order / Orga.	5	Efficiency	5	Cooperation	4	Maintainability	2
Abstract thinking	4	Correctness	3	Diversity	2	Comprehensibility	1
Critical thinking	2	Scalability	3	Compliance	2	Flexibility	1
Attention	1	Safety	3	Fairness	1	Documentation	1
Credibility	1	Performance	3	Helpfulness	1		
Performance	1	Cost-effectiveness	2				
Respect	1	Stability	2				
Self-reliance	1	Runnability	1				
Inquisitiveness	1	Interoperability	1				
Determination	1	Compatibility	1				
Reliability	1						
Sum	**26**		**31**		**15**		**12**

five, low to high), followed by an assessment of a four (7) and two or five (1 each). 28 students report having already used five or more programming languages, 6 four, and 5 three.

Codes and categories are shown in Table 4.3. The personal values category contains the most unique values (13). These deal with personal expectations of the figure of the computer scientist, such as determination or credibility, and thus also describe character traits. In contrast, values from the system values category were named most frequently (31). This includes values that are expected of systems developed by computer scientists, like programs and other artifacts. In contrast, social values include only a total of 15 mentions, distributed among seven different values. The group includes the values that should prevail in social interaction between computer scientists (and beyond). The fewest mentions altogether (13) and unique values (5) are classified in relation to the creation process: This refers to the process of creating software (or other artifacts) as a process in time, which therefore imposes certain requirements.

Looking across the groupings at the values mentioned most frequently (more than five mentions), these are: Privacy (7), Responsibility (7), Transparency (7), Efficiency (5), Order/Organization (5), and Equality (5).

Discussion The survey identified a large number of values. Some of the values are not very surprising, as they seem *somehow* self-evident for computer scientists, such as the frequent mention of efficiency, which is often considered in studies due to a focus on optimization. However, other frequent mentions of values, such as responsibility, privacy or equality, are probably related to the context of the survey in a lecture on "Ethics in Computer Science"; responsibility in particular was the explicit topic of the lecture only two weeks before the survey. At the same time, it is noticeable that the values are on different levels: There are very general values, such as efficiency or collaboration, and very specific ones, such as scalability or interoperability. On the one hand, this makes it difficult to compare the values with each other, but on the other hand, it is not surprising given the rather general question. On the contrary, it is interesting to see the wide range, which is equally complemented by different levels of detail in the values.

In particular, the grouping into the four categories shows that a strong focus of the values lies on system values and personal values of the computer scientist. Social and process values drop off significantly in their number and versatility. To some extent, this fits the common stereotype of the computer scientist as a "loner" (s. Ch. 1) who is primarily interested in things rather than people [MCG+21].[5] However, the number and variety of values mentioned overall proves that computer science is not just a technical and value-neutral science: Computer scientists are given a variety of values during their education about how systems, processes,

---

[5] Marcher et al. [MCG+21] cite this observation for male students, but as most participants in this survey where male it seems appropriate to reference their work in this context as evidence.

social interactions, and themselves should be. These formative value concepts apparently function as both inclusion and exclusion criteria [LBR+19].

Since this study presents preliminary results, it is important to also highlight some critical aspects. The categorization of values followed the suggestion made (independently) by two groups of students (who were not taught by the same lecturer); it was adopted overall for the available material in order to remain as close as possible to the student considerations. A closer look at the grouped values makes clear that the categories are not without overlap: For example, social values and process values cannot always be clearly separated, and the same applies to process and system values. Accordingly, the categorization should rather be regarded as a rough orientation, which should be the subject of differentiation and clearer separation in further studies. In addition, the selected group of students is, to some extent, a convenience sample of students who were available through participation in the lecture "Ethics in Computer Science". As already addressed above, participation may have influenced value selection. However, it is difficult to imagine repeating the same survey outside of this context, since attending this lecture gave the students a fairly precise idea of the underlying concept of values, which would be more difficult to establish with a different target group. If the study is repeated with other groups, it is essential to first establish and practice the theoretical foundations in order to be able to talk about values in a meaningful way. That this establishment worked well in the given case is also shown by the fact that the result actually consisted predominantly of values (even if – as mentioned above – there was some noise and fun answers).

The survey should be repeated with further target groups, taking into account the problems mentioned above. Conceivable target groups would be advanced beginners at the beginning of their training and professionals in order to show a progression in the development of values. Furthermore, at the moment the results stand very much for themselves, an interpretation is difficult without comparison with other subjects. Therefore, it would be interesting to repeat the same survey with nearby subjects, such as mathematics or other engineering subjects, and distant subjects, such as the humanities. Only in this way could one really state a specificity of the computer science values and their distribution, which in the above interpretation is initially based only on a pure comparison of the group distributions and mentions.

Only through further research will it be possible to talk about the normative preconditions of the subject. Nonetheless, this preliminary investigation has prepared the ground and laid the conceptual foundations for pursuing these and similar questions further.

4.3 Conclusions

This chapter had three goals: (1) to substantiate values as part of identity, (2) to take values per se seriously and to differentiate them in order to see what is meant by values of computer science in the first place, and finally (3) to theoretically substantiate both views on the value perspective. All three goals were achieved – with a preliminary character, as emphasized before.

What does this mean overall and what can be learned from it? It has been already mentioned that values play a role in retention: Students enter university with preconceived values; if these are not met, or if what they encounter in their studies even contradicts them, this can have a deterrent effect, similar to a mismatch of conceptions, and cause them to drop out. This observation has been made especially for students of historically marginalized groups, because they often take up studies with pronounced communal values expectations. Values in this case are located at the intersection of diversity and retention and thus play a prominent role in both problems.

Despite their supposed importance for discussions about identity, social cohesion and the problems of retention and diversity – which are theoretically based on the fact that identities are closely linked to social communities, each of which has a canon of values as a shared resource –, values have not yet played a major role in CER. At this point, there is potential and need for further research in this area.

Chapter 5

Computer Science Practices

This chapter reports the results of an interview study not previously published. It lays the foundation for a pragmatic notion of practices and illustrates its use through the accounts of computer science professionals.

The Czech microbiologist, physician, and science theorist Ludwig Fleck, a pronounced practitioner, reflected on his discipline and – based on this – on science and the generation of knowledge as a whole: In his view, medicine is based on an abstraction, a species concept of disease, which has a higher degree of fictionality than in other fields of knowledge [Fle19a]. For him, this results in a large gap between theory and practice within the discipline: What the medical student learns in medical school is something completely different from the activities he or she performs in the profession [Fle19a, p. 44].[1]

These activities are primarily characterized by a specific perception that allows certain phenomena to come into focus first. *Seeing* itself, Fleck states, is a complex process that is very much tied to experience and the exercise of certain practices [Fle19b]. He distinguishes this from *looking*, the mere physical, non-interpretive process of perception. Seeing in this sense is something that has to be laboriously learned and practiced [Fle19c]. Fleck describes the development of seeing in terms that immediately bring to mind the progression of an apprentice to a master in a community of practice described by Lave and Wenger [LW91] (s. Ch. 3):

> The less the education of the observing layman, or, rather, the more remote it is from the education of an expert in our field, the more different the image he sees from what the expert sees, the more remote the description.[2] [Fle19c, p. 217]

Fleck's initial starting point seems to be primarily cognitive content, which can be explained in terms of Gestalt psychology [Fle19b, p. 395]. The acquisition of these contents constitutes learning, but at the same time sets narrow limits for exchange and understanding. The science theorist characterizes this as two interconnected difficulties, which consist on the

---

[1] I will refrain from drawing explicit parallels between computer science and this view of medicine here, but I would like to point out that they exist.

[2] The original text is in German. This translation as well as the following quotations from Fleck were done by the author of this thesis.

one hand in the fact that – in order to become part of the discipline – a certain standardized training is necessary, and at the same time the communication and understanding of the necessary contents is difficult even for trained experts [Fle19c, p. 219]. Thus, in a sense, it is a matter of unconscious knowledge, which Fleck summarizes under the concept of *thought style* (*Denkstil*). Each thought style is, in turn, part of a *thought collective* (*Denkkollektiv*), for "[w]e look with our own eyes, but we see with the eyes of the collective figures whose meaning and range of permissible transpositions the collective has created." [Fle19b, p. 400] The thought style is the result of both practical and theoretical training, it is passed on from teacher to student, it represents certain traditional values which, qua tradition, are "subject to a specific historical development and specific sociological laws" [Fle19c, p. 219].

Again, the notion of thought style borders closely on the concept of communities of practice, but goes beyond in scope and reach: Fleck, as a scientific theorist, sees the epistemological implications of his conceptualization. Thus, thought style and collective define *what can be known*. The thought style is shaped by a thought collective, but its transmission is partly *tacit*, through the transmission of and in practices [Fle19d]. Practices thus play a formative role in the acquisition of knowledge in a certain domain. Fleck illustrates this by the reading of microscopy images that are misinterpreted by lay people: They look but cannot see; they are not able to identify specific patterns. Only an expert trained on many images can identify interesting observations and deviations from the norm, which can then be used for further knowledge acquisition, classification, and interpretation [Fle19d].

If practices are understood in this way, then they play a central role in understanding a discipline: they define who can see themselves as *such a person* (i.e. member of the discipline), since practices are passed on in teacher-student relationships and are thus exemplified by important role models. As long as they are not *esoteric* practices, but practices that are known to outsiders (without the necessity of being able to practice them oneself), they allow identification as a member of a thought collective or community of practice, for other members as well as for outsiders. And they shape what can be known at all, which gives them an important epistemological function [Fle19e]. The fact that they are handed down and are socially anchored has also been recognized as an important criterion in socio-cultural theories of identity, which often include a play on the notion of practices [Wen99].

Nevertheless, as shown in Ch. 3, it is unclear which practices are part of computer science. This desideratum will be tackled in the following, through the description of an interview study dedicated to the elaboration of practices in computer science. Since these are again previously unpublished results, the description follows the classical structure of research papers, in which first the theoretical background is presented and in particular a pragmatic concept of practices is developed. This is followed by a description of the method used, results, and a discussion of the findings. The chapter concludes with a discussion of the practices perspective.

5.1 A Theory of Practice

In recent years, the study of practices has seen growing intensity in science education, as well as in some subject didactics and human computer interaction (HCI). Examples of this are provided by chemistry didactics (e.g. [BWJ+06; CS18]), where practices have been the focus of increased reflection both as part of teaching and as part of educational research. In HCI, Kuuti and Bannon [KB14] have presented a programmatic paper entitled "The Turn to Practices in HCI: Towards a Research Agenda" to further develop research in this area.

This trend also manifests itself in an increased meta-theoretical engagement with practices and the concept of practices. Miettinen et al. [MSY09] divide the concept of practices into three main strands of theoretical engagement: cultural-historical activity, exemplified by Engeström et al. [EMP99], socio-cultural approaches, as pursued by Chaiklin and Lave [CL96], Lave and Wenger [LW91] and Wenger [Wen99], and the pragmatists' theory of action, exemplified by Joas [Joa93]. The first two understand practices as a "creative activity mediated by tools and cultural artifacts and as a process in which humans simultaneously create both themselves and their material culture." [MSY09] Lave and Wenger's [LW91] theories have already been discussed in the context of identity (s. Ch. 3), but since practices are a central aspect here, we will return to this below in a discussion of Wenger's [Wen99] further development of the theory.

Practices, as understood primarily in the first two major strands of theory, are all about context, including the social and historical circumstances, but also our everyday, lived experiences, the material environment, artifacts, affinities and motivation: "Practices are wholes, whose existence is dependent on the temporal interconnection of all these elements, and cannot be reduced to, or explained by, any one single element." [KB14] A focus on practice forces researchers to look at different levels of analysis and to change the zoom level between micro (personal level), meso (routines shared by practitioners) and macro (institutional and social contexts) [MSY09]. For this reason, Lave bitingly states that "'decontextualized learning activity' is a contradiction" [Lav09]. This has serious implications for research, as it clearly complicates the study of practices when the particular context of the practices must be considered. Therefore, Nicolini states that "studying practices through surveys or interviews alone is unacceptable." [Nic12, pp. 218]

This situation seems aporetic: How can practices be researched when two of the most common research methods are insufficient to do so? Lave and Nicolini's concept of practice certainly has its justification, but at the same time it makes such high demands that it is hardly suitable in practice. In the following, I will therefore try to develop a more pragmatic concept of practice. To do so, it would be helpful to take a look at the uses of the concept practice as they occur in CER. As discussed in Chapter 3, however, little (at least in the context of identity) can be found on this. One exception is the work of Peters, discussed in paper IV.

However, the practices she identifies are quite general in nature and at such a high level of abstraction that it makes concrete application difficult.

Another exception is the work of Bergstrom and Blackwell [BB16], who have identified a set of practices of programming. The theoretical background they have chosen for this draws on Kuuti and Bannon [KB14], where practices are understood as socio-culturally situated, as “routines consisting of interconnected and inseparable elements: physical and mental activities of human bodies, the material environment, artifacts and their use, contexts, human capabilities, affinities and motivation.” [KB14] In this very sense, as quoted above, they are wholes that result from the contextual interconnection of all these elements. Methodologically, however, Bergstrom and Blackwell do not proceed empirically, but on the basis of a (non-systematic) literature review. This allows them to identify six different practices: software engineering, bricolage & tinkering, sketching with code, live coding, hacking, and code bending. Completeness is not the authors’ goal but rather to demonstrate a diversity of programming practices. Following the identification, the authors discuss some further observations related to these practices, some of which are also of interest for further consideration here: The authors thus assume that practices often do not occur (or are performed) alone, but in combination. Therefore, it is difficult to establish overlap-free theoretical reflection and definition (“[P]ractices need not be mutually exclusive.”), which is evident in the authors’ discussion of whether debugging is a practice. In addition, Bergstrom and Blackwell [BB16] emphasize the importance of considering different practices in relation to the teaching and didactics of computer science, because presenting the range of practices of programming ensures that not only future programmers, but a broad range of people can be addressed. Software engineering is not the only *correct* way to program. What is striking about the authors’ inventory of practices, however, is that very little internal differentiation in practices exists or is even discussed. The practice that the authors see as the most central – software engineering – and that probably most often represents programming, is, according to its definition, very general and the degree of abstraction is obviously on a completely different level than, for example, the very narrowly defined code bending.

In addition to the marginal discussion of practices in CSER, there are some neighboring concepts that deserve attention in order to sharpen the conceptual boundaries of the practice concept. For example, Zendler et al. [ZSK08] have identified the most important processes in computer science. They understand processes as “scientific methods of individual scientific disciplines which are performed in order to gain knowledge (e.g., observing and experimenting in biology; deducing and inducing in mathematics; observing and interviewing in psychology)” [ZSK08]. They draw on Parker and Rubin: “There are a variety of processes through which knowledge is created. There are also processes for utilizing knowledge and for communicating it. Processes are involved in arriving at decisions, in evaluating consequences, and in accommodating new insights” [PR66, p. 2]. Processes are understood by Zendler et al.

[ZSK08] as constructivist processes of knowledge acquisition and are explicitly examined as a counter-model to Schwill's "fundamental ideas" [Sch93]. The authors criticize the lack of reliable empiricism, as well as the missing connection to the methods of the discipline, as the "ideas" are purely content-oriented. To identify the processes of computer science, the study presented 128 experts (from Germany only) with a list of 44 general processes from the literature for evaluation. 24 experts made such an assessment. The results were clustered, and the clusters yielded the processes most relevant to computer science. According to the authors, the results are statistically stable and valid. The most important cluster contains the following processes: problem solving and problem posing, classifying, finding relationships, investigating, analyzing, and generalizing; another cluster identified as important by the authors contains: communicating, presenting, collaborating, questioning, ordering, and comparing. Thus, even the processes identified in this way are somewhat general and not necessarily specific to computer science [ZSK08]. Processes share with practices the practical approach, which distances itself from a purely content-focused examination of a discipline. Practices focus on social, while processes focus on cognitive aspects. Zendler et al.'s [ZSK08] criticism of "fundamental ideas" as insufficiently empirically grounded can also be understood in this context as indicating that normative ideas enjoy a somewhat undeservedly high status here.

Another, neighboring concept to practices is the concept of competences. The modern interpretation of competences as cognitive abilities and skills tied to solving specific problems, as well as the necessary motivation, volition, and social readiness to apply the ability/skill successfully and responsibly in variable situations [KH08] is mostly centered around an individual person and removes in part the social and historical context as well as the material conditions that are otherwise central to the concept of practices. The historical concept of competence according to McClelland [McC73], on the other hand, goes very much in a direction reminiscent of Wenger's [Wen99] discussion of the concept and is particularly evident in the policeman example:

> If you want to test who will be a good policeman, go find out what a policeman does. Follow him around, make a list of his activities, and sample from that list in screening applicants. Some of the job sampling will have to be based on theory as well as practice. [McC73]

My concept of practices is based on Wenger [Wen99] and McClelland [McC73]: In Wenger's sense, it is something that is socially and historically anchored, taking place in a community of like-minded people. Practices are actions that (can) include social, historical and material resources, even one's own body – they are "embodied". Different "maturities" of mastery of practices can be identified within a CoP [Wen99, pp. 45]: Newcomers who are still on the periphery perform a practice differently from the old-timers in the center of the community. For me, this is also a definitional property, since different skill levels can be identified for practices. Moreover, they are repetitive activities; they are activities that draw on material and ideational shared resources – the shared repertoire – of the community. Practices are normative (s. the discussion of Appiah [App18] in Ch. 4) and, as norming and normed, are subject to change and metapractices (i.e. discussions or arguments).

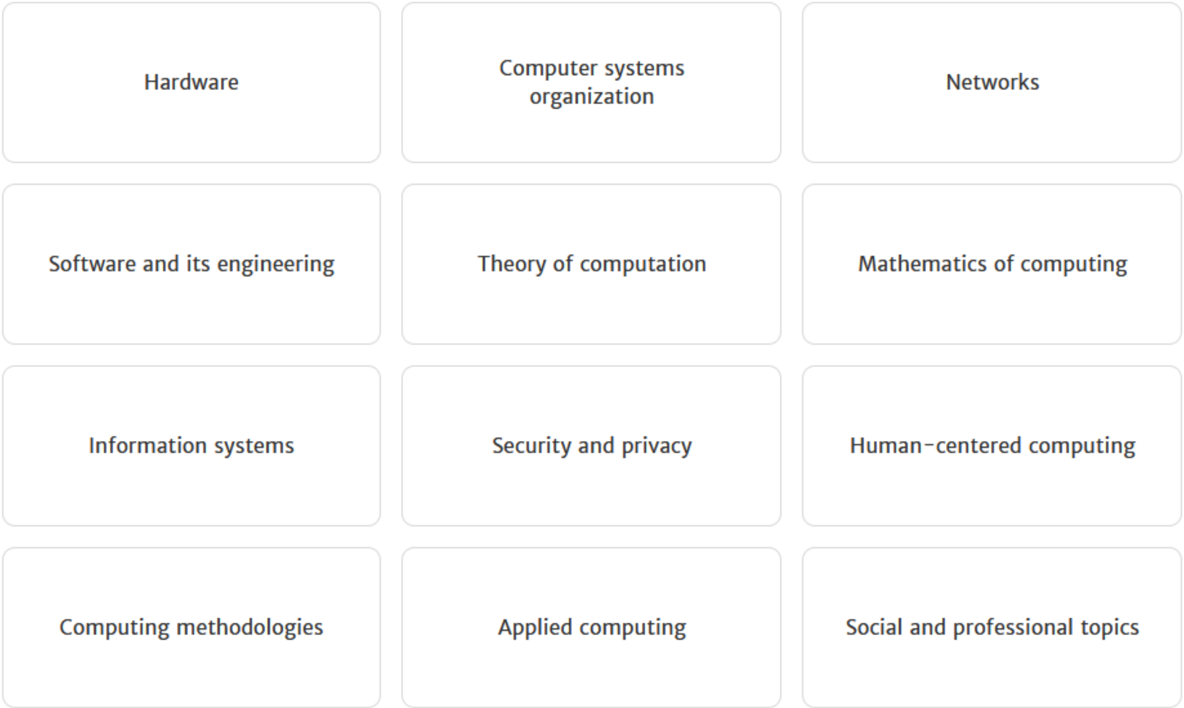
This notion is particularly inspired by McClelland [McC73] in that the observational aspect plays a significant role: We can follow or ask someone what they consider an activity of their working environment and infer (if we ask various people) what practices constitute the community. Practices have an objective or at least intersubjective character, because they are part of the social inventory of a community and as such are reified or have reifying consequences. Thus, language practices develop that refer to practices or even name them as practices.

An example of this is "meetings", since talking to another person is not a practice to begin with. Exchanging ideas with one's office neighbor on a regular basis about a particular aspect of computer science is a practice, because that involves a shared knowledge base that can be more or less (deeply) mastered. Meetings are a practice if they are held recurrently for a specific reason (sprint meeting, product planning) and use methods that can exist in varying degrees of sophistication (again: sprint meeting, but also agenda creation, etc.). This refers back to the shared resources of a community, but at the same time contributes to it by creating named processes and workflows. This is illustrated by another example: Testing software is a practice, because it is a repetitive and socially anchored activity that draws on and produces methods (defined procedures such as Test Driven Development) and other resources (such as libraries and frameworks that facilitate the writing and execution of tests), and can be mastered and observed to varying degrees.

Taken together, the definition of practice used in the following is: Practices are repetitive, socially anchored activities that draw on the shared inventory of a community, but also performatively confirm and extend it. They tend to have varying degrees to which they can be mastered (and thus implicitly include some power dynamic because there is a difference between teachers or those who know and learners or those who do not know). They are also observable because recourse to a shared inventory involves fixed processes, the use of tools, or language practices.

5.2 Interview Study Method

Sampling The goal of the interviews is to obtain an overview of common practices in computer science that, while not expected to be exhaustive, should have a satisfactory breadth and, in that respect, be exemplary as a pattern for adding further practices in subsequent surveys. On the one hand, this breadth is to be understood in disciplinary terms: All sub-disciplines of computer science should be covered or represented. On the other hand,



Hardware	Computer systems organization	Networks
Software and its engineering	Theory of computation	Mathematics of computing
Information systems	Security and privacy	Human-centered computing
Computing methodologies	Applied computing	Social and professional topics

Figure 5.1. All first order subcategories of the ACM Computing Classification System (ACM CCS).

care should be taken to interview individuals of different genders, who have had different educational backgrounds, and who work in business, as well as in academia and the public sector.

It can be assumed that the practices are in a certain sense quite homogeneous, since they exist in the context of a globally active research and working community (see, for example, the open source movement); therefore, a geographical restriction to northern Germany or even Schleswig-Holstein seems unproblematic for the interviews. The views of the subject of computer science represented in the literature of the Nature of Science or *Computer Science* Science Theory are mostly limited to a very rough and abstract subdivision of the discipline into subdisciplines (e.g. [TS08; Ede07]); moreover, areas, especially in the field of social issues or HCI, are often not considered. Thus, these divisions cannot serve as a basis for recruiting interview participants.

Instead, the ACM Computing Classification System (ACM CCS) is used, more specifically the highest order concepts contained therein, s. Fig. 5.1. The aim of the sampling is to interview at least two representatives from each of the mentioned domains to practices (the domain "General and reference" is disregarded). Since the CCS is used to classify all publications of the ACM, this should result in a broad picture, after all, the CCS pursues the goal of being able to classify everything that is publishable in the field of computer science. The interview partners should be chosen in such a way that there is a relative balance in terms of gender and activity in business/science. Overall, then, a mixture of *criterion-i* and *snowball* [PHG+13] was used for sampling: First, criteria were established to identify interview participants, namely fit with a subcategory of the CCS. After an initial selection of participants, they were asked about further contacts, and so on.

Since the ACM sees itself as a representative organization of both business and science, no tendency to distort the results is to be expected in this respect. However, it is still a matter of a certain (American, Western, etc.), possibly even normative point of view on computer science. This positionality, as well as the origin of the interview participants, has to be critically considered in the evaluation of the results and in the context of the limitations.

Interviews The interviews were conducted partly in person and partly online via the BigBlueButton software, and the conversations were recorded. The actual interview was preceded by a pre-talk in which a rough, but not too detailed overview of the research objective, as well as a briefing on data protection and the rights of the interviewee were given. The audio recording was also followed by a brief debriefing in which the interviewee was asked for impressions and feedback on the interview.

The actual interview was semi-structured. Although there was a prepared list of questions, this was expanded by spontaneous additional and follow-up questions as the interview progressed. The questions asked in each case were:

1. When you are asked what you do for work, what do you answer?
2. If you were to give this activity a label, what would be your answer?
3. Would you say that you identify as a computer scientist? (intended closed question)
4. (Inquiry) Why/why not? As what else?
5. If I were to accompany you in your everyday work: What activities could I observe that you do (again and again) and that you consider to be firmly part of your <identification>?
6. Is there anything else on the subject that we haven't addressed?

Overall, the list of questions was designed to make the interviews as time-efficient as possible: Since the respondents were professionals and potentially very busy, it was an explicit goal to limit the entire interview, including pre- and post-interview, to about 30 minutes.

Coding Mayring's [May14] content analysis was used to code the data. An inductive coding method was chosen to allow for an open and exploratory analysis of the interview data and to ensure that the results were based on a thorough data evaluation. This approach allows for the identification of practices and comprehensive coverage of the variety of topics covered in the interviews. To this end, the interview data were first transcribed in a DSGVO-compliant manner using the f4xtranscript software[3], and then the transcripts were proofread and corrected. The transcripts were then loaded into MAXQDA 2022 and the sections structured by the research questions were color-coded. This was followed by familiarization with the entire data set to develop an understanding of the topics covered in the interviews.

No predefined categories or codes were used to inductively code the interview data according to Mayring's methodology. Instead, recurring patterns, key words, and themes were identified in the transcripts. These preliminary categories were iteratively developed and further refined to ensure that they adequately reflected the diversity of concepts and ideas discussed in the interviews. From these, categories were derived that were predominantly aligned with the research questions. For the purposes of this study, only the categories related to the identification of practices are considered.

The coding of the interview data was done systematically by assigning one or more of the identified categories to each relevant section of text. Once coding was completed, the data were interpreted by analyzing the meaning and relevance in terms of the definition of practices given above. From this, an initial set of possible practices was first developed, which were then grouped using the *Creative Coding* tool in MAXQDA. The grouping was done thematically; the individual group titles do not themselves represent a practice.

Transcription, coding, and grouping were prepared by a single individual, the author, for presentation in this thesis. This has implications for the quality of the analysis, which will be discussed in the limitations section.

5.3 The Practices of CS: Results

20 interviews were conducted for this study, thus, not all twelve subgroups of the ACM CCS could be covered by two interviews. This concerns the groups computer systems organization, security and privacy, hardware and computing methodologies. Interviews were conducted between November 2022 and July 2023, a majority (12) online, a few in presence (8). The recordings have an average length of 17 minutes, 48 seconds (standard deviation: 5 minutes, 30 seconds). All interviews were conducted in German.

To ensure the privacy of interview participants, descriptive information on the interviewees is provided without attribution to participants and only in summary form: 6 of the participants are female, 13 male; one interviewee identifies as non-binary. The youngest participant is 24 years old, while the oldest is 58 years old; the average age is 36. Participants have different educational backgrounds and current occupations: Participants from the business sector (8)

---

[3] `https://www.audiotranskription.de/f4x/`

work independently (1), in small (2) or medium-sized (2) companies, and some in large international corporations (3). With regard to the participants from academia (6), professors (2) as well as research assistants (4) are represented. Participants from the public sector (6) include teachers (2) as well as employees of government agencies (4).

The participants educational background also differs greatly: Some have undergone vocational training (1), dropped out of computer science study programs (2), had lateral entry into the discipline via occupation (6), or completed master's (5) or PhD studies (6).

According to the method outlined before, 56 practices in eight groups were identified (s. Fig. 5.2). They will be explained group-wise (in descending order by number of codings) in the following:

Social This group is comprises the following practices: Networking, Discussions with colleagues, Meetings, Interface between departments, Acting as influencer, Helping with (technical) problems, Communicate/explain technology, Checking mails/messaging services, Organizing events, Delegate.

This group of practices includes everything that is primarily aimed at communication and the organization of social interaction, without having a specific purpose. It includes the code that was mentioned most frequently (16) and by the most participants (11): *Discussions with colleagues*. As already explained above as an example, this does not mean an unorganized, random conversation, but recurring, professional discussions that have a certain structure. If this structure is very pronounced, then it can include, for example, a predefined agenda or minutes, as was reported by participants. Closely related to this are *Meetings*. These are practices that involve a larger group of people, especially those from outside the field. They differ from the *Interface between departments* practice, which is also quite close, in that a common project or a common concern is pursued at the center of a meeting, while the interface practice focuses on mediation between the departments, i.e. the subject matter changes in terms of a transmission function. This requires a modified recourse to the shared inventory, as reported by interview participants: The language used has to be modified, additional means such as visualizations or documentation have to be drawn upon. The practices *Acting as influencer*, *Helping with (technical) problems* and *Communicate/explain technology* can also be considered as a subgroup because they each aim at communicating content or explaining technology. The remaining practices, which revolve around everyday communication, reading e-mails, organizing events, and delegating tasks, mentioned by several interview participants in leading positions, have a more organizational background.

In summary, the group "Social" not only contains the most frequently mentioned code, but also combines the most codes. In terms of the number of codes, it is the second largest group (after "Meta").

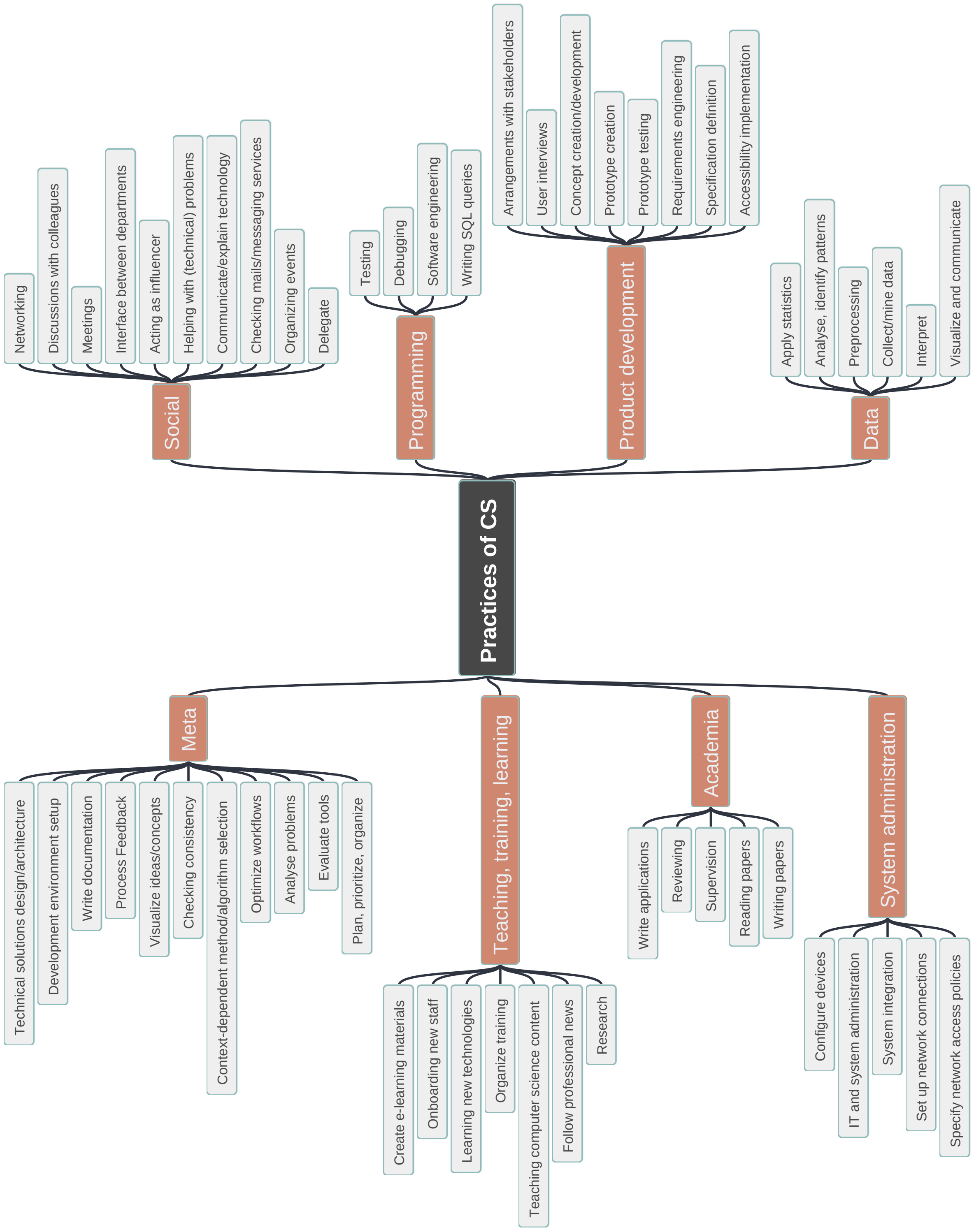


Figure 5.2. Overview of computer science practices identified in the interview study: the leaves (light gray) name the practices, while the inner nodes (orange) indicate the grouping.

Meta This group is composed of the following practices: Technical solutions design/architecture, Development environment setup, Write documentation, Process Feedback, Visualize ideas/concepts, Checking consistency, Context-dependent method/algorithm selection, Optimize workflows, Analyze problems, Evaluate tools, Plan/prioritize/organize.

The Meta group combines practices that are of overarching importance. Many of the practices mentioned here can be – in contrast to the practices in e.g. "Social" – regarded as actually specific to computer science. For example, *Technical solutions design/architecture* can be found, which was regarded as genuinely computer sciency by the interviewees. This applies to further practices such as *Development environment setup*, which is a recurring activity for software developers, which at the same time requires specific knowledge depending on the development environment and must therefore be learned. *Checking consistency*, *Optimizing workflows*, *Analyzing problems* and *Evaluating tools*, on the other hand, are frequently mentioned as overarching skills of computer scientists, which they attribute to themselves in their work contexts, but also state beyond that for their everyday life. Writing documentation and visualizing ideas and concepts are again communicative skills, but they do not necessarily have to be used in an intersubjective context – they can also serve to secure results or to communicate with oneself.

The "Meta" group contains the most diverse codes. *Technical solutions design/architecture* (4) and *Analyze problems* (3) were mentioned by more than two interview participants, while the other codes were each mentioned once or twice.

Programming This group encompasses the following practices: Testing, Debugging, Software engineering, Writing SQL queries.

"Programming" contains only a few codes, but they all came up several times in the interviews. In particular, *Software Engineering* was also mentioned by a particularly large number of participants (8) as a characteristic of computer science. In naming the code, I followed Bergstrom and Blackwell [BB16], who use software engineering to refer to the type of programming practiced in professional contexts as opposed to less or differently organized types of programming. Other types of programming or subtypes are found in the present material as *Software testing*, *Debugging*, and *Writing SQL queries*, each of which brings with it its own tools, approaches, and vocabulary. While testing and debugging are readily apparent as practices in their own right, in part because of the examples developed earlier, I chose to code *Writing SQL queries* separately: according to the interview participants' descriptions, it is a highly exploratory activity that has a different goal than standard software development. Instead of producing software, the focus here is on creating a specific subset of data.

Taken together, while "programming" is the smallest group in terms of the number of distinct codes, with *Software engineering* it has the second most frequently mentioned code overall. Nevertheless, it is surprising that a discipline often associated with programming produces so few diverse practices of programming. I will return to this point in the discussion.

Teaching, training, learning This group is composed of the following practices: Create e-learning materials, Onboarding new staff, Learning new technologies, Organize training, Teaching computer science content, Follow professional news, Research.

The group of interviewees included teachers at schools as well as teachers from universities. Accordingly, it may come as little surprise that there is a group with teaching practices dedicated to teaching computer science content and creating (e-)learning materials. In addition, however, there are other practices that show significance across the board: For example, new employees have to be trained and further training has to be organized. But also self-education by learning new technologies, following professional news, and researching tools and methods play a role as organized practices. Here, interviewees report on how they make time for this in their daily work schedule and what resources (e.g., news portals) and methods they use to self-organize their learning activities.

The group "Teaching, training, learning" contains seven different codes, of which *Teaching computer science content* was mentioned most frequently (4).

Product development This group encompasses the following practices: Arrangements with stakeholders, User interviews, Concept creation/development, Prototype creation, Prototype testing, Requirements engineering, Specification definition, Accessibility implementation.

This group of practices occurs mainly in the business domain and includes various activities related to the creation and development of a product. There are again specific communicative practices, such as conducting user interviews and agreements with stakeholders, but also abstract ones that deal with the conception of the product (*Concept creation/development*, *Requirements engineering*, *Specification definition*). The latter again draw on their own set of methods and tools, which are frequently located by the interview participants in the context of Agile Development. Three further codes deal with the implementation of ideas: the creation and testing of prototypes, and the implementation of accessibility. This code comes from a participant who works in the public sector, where accessibility is mandatory for software applications due to the implementation of European legislation. In the interview, the participant describes how this necessity has evolved into a workflow that includes an internal review board, the creation of automated reports, and the organized processing of the resulting requirements.

"Product development" includes eight practices, of which *Concept creation/development* and *Requirements engineering* were mentioned most frequently (3). Codes from this group, as noted, come mostly from interviews with computer scientists working in business.

Academia This group is composed of the following practices: Write applications, Reviewing, Supervision, Reading papers, Writing papers.

Like business, academia has developed its own set of practices that will be familiar to anyone who works in science: Proposals and papers need to be written, articles and papers need to be read and reviewed. In addition, all professional levels report on how they supervise junior academic staff (i.e., undergraduate or graduate students), each of which can also be considered a distinct practice with a specific set of resources.

This group includes five individual codes. Both professors in the interview pool also mentioned academic self-governance as a recurring activity. However, since this is too broad a catch-all term, encompassing very different things, it was not included here.

Data This group is composed of the following practices: Apply statistics, Analyse/identify patterns, Preprocessing, Collect/mine data, Interpret, Visualize and communicate.

"Data" groups practices are organized around data and cover the entire workflow of data processing: from collection, pre-processing, application of statistical methods and identification of patterns, interpretation, to visualization and communication of the data. What is most interesting about this group is that it can be found in all sectors (business, academia, public sector). It is also a very interdisciplinary group of practices, as people from different backgrounds describe and use the same practices to go about their daily work. In part, this similarity was made explicit by participants in the interviews when they noted that their work is not so different from that of other computer scientists who seem to come from a completely opposite field.

Of the eight codes in this group, analyzing and identifying patterns was mentioned most frequently (3).

System administration This group contains the following practices: Configure devices, IT and system administration, System integration, Set up network connections, Specify network access policies.

This last group includes various activities that revolve around the administration of IT systems, configuring and setting up devices, integrating systems and system landscapes, setting up network connections and specifying network access policies. These practices are marginal in the interview material in that they were mentioned by only three individuals, all of whom work quite hardware and/or network oriented.

The five codes in this group were mentioned least frequently overall, with only IT and system administration (3) having more than two mentions.

5.4 Why Practices Matter

Discussion The grouping of practices shows clearly that there are a number of practices that can be assigned to specific roles or areas. This applies to "Academia", "Product development" and "System administration". In addition, however, there are also practices that are more overarching and affect everyone. These are mainly located in the groups "Meta", "Social" and (partly) in "Teaching, training, learning". The latter are particularly important for the perception of the discipline by others and by the discipline itself. "Programming" and "Data", on the other hand, cannot be clearly assigned to either the specific or the general area.

The practices in "Meta" can be well connected to the processes (see above): Zendler et al. [ZSK08] list problem solving and problem posing, classifying, finding relationships, investigating, analyzing, and generalizing as processes of computer science. "Meta" contains with *Technical solutions design*, *Checking consistency*, *Optimize workflows*, *Analyze problems*, as well as *Plan, prioritize, organize* some general practices that show a high similarity to the processes or directly refer back to them. These very general or meta-practices are therefore closest to the cognitive processes often associated with computer science and reflected in Zendler et al.'s processes [ZSK08]. The practices in this group can also be identified across all fields (business, academia, public service) and subdisciplines. If there is a unifying core of practices in computer science, it is to be located here, although some practices (e.g. *Development environment setup*) may be too specific to be included in this core.

Practices in the "Teaching, training and learning" group were also identified across the board, not just those explicitly concerned with teaching computer science content. This may be due to the fact that computer science, even more than other fields, is subject to rapid changes in content and therefore requires lifelong learning. However, not only individual learning is evident in the practices identified, but also the need to communicate computer science content, whether through visualizing concepts, writing documentation, explaining technology, or interfacing departments. Computer scientists in practice are thus often communicators and mediators of computer science or computer science content. This aspect of the subject is currently hardly reflected in curricula and is only specifically addressed in teacher training programs.

Regarding the "Social" group, it is interesting to note that when asked about the percentage distribution of working time, most of the participants clearly spend the most time on the practices of this group. Moreover, for a part of the interviewees it represents the real core of their work, not programming, handling data or administering systems. This may be because there is a small bias in the interviewee population towards very highly skilled individuals performing managerial tasks. On the other hand, the phenomenon is also evident for and was mentioned by "ordinary" software developers. The importance of social practices is thus significantly higher than expected and paints a picture of computer science that clearly contradicts common clichés of introverted nerds: The work of a computer scientist in any field has a very high communicative and mediating component.

The practices of the "Programming" group are difficult to further differentiate from the given material, similar to what can be observed in Bergstrom and Blackwell [BB16]. Such differentiation would require more specific inquiries or a study using a different methodology, such as participant observation. Overall, however, the importance of programming for the professional computer scientist is lower than one might expect. According to respondents, even full-time software developers do not spend the majority of their time programming. At the same time, programming was mentioned several times by interviewees as a distinguishing criterion of computer science, often ex negativo. For example, one participant who primarily implements systems administration practices in their professional day-to-day life questioned whether they could be considered a computer scientist, since, after all, they does not program. At the same time, it was noted several times that it would be desirable for computer sciences to be more closely associated with these very practices.

With regard to the group "Data", it becomes apparent that the practices mentioned here are the same for e.g. digital humanities as for "hardcore" machine learners[4]: All of them could theoretically find themselves in these practices. This is also shown by the interviews. Nevertheless, different practitioners feel differently comfortable identifying themselves as computer scientists. This suggests that the practices alone are not sufficient for identification. It is evident in the interview data that educational background plays a major role: Even if similar or even the same practices are performed, identification is easier if one also has a formal degree as a computer scientist.

The question of connections between practices to retention and diversity is much more difficult to answer for practices and almost necessarily requires recourse to one of the other perspectives, conceptions, identity and/or values. Practices are part of and draw on the shared repertoire of a community. This ensures that they are an externally visible and observable feature of a discipline. Thus, performing or mastering a practice associated with a discipline ensures that one is perceived as a member of the community; moreover, practices are associated with a community. Both of these are important for questions of *Conceptions* and *Identity*, as seen in the previous chapters.

In the context of retention and diversity, practices gain relevance when individual practices are invoked as particularly salient markers of inclusion or exclusion. A strong example of this, already criticized in the past [PP13], would be the strong focus on programming in CSE: Based on the catalog of practices given above, within which programming plays a lesser role, it is questionable whether programming should be used as a criterion of distinction in computer science. Put differently: The very notion of a closed catalog of computer science practices is something that could be off-putting and exclusionary. Instead, practices should be

---

[4] In the interviews they were mentioned by people from both (and more) fields.

seen as open and as a way to create identification that is expandable and inclusive. Viewing diversity in the practices of the discipline as an ideal and a positive attribute creates greater opportunities to attract, retain, and bring a diverse group of people into the discipline.

A consideration of the similarities and differences between the practices perspective and the other perspectives leads primarily to the realization that the practices complement the other perspectives: As discussed, conceptions of the subject are very much attached to underlying, (publicly) perceived practices. Identity is strongly tied to the performance of practices, so much so that one of the best-known theories places Communities of Practice [LW91] at its center and views learning as the acquisition of practices, which simultaneously involves the formation of a professional identity. Values come into play only when they normalize actions [App18]; but repetitive, central actions within a professional community are practices. Practices are thus closely related to all other perspectives and hold a somewhat special role. This argument, and the precise role of practices, will be explored in more detail in the next chapter, when the perspectives are contrasted and distinguished from one another.

Limitations The background of the concept of practices developed here are socio-cultural identity theories. These emphasize the dependence on specific cultural and social contexts, which per se develop and can never be regarded as complete. The corpus of practices identified in this way can therefore not be considered closed – not even on the basis of theoretical considerations. Moreover, it is itself dependent on the context in which it was collected: The practices presented here only give an insight into practices at a certain point in time, in a Western European industrialized country.

Furthermore, the way in which the interview participants were sampled imposes certain limitations. Although the method used attempted to ensure that a wide range of the discipline was represented, the individual representatives of the sub-fields were primarily recruited through contacts or recommendations (snowballing). This essentially means that an existing social network was used, which entails certain limitations and biases. This can be seen, for example, in the fact that there is a clear bias toward high academic degrees, even among the interview participants from the economy. In contrast, the educational background through vocational training, which is strong in Germany, is hardly represented.

Finally, the data as presented here were collected by only one person. This person conducted, transcribed, corrected, coded and interpreted the interviews. Although interim results were repeatedly discussed with colleagues in the process and presented at seminar meetings, this has implications for the reliability and validity of the data. These implications may not be too severe, since the coding of the data was very close to the material, and only grouping was done as a further abstraction. The identified practices themselves are to a large extend of no great surprise value and, like e.g. programming, problem analysis and system administration, belong to the public image of computer scientists. Nevertheless, this limitation should be taken into account in the further use of the results; prior to a peer-reviewed publication, quality assurance of the results should be ensured by a critical review and recoding of the data by one or more other persons.

Conclusions The study presented here shows the diversity of practices computer scientists perform. This diversity contradicts the often narrow image of computer science as applied mathematics, programming, or even machine learning. It should be made clearer in teaching that computer science is a diverse subject. This does not require that all practices themselves become the subject of teaching: The subject need not be watered down in this sense. But it would be desirable if diversity was given more visibility.

For this, it would also be important to conduct further research in this area, because the results presented here are only a first beginning. In the future, the results should be deepened and expanded so that a larger picture of practice diversity emerges. In addition, it has been pointed out in the limitations that practices have socio-cultural dependencies, so they can and will change over time. Accordingly, studies surveying practices should be conducted in other contexts and with some regularity to reflect the changes that occur within the discipline.

In addition, it would be desirable to include the other perspectives in overarching studies: How, for example, do novice practitioners' conceptions of the subject relate to actual practices? Is there much difference? Which practices primarily reinforce identity as a computer scientist or lead to perception as part of the discipline? Ultimately, it would still be desirable to use the insights gathered in this way to develop tools and interventions that move away from a normative and purely content-focused approach to a more permeable and less exclusive view of the discipline.

Chapter 6

Perspectives on Computer Science

The idea of different perspectives on computer science, expressed in the different research approaches and concepts presented in the introduction (s. Sec. 1.1), has been with me since I worked on the identity review in 2021 [GGG+21]. In order to get an initial overview and to illustrate the connections and content overlaps between these perspectives and concepts, I designed a visualization (which has now been revised several times since the first draft) that arranges them in a two-dimensional grid according to how they fit between self and world view on the one hand, and descriptive or ideal/normative characteristics on the other (s. Fig. 6.1).

In this visualization, it is clear that perspectives actually form an epistemological continuum, performing distinct tasks in different dimensions and thus opening up different areas of reasoning about computer science. The previous chapters have discussed four such perspectives and how they contribute to the constitution of their intersections (s. Fig. 6.1, darker areas): What do the perspectives have in common, how do they differ? In the following,

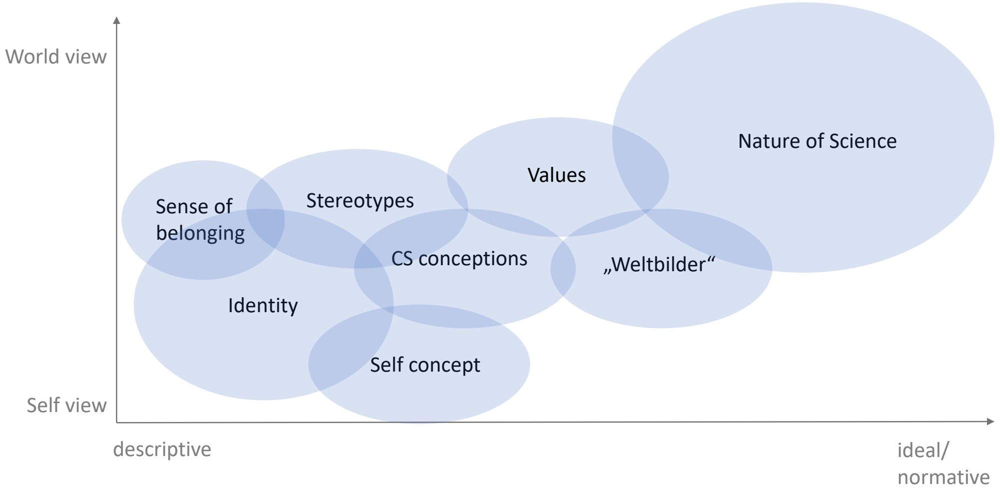


Figure 6.1. Different perspectives on computer science placed in a coordinate system in the dimensions “Self view — World view” (y axis) and “descriptive — ideal/normative” (x axis). The placement of each perspective is exemplary and open for discussion, it serves to illustrate the general idea.

I summarize the results and, through the summary, answer the research questions raised in Sec. 1.2:

RQ1: What theoretical affordances does each perspective provide to examine challenges related to retention and diversity?

Conceptions, as presented in Ch. 2, have the potential to be used as the basis for interventions that can specifically support groups of students by creating typifications. Such interventions would primarily aim to align students' conceptions with a normative image of computer science, facilitating their entry into the program and thereby address the problem of retention. However, the normative conceptions may also deter students who already feel excluded, thus jeopardizing diversity. It was also discussed that students' conceptions themselves are insufficient as a theoretical framework to explain complex phenomena over time. A different framework is needed to address such issues. Conceptions thus have their own theoretical niche, but they needs to be supplemented with other perspectives to address broader questions.

The systematic literature review (s. Ch. 3) has shown that identity is used as an analytical lens to look at problems of retention as well as diversity. This versatility is grounded in a variety of theories that have different foci and has already been appreciated by other researchers (as cited above, "Identity is a lens that is adjustable" [Dar16]). The explanatory possibilities that identity theory and the identity perspective offer to computer science are thus undisputed. At the same time, it is noteworthy that there are only a limited number of interventions and instruments based on this. Possibly this is due to the fact that socio-cultural theories are skeptical of an essentialist view [SP05] of individuals that might be expressed through categorizations: A theory that aims to comprehend each individual in its versatility and as a unique intersection of different identities is inadequate to provide the basis for typification [DKG+23]. Thus, for the identification and accurate description of problems, this perspective is well suited, but should be complemented by other perspectives for the development of interventions and tools.

Values as a perspective in their own right have not been well researched to date, and only preliminary results were presented in Ch. 4, but there is a case for further investigation of this perspective: Values are part of a community's shared repertoire that is drawn upon to assign or deny meaning to things. If it is not part of a community to recognize diversity and polyphony as a value (or simply other values are more prominent), then this could explain why change in a professional culture is slow and a problem such as diversity, which as noted above (s. Ch. 1) has been an issue for decades, is not improving. It has also been discussed in Ch. 2 that students who are members of historically marginalized groups are more likely to turn their backs on their studies because they see no possibility of finding the values of their community in the subject of computer science. Thus, this perspective is at the intersection of retention and diversity. At the same time, this perspective needs to be complemented by other perspectives to explain how values are lived and realized within a community.

This task could be accomplished by the practices perspective. Practices are normed by values, so that within a discipline it becomes possible to judge whether something has been done well or badly. This has become clear in the examples of testing and the selection of programming languages (s. Ch. 4). Practices form an important part of the public image of a discipline and thus also shape the conception of novices: this is particularly evident, for example, in the type of *Programmer* mentioned in Ch. 5, but also the types of *Mathematician*, *Technician*, and *Interpreter* are tied to certain expectations of practices that a computer scientist engages in. It has also become clear that students' accounts of engaging in practices acted as "aha" moments and thus directly became identity-shaping narratives. The importance of practices for identity is also evident in the interviews with practitioners in academia and business, many of whom made their identification directly dependent on the practices they engaged in. Practices, then, are a perspective that can be found as part of all other perspectives, as discussed for RQ2. In terms of retention and diversity, practices are relevant in that perceptions of a particular canon of practices can have inclusive and exclusionary effects. If an existing set of practices is understood normatively (they exist and only those who practice them are informants) this has the potential to be a deterrent.

RQ2: How do the perspectives interrelate, and what are the points of convergence and divergence among them?

Some of the linkages between perspectives have already been addressed in the response to RQ1, so there are minor redundancies for the response to RQ2. These mainly concern practices, because, as RQ1 showed, the nature of practices as perspectives is to illuminate parts of other perspectives. That being said, I will base my discussion on the findings developed in the previous chapters.

The connection between conceptions and identity is, first and foremost, a difference in self image/world image emphasis: Conceptions attempt to focus on learners' worldview in relation to the discipline of computer science, while identity examines students' self image as members of the discipline. As a result, the perspectives are not without overlap and are not completely separable, as could be seen in the discussion of the personified naming of types as programmers, mathematicians, etc. Nevertheless, identity is a more comprehensive perspective, because it includes not only the relation between individual and subject, but also other relations (individual — self, individual — persons outside the subject, etc.). This makes it possible to explain more complex phenomena. In contrast, conceptions are "easier to handle" as a theory. Both perspectives complement each other quite well, depending on the question at hand. As shown, this is true for retention, while diversity requires more complex explanatory models.

Conceptions and practices also have a close connection: As noted earlier, public perceptions are shaped by activities such as programming, hacking, or wrenching on hardware. These are correspondingly reflected in the conceptions that students bring with them to university (as discussed in particular in I and II). Moreover, it is evident that the types of conceptions identified in Ch. 2 are organized around a set of practices, which is already clear from the respective naming (and would not only apply to this study, but also, for example, to Hewner's [Hew13] findings). The Differentiated Picture or Broad View are organized around practices as well, except that in this case they represent the intersection of the various other types.

As shown, the relationship between identity and values is that values, in the first of the two understandings outlined in Ch. 4, can be understood as a distinct feature of identity, as a subconstruct in operationalizations. Values can promote identification if they are shared, or serve as a criterion of exclusion if conflicting values are present in the community or other values are given priority. Accordingly, an analysis of values can help to shed light on the understanding of identification in a community. Moreover, it is difficult to gain meaningful insights into the values of computer science by studying the identity of a single individual: Researchers need to look at the whole community. Thus, the values perspective is to be understood in particular as complementary to the concept of identity, not vice versa. Values, as seen, form a part of the repertoire of the community of practice of computer science. They are therefore in part independent of the individual and their meaning for the identification of the individual only comes into play through the actualization in actions.

This highlights a feature of the connection between identities and practices: Performing practices in meaningful actions creates an identification of the individual with the corresponding discipline. The importance of practices for identity was initially established by Lave and Wenger [LW91] and later elaborated by Wenger:

> There is a profound connection between identity and practice. Developing a practice requires the formation of a community whose members can engage with one another and thus acknowledge each other as participants. As a consequence, practice entails the negotiation of ways of being a person in the context. This negotiation may be silent; participants may not necessarily talk directly about that issue. But whether or not they address the question directly, they deal with it through the way they engage in action with one another and relate to one another. Inevitably, our practices deal with the profound issue of how to be a human being. In this sense, the formation of a community of practice is also the negotiation of identities. [Wen99, p. 149]

This connection is so fundamental that it was considered a prerequisite of my own research.

The values and conceptions perspectives have not been considered in detail here, and the provided elaborations fall short in establishing a clear connection between the two concepts. The different types of conceptions do not seem to enforce or realize certain values. On the other hand, neither are the values themselves linked to specific conceptions. So, with the current state of knowledge, there seems to be a connection of the perspectives only transitively through others, such as those of identity or practices.

That values and practices have a connection has already been discussed and is easily comprehensible: Values standardize behavior and thus the realization of practices. They allow a statement about whether a practice was performed well or badly, and thus provide the basis for talking meaningfully about the execution of actions in computer science. Another connection that is less obvious is that both values and practices can be inventoried: They can be collected, listed, and then used to develop interventions and tools, for example, through qualitative research. In doing so, however, there is always the risk of creating an overly static picture, which in turn may develop normative force.

During the discussion of the research questions it should have become clear that if there is something like a connecting link between the perspectives, it is most likely to be *practices*. The can, similar to values, be considered as a perspective of its own (as discussed in Ch. 5), but also always as *part of* other perspectives. Practices show a connection not only to conceptions, values, and identity, but also, as shown in Ch. 5, to epistemology and philosophy of science. Moreover, as in the case of the connection between values and conceptions, they may function as a transitive link.

Practices may have a tacit component, as suggested by Fleck (s. Ch. 5), but they are nonetheless observable and thus in some sense objectively measurable. They might therefore offer a way to make other, harder-to-measure perspectives (such as identity) more accessible. What this might look like is illustrated by a tool developed by Mercier et al. [MBO06] called *Fluency Building Items*, which contains, for example, the following: written code using a programming language like C, Java, Logo, Perl; made a database; built a robot or created an invention of any kind using technology; created an animation or cartoon [MBO06]. Students who were presented with the instrument were asked to indicate which of these activities they had done at least once. Mercier et al. were able to demonstrate a statistical correlation between the number of corresponding activities and identification as a computer person [MBO06].

Practices might also be linked to competencies, as shortly discussed in Ch. 5. This opens up the vast field of research in this area for further investigation. However, there is a caveat to these possibilities that should be emphasized again: Reification into instruments runs the risk of such practices being perceived as the normative inventory of the discipline, creating a closed picture that is explicitly not intended to be the goal. It is necessary to understand such instrument developments, especially in the context of identifying computer scientists, as temporally and locally situated means that require regular updating. They serve descriptive, not normative purposes.

Chapter 7

Limitations and Positionality

Since each of the papers used as the basis for this thesis has its own section on limitations, only some common and fundamental aspects will be discussed here.

The research presented was conducted at the University of Kiel, and thus at a medium-sized German university located in a major city, but only one of few in a mostly rural state. This places certain limitations on the validity of the studies that have been conducted on students and faculty members. With respect to I and II, this does not appear to be problematic, as some of the results already replicated older research by Hewner [Hew13] (s. Ch. 2), indicating reliability.

More fundamental are two conceptual difficulties that affect all studies, as well as this thesis: First, diversity is understood across my writing to mean mostly "too few women". Diversity would have to take a much broader view and also address questions about the representation of other historically marginalized groups within computer science. It is widely recognized that the current representation is inadequate, and that the situation is much worse than that of women [RLB18]; on the other hand, women represent the largest group that is not represented and, as is clear from the introduction, there are enough robust figures to speak meaningfully about the situation (s. Ch. 1). Diversity should not be misunderstood to mean that it is only about women, but looking at women is a first step. In this context, it should also be noted that important questions of intersectionality are completely ignored, e.g.: What is the situation of women with a migration background in computer science? It is known from research that such intersectional questions are not simply the sum of different problem situations, but that there are specific, unique questions that are linked to this situation [DK20, Ch. 1]. The above observation that research at the University of Kiel poses certain difficulties because of the relatively homogeneous student body also applies here.[1]

Second, I have referred to *computer science* throughout the thesis. In German usage, however, the term *Informatik* is more common as a disciplinary designation. While this issue has been addressed in each of the publications, it also extends to this thesis and even more so as its perspectives on the discipline are discussed. There may be differences in detail between the understanding of Computer Science and *Informatik*, but what matters to me and here is that it encompasses one and the same community of practice, which can be

---

[1] This is a well documented fact due to the extensive investigation into the student body during the KOI project detailed in Ch. 2.

demonstrated by several observations: Computer scientists publish at the same conferences worldwide, mostly under the umbrella of ACM and IEEE, there are close ties not only in research but also in virtually all other areas, such as sharing open source code via Github, sharing problems at Stackoverflow, etc. All this does not concern either computer scientists or *Informatiker:innen*, but both groups equally: There is a basis of common exchange that spans conceptual difference. In this respect, I think it is justified to speak of one discipline, because even though there may be different foci or centers, there are enough unifying practices and a *shared repertoire* (s. Ch. 4).

7.1 Positionality Statement

In recent years, as part of the reflection and critical examination of bias in research, it has become established to disclose one's own positionality and situatedness in the form of a positionality statement [Hub21; RBD+20; HRO21]. Especially in relation to qualitative research, I find it profitable and necessary to reflect on one's own positionality, since qualitative research is, in my eyes, sense making, and what sense is generated depends heavily on the personality and conditions of the researcher.

My background is that of a white, middle-aged, cis-male living and working in a big city, in the north of Germany and thus in one of the richest countries on earth, which is strongly influenced by western culture and lifestyle. Besides a degree in computer science, I studied philosophy and worked as a software developer in business for several years. This background has shaped me to have a strong bias towards theoretical and qualitative work, although I recognize the value of quantitative research and try to give each of the different research streams their due place in my perception.

Through my research and teaching, I have come to believe that diversity and equal access to knowledge and science are necessary conditions for the functioning of modern societies and democracies. This is especially important for computer science, as our discipline provides the means that shape modern societies. This attitude is reflected, for example, in the effort to select interview participants in such a way that as many different perspectives as possible are brought up. This is in direct contradiction to a representative sample, in which white men would necessarily have to make up the majority of the interviewees. Moreover, this also forms the background to the discussion of descriptive and normative images of computer science, which has been addressed several times in this thesis, because I firmly believe that the current conditions did not arise by chance and are not immutable.

Nevertheless, I do not understand my research as activist research. With this statement I try to make transparent the personal background against which my research has been developed. Even if a similar statement is missing in the individual studies, I have tried to make every step of the research transparent and to include different perspectives in the analysis of my data by collaborating with other authors.

Chapter 8

Conclusion

Etienne Wenger's [Wen99] book, which has been referred to at various points in this thesis, sketches the following picture of a social theory of learning on its first pages:

Learning, according to this picture, is not an individual, primarily cognitive process, but a meaningful exercise of practices in a community of like-minded people to which the learner feels a sense of belonging. Fig. 8.1 explicitly features the perspectives of identity and practices, while meaning resembles the values perspective, and views of the discipline in the form of conceptions have a strong connection to the community aspect. This work has not organized these concepts around the central concept of learning, but has instead asked how far retention and diversity can be explained and influenced by conceptions, identity, values and practices. However, it is easy to imagine that the perspectives could be applied to other "centers". What

Figure 8.1. This visualization by Wenger [Wen99, p. 5] is originally captioned as: "Components of a social theory of learning: an initial inventory."

do values tell us about competencies, identity about motivation, practices about interest?

This thesis focuses on two central issues in computer science education research, namely retention and diversity, and attempts to make a contribution by showing how – from a theoretical perspective – both issues can be addressed. The individual contributions will be summarized below, followed by a discussion of questions that could be subject of future research.

8.1 Contributions

The study of identity, conceptions, and practices, as presented earlier, has yielded contributions in various ways, as outlined below.

Reproduction and extension of CS Conceptions research (Ch. 2) The conceptions with which novices approach computer science have already been extensively researched. However, a replication of the results has been lacking so far. Paper I replicates observations from preceding publications, especially with regard to the research of Hewner [Hew13] and in a differing cultural context (the original results come from the USA). In addition, the system of types of conceptions presented by Hewner could be supplemented by another type. Furthermore, the results of the underlying interview study provide the material to bridge identity theory and CS Conceptions research.

Identification of important lectures and extracurricular events that shape conception development during the first semester (Ch. 2) By comparing type changes over the course of the first semester through two successive rounds of interviews at the start and end, it was possible to identify events important to students for type change at the beginning of their studies. On the one hand, these were lectures and events that contained knowledge that had little to do with the previous type. On the other hand, there were "aha" moments, i.e. formative events that had a particular impact on the conception of computer science. These include, for example, programming projects in college or outside.

Connection between types and normative conceptions (Ch. 2) Through a third round of interviews – not with students, but with first semester instructors – and critical reflection on our own methodology, the connection between normative conceptions and students' typification was elaborated. This can be seen in an idealization of the Broad or Differentiated View of computer science, which can be found in the literature as well as in the statements of (especially central) lecturers. Among other things, this critical reflection has pointed to the theoretical gaps in this perspective, which point to an investigation of identity.

First comprehensive examination of the use of the concept of identity in computing education research (Ch. 3) Although the concept of identity is increasingly used in computing education research, there has been no overview of current research as it exists, for example, in mathematics or engineering. The lack of such a compilation not only makes it difficult for new researchers to get started with this particular aspect of research, but also prevents or at least hinders inter- and intradisciplinary exchange. This thesis and the publications III and IV fill this gap and provide a comprehensive overview of the literature published in this area. By conducting a qualitative content analysis, the results thereby go beyond a mere synopsis of the literature, but contribute to theory building itself. The category and code systems developed can serve as a basis for (intradisciplinary) discussion and debate in the future. Moreover, the presentation of the historical and conceptual roots of the concept of identity makes an important contribution by providing a quick introduction to an otherwise complicated theoretical landscape.

Identification of practices as a noticeable desideratum (Ch. 3) Through the systematic analysis of the literature against the background of the theoretical argument, it has become clear that there is a significant desideratum in research: While most research in the context of identity takes place in relation to socio-cultural theories, especially Lave and Wenger's [LW91; Wen99] theory, there is, however, no accurate description of computer science practices. A description of the practices of computer science therefore represents a noticeable desideratum that could be empirically proven in the literature review.

Theoretical exploration of CS Values (Ch. 4) To the best of the author's knowledge, there is no systematic discussion of values in computer science in CER, apart from the few articles published in connection with sense-of-belonging, which are referenced in Ch. 4. Accordingly, the theoretical exploration of the topic and justification of values as part of identity and as a distinct perspective, represents a meaningful contribution to the research.

Empirical extension of Mahadeo et al.'s instrument (Ch. 4) Another outcome of the systematic literature review was the theoretical extension of Mahadeo et al.'s [MHP20] construct of Computing Identity by a fourth sub-construct based on the existing literature. Chapter 4 presented the preliminary results of a survey study, analyzed through a confirmatory factor analysis, which justified a corresponding extension of the instrument to measure identity.

Identification of Computer Science Values (Ch. 4) Another preliminary study used a qualitative group survey to identify the values of computer science as perceived by advanced beginning students. This opens up, for the first time, a discussion of which values in computer science are currently given special consideration in teaching; a discussion that will further provide a basis for planning interventions and subsequent research.

Theoretical development of the concept of practices (Ch. 5) The above-mentioned desideratum of practices in computer science was addressed in another study, which was first devoted to clarifying the concept of practices. This theoretical contribution consists of a pragmatic notion of practices, borrowed from Wenger [Wen99] and the competence literature (especially McClelland [McC73]), whose use in research is less presuppositional than other conceptualizations represented in recent literature.

Differentiating practices through an extensive interview study (Ch. 5) Based on the refined concept of practices, an extensive interview study with practitioners from business and academia working in a wide range of computer science fields identified key practices and discussed their relationship to an identification with the discipline. On this basis, interventions and further research can be planned in the future.

Linking and differentiating the various perspectives (Ch. 6) The four perspectives of Conceptions, Identity, Values and Practices, which were differentiated in the thesis, were for the first time placed in a common context and discussed against the background of the problem areas of retention and diversity. This theoretical development has contributed to further research by clarifying the terminology and identifying the complementary and differentiating characteristics, which will facilitate the use of the concepts in further research.

8.2 Future Research

With regard to future research, the individual papers, as well as Ch. 4 and Ch. 5, offer suggestions for the individual perspectives. Accordingly, this section deals solely with proposals that relate to overarching issues. In addition, I have taken the liberty of formulating not only suggestions but also wishes for further research that are, however, unlikely to be implemented due to the complex research designs that they require.

It would be highly desirable and a reasonable goal of future research to conduct comparative case studies across different educational institutions to assess the integration of perspectives depending on the institutional context. For the German context, this concerns, for example, the comparison of universities, universities of applied sciences, but also the vocational education in the field of computer science. In addition, it has become clear from the figures reported in Ch. 1 that there are considerable differences between the various computer science programs in terms of diversity: How is this reflected in students' conceptions, identities, values, and practices? Where are the differences and what can computer science learn from "hyphenated" (e.g. media or bio informatics) majors?

The observation made in Ch. 4 that students are not able to give an assessment of themselves through the eyes of faculty points to another possible research direction: What if making prospective teachers aware of their central role in shaping the identities and values of their charges became an integral part of teacher education for schools and colleges? Perhaps the mere encouragement to reflect on such issues could make teachers sensible for their prominent position and thus influence their teaching. In addition, the teacher-student relationship could be specifically examined: For example, what are the effects of the teacher addressing his students as computer scientists, giving them regular feedback on their learning progress, etc.? The last point in particular, while identified e.g. as an important factor in learning progress in the Hattie study [Hat08, p. 243], has not yet been examined in detail for its significance in the context of identity development.

Educational institutions also have an identity, a self-image that is expressed through the self-image of faculty and public communication, but also through policies and curricula. What influence does this shared repertoire have on learning success, the culture of the institution, and the social climate? In Ch. 2 it became clear how the curriculum has shaped students' perceptions of the discipline; Ch. 4 discussed, that a course such as "Ethics in Computer Science" provides an understanding of values and instills certain value concepts in students, such as professional responsibility. It would be an interesting subject of future research to pursue similar questions and, in particular, to explore what elements of the course or training influence the development of students' conceptions, identities, values, and practices and to what extent. This is likely to be primarily through the courses, but, as noted, also to a large extent through other institutional settings.

All of these issues, and especially questions resolving around conceptions, identity, values, and practices, are directly dependent on the subject culture, the change of which is a continuous and difficult-to-observe process, just as individual identity development occurs over time. Therefore, long-term studies that include and integrate qualitative and quantitative methods are needed to focus on these phenomena. The KOI study presented in Ch. 2 was accordingly on the right track, but its research questions were too broad in scope to yield specific results.

Appendix A

Papers

The papers on which this dissertation is based are cited in full below. The cited version corresponds to the manuscript versions accepted by the respective conferences and journal.

For the published version, see the citation notes or DOIs.

A.1 "It's like computers speak a different language" Beginning Students' Conceptions of Computer Science

Gregor Große-Bölting, Yannick Schneider, and Andreas Mühling. 2019. It's like computers speak a different language: Beginning Students' Conceptions of Computer Science. In Proceedings of the 19th Koli Calling International Conference on Computing Education Research (Koli Calling '19). Association for Computing Machinery, New York, NY, USA, Article 2, 1–5. `https://doi.org/10.1145/3364510.3364527`

“It’s like computers speak a different language”

Beginning Students’ Conceptions of Computer Science

Gregor Große-Bölting
Kiel University
Kiel, Germany
ggb@informatik.uni-kiel.de

Yannick Schneider
Kiel University
Kiel, Germany
ysc@informatik.uni-kiel.de

Andreas Mühling
Kiel University
Kiel, Germany
andreas.muehling@informatik.uni-kiel.de

ABSTRACT

Students who enter university to study computer science have varying motivations, prior exposure to, and knowledge about computer science (CS). Accordingly, they have a great range of ideas about what their chosen subject *is*, often, as previous research has shown, rather unclear ideas. As part of a longitudinal research project, we conducted surveys and interviews with freshmen and thus collected students’ ideas of what computer science means to them. As a result, we present a classification of five prevalent types of beginning students that we identified qualitatively and quantitative support for the types. The types align partly with previous research results. We discuss the types and their potential (dis-)advantages in their CS studies.

CCS CONCEPTS

• **Social and professional topics → Student assessment**; *K-12 education*; *Adult education.*

KEYWORDS

student conceptions, beginning students, retention, qualitative analysis

1 INTRODUCTION

As in many other subjects, students in their first semester of computer science[1] have very diverse biographies. This leads, maybe more than in many other subjects, to a broad range of prior experiences with, assumptions about, and ideas of the subject when entering university. Is computer science perceived as being akin to natural sciences or to engineering? Is the study program seen as a mere training in programming or as a course in advanced, applied mathematics? As good teaching will strive to integrate each learner’s prior conception and existing knowledge structure, this situation poses a challenge for universities: How do we integrate the material taught in the first semester(s) meaningfully in such a diverse backdrop of ideas? How do we show “the bigger picture” beyond all the basics of the first courses in order to present students with opportunities of correcting their existing notions about our subject?

Even more, existing research suggests that incorrect perceptions of computer science deter students from enrolling for this subject in the first place and, if they do, will influence both their motivation and their success negatively [10, 11, 14]. Since dropout rates in computer science programs are usually reported as high [8] we think it’s worthwhile to investigate students’ perceptions of computer science, how they develop in their first semesters and how they influence students’ progress and success in order to adapt teaching and/or advertising strategies accordingly.

The results presented here are part of a longitudinal research project (KOI) that investigates how a cohort of students of computer science at Kiel University develops throughout the first three years of their Bachelor’s program in order to find out how local study conditions can be improved. As a part of this project, this paper is focused on the following research questions:

- **R1** What different types of perceptions can be observed among our population of beginning students and how are those types distributed?
- **R2** Is it possible to replicate previous research with regard to different types of perceptions of computer science and differences between students’ and professionals’ perceptions?

So, or goal was both to try reproducing prior research results and to lay a foundation for a long-term study of the changes in the understanding of computer science.

2 RELATED WORK

The understanding of the subject and its influence on the success of students has previously been examined by a number of authors. Hewner [10] conducted 37 interviews with students and advisors and analyzed them with Grounded Theory. He interviewed students from three different colleges throughout their studies and was able to identify three different views of the students on computer science, which he calls *Theory View*, *Programming View* and *Broad View*. The *Theory View* is guided by a mathematical or theoretical understanding, while the *Programming View* focuses on the activity of programming. The *Broad View* understands computer science as a science composed of different disciplines (e.g. theory, robotics, programming).

Funke, Berges and Hubwieser [6] surveyed 217 students in a biographical study to find out gender-specific differences in the perception of computer science. Their survey was conducted at the beginning of an introductory course in computer science and contained two items that were answered as free text and evaluated using Mayring’s qualitative content analysis [15]. They were able to identify different perceptions of computer science in relation to gender. The authors also note that many of the participants in their study seem to have no idea either of how computers work or of typical computer science topics, even though the majority of the participants had contact with computers since their earliest youth. Both Funke et al. and Hewner refer to Greening [9], who found

---

[1] Our research has taken place in the German-speaking area, where the term “Informatik” is used. It is sometimes translated as “informatics” but we chose to keep the more frequently used “computer science” (CS) and treat the two terms as equivalent.

that the majority of the students he interviewed in his study were not able to give a meaningful definition of computer science.

Knobelsdorf and Schulte used the biographical method to assess perspectives on the use of computers and attitudes towards computer science [13]. In this way, they collected self-images and world-images in addition to ideas on the field [22]. Their result is that, initially, there is a variety of different ideas [13], but these can be divided into three broad areas: Those of *use*, *professional use* and *design*. The authors express the assumption that a transition must happen from *use* to *design* during the course of study, but this step must be taken by the learners themselves: It does not help students to give an explanation of what computer science is, students have to experience it for themselves; world-image, self-image and habits are too closely interwoven [22].

Peters [17, 18] investigated in a longitudinal survey how students experience participation in computer science and how they identify as computer scientists. First, 120 essays written by students enrolled in an information technology and computer science program were collected. Then, over the following three years, 61 interviews were conducted with these students regarding their choice of study, future career, study experiences, and perception of computer science as a discipline. Peters discovered seven different ways of student participation in computer science: using, learning about technology, creating, problem solving, problem solving for others, creating new knowledge and contributing to social endeavours [18].

Biggers, Brauer and Yilmaz [2] investigated why computer science students leave the field. They sent questionnaires with open and closed questions to students and graduates of Georgia Tech and divided the answers into students who stayed in computer science for their major ("Stayer") and those who changed ("Leaver"). One of their open-ended questions was: "How do you define computer science/computing?" One result of their analysis was that Leavers often refer passively to the field and often refer negatively to the working conditions and environment. It was particularly striking for the authors that graduates often refer in their definitions to what computer science is *not*. For example, it is neither information technology, nor the repair of hardware, nor just coding. Exactly these definitions, however, were often found among the Leavers. Answers related to the efficient solving of problems or to the fact that computer science is a mixture of topics related to hardware and software were found with about the same frequency in both groups.

Schulte and Budde [21] question whether it is even necessary to start with a descriptive idea of what constitutes "the nature of the discipline" in order to create successful teaching. The (normative) hybrid interaction model presented by the authors is based on the question, what is best for the next generation. Nevertheless, they also assume that the descriptive conceptions and ideas can be used meaningfully for a precise design of computing education.

3 METHODOLOGY

Within our larger research project (as mentioned above, s. 1), all students of the first semester of computer science at Kiel University were asked, among other surveys, to fill out a questionnaires (distributed online) at the beginning of the semester. Additionally, interviews were conducted with a small group of students at the beginning of the semester. All data except for interviews were conducted anonymously with a self-constructed code allowing us to match responses from different sources and all data was collected voluntarily, however some parts, including taking part in the interviews was rewarded with a small Amazon gift card.

3.1 Open-ended questions

A large number of items have been presented in the first questionnaire: In addition to basic demographics, these included e.g. their final grades of school or prior programming experience. This questionnaire was completed by a total of 350 students.

Among the items was a free text question asking to "[g]ive a short definition of the term computer science that you think is appropriate". 310 students gave such a definition. In order to evaluate this question, the answers were graded independently by two of the authors as 0, 1 or 2 points. One rater rated without a manual, the other rater used a definition derived from literature as a comparison. The evaluation of the two raters shows a high agreement, with an inter-rater reliability of 0.79 (Cohen's κ). The reference used by the rater was created by encoding definitions from experts [20]. We used definitions of the ACM [5], of the German equivalent GI [7], of the Computer Science Teachers Association [23], a definition of a standard (German) textbook [3], and the description of a renowned German university [19]. In a second step, the consensus between these expert definitions was determined. The result of this process was that, while each definition in itself was more complex than the consensus, a minimal valid definition of computer science must contain that it is a *science* that either deals with *computers* or does *data or information processing*. A definition of a student that contained these concepts was given one point. A two-point definition had to additionally name any additional concept, such as e. g. *efficiency*, *algorithms*, *math* and *logics* or *social relevance*. All expert definitions would score 2 points, but also other, simpler ones can easily achieve this value.

3.2 Interviews

Interviews have been conducted with 14 students. They were selected from 125 students who indicated in the initial questionnaire that they are willing to participate in an interview based on their answers to the open-ended question, with the aim of hearing a wide range of perspectives. All interviews were conducted in a semi-structured setting by the first author and recorded. Participants were offered to conduct the interview anonymously, using only a pseudonym, however none of the students chose to do so.

The interviews lasted between 12 and 49 minutes, with a mean interview length of 28 minutes (standard deviation: 12 minutes). Seven participants were male, seven female. The participants were also a mixed group with regard to their background: Four have already successfully completed another course of study, while for three it is the second after an unsuccessful other course of study. Seven have just finished school and are beginning their first course of study.

The interview touched upon three subject areas: existing ideas of computer science, previous learning routine and overall changes in students' lives due to the start of their studies. With regard to the questions on the ideas of computer science, the interviewees were first read to their own short definition from the survey. They were then asked to elaborate on this definition. Depending on the answer, a number of follow-up questions were asked. One area that was always taken into account in the questions was their specific understanding of data and information and the interrelationship of these with computer science.

All interviews were coded following Mayring's approach of qualitative content analysis [15] independently by two of the authors. Both inductively developed a category system and then compared it to the other. After the first run there was already a high level of agreement, which was then refined by discussions. Out of these categories, five types of students were extracted. To validate these types, we matched the open-ended answers from the questionnaire to the types and were successful in always identifying one predominant type for all of the 310 responses. The first 50 answers were evaluated by two of the authors and the agreement was determined. Since the agreement was high (0.81, Cohen's κ), the remaining 260 answers were divided between the two coders and manually assigned to one of the types.

4 RESULTS

The evaluation of the interviews revealed five different types of notions about computer science. Through our non-random sampling method, we expect that the types presented here provide a good representation of the population under consideration.

4.1 Creator

A "creator" regards computer science primarily from the point of view that it provides a means to influence the real world. This influence usually consists of the fact that (some) problems can be solved particularly well with computer science. For her, programming is the most important part of our subject (all quotes from interviews are translated by the authors and the gender used in the description does not indicate the gender of the interviewee):

> Q: What is computer science for you?
> A: Well, definitely programming, creating programs, writing programs. Learning of one, if not several programming languages or at least the system behind a programming language. (26-15)

She also considers algorithms and the design of efficient algorithms to be important. Creators are less scientists and more engineers. In this group computer science is often perceived like a school subject: You have to learn a certain set of things and take exams on them and at the end you are a computer scientist in order to go and find work. There is no sign of them understanding the openness of a science. We have also identified a specific subgroup of the creator that we dubbed "idealist". Her goal, as the quote below is intended to illustrate, is to exert a positive influence on human problems on a large scale, such as climate change or global nutrition.

> So the question of how mobility can be regulated, how to get a grip on it, how to get environmental pollution under control. I think these are all algorithmic questions that you have to solve somehow, if you want to have a reasonably good life here on earth. (15-12)

4.2 Mathematician

The "mathematician" has a mathematical, or rather theoretical, view of computer science. For her, abstractions and systems are more important than specific applications. She is not interested in solving problems, but in solving puzzles. Neither programming, nor the technical side of computing machinery plays an important role for these types. But although they do not play a major role for them, all mathematicians have expressed satisfaction that with computer science they are able to try out mathematical constructs and see direct results.

> At the beginning I was a bit afraid because I had never programmed anything and I don't know anything about computer games, that I can't keep up. But now that I've noticed that even there it's actually just mathematics, to think about how to get there, to write mathematically, yes, a manual like that. (22-11)

4.3 Interpreter

The interpreter regards computer science as a "translation activity" between man and machine. For him, the machine is something that is almost sentient and can be understood and "addressed", but in a different way than humans. It is therefore necessary for someone or something to take on a mediating position. This is the task of the field of computer science and the computer scientist.

> It's like computers speak a different language. That's how I always imagined it. Because I never understood exactly what was happening. I only saw what was happening. It's like, for example, two people talking and suddenly one of them makes a somersault and the other doesn't know why. And then I just learn the language to understand why he did the somersault. And so it was with the computers. (19-14)

As with other types, there are echoes or references to this type even for interviewees that we did not take to be a translator. They express themselves in two ways: An emphasis on an interface function of computer science between man and machine and through the humanization of machine behavior.

4.4 No clear picture

> Q: Can you tell us a little bit about how you came to this definition or answer?
> A: Mostly because it was asked in the middle of the lesson and maybe twenty minutes before we had just done that. (14-14)

Some of the students had given a written definition, but could not elaborate on it on request. It was the textbook definition [3] that had been discussed in one of the lectures shortly before. The students who had given this kind of definition were able to replicate the definition, but could not explain its content. In addition, when asked, they could not give their own idea of what their chosen subject of studies was actually about.

4.5 Differentiated picture

A student with a differentiated view of computer science sees the subject as a mixture of different disciplines. For them, theory and programming are equally important and possibly only two of many parts that make up CS.

> So I think it's a mixture of... well, the hardware and the software. On the one hand, we're programming now during our studies and at Computer Systems we get to know how a computer works at all. And maybe that's the connection between both sides. (26-14)

It is the least common among our participants and is represented by only one student. We took it into account despite the small amount of evidence, because we know from the next rounds of interviews that have been conducted already but not yet fully coded that it will become more important.

4.6 Analysis of short texts

The ratings of the answers to the open-ended question are distributed as follows: The vast majority of 219 students have zero points, 83 students achieved a definition with one point, and only 8 students received two points for their answer.

This is the distribution of the five types among the students who filled out the questionnaire: the largest part were assigned the type *no clear picture* (172). The types of *mathematician* (30), *interpreter* (32) and *differentiated picture* (29) are represented by approximately the same number of students. The *creators* are represented somewhat more frequently (48).

5 DISCUSSION

The analysis of the short definition from the questionnaire replicates a finding from literature: Students in their first semester are not able to give a good definition of what computer science is or have only an inadequate conception of the subject. Inadequate in our context means that it is very different from the definitions of experts in the subject. Besides it replicating prior results, there are two reasons that make the result particularly interesting in our specific research context: (1) Our survey was taken shortly after a definition of CS had been presented in lecture. As noted in the related work, our results may no be surprising as students did not have enough opportunities to discover their "own" conception of CS. However, as we know from the other survey results, the majority of the respondents, 247 out of 350, had computer science in school. So in theory they should have had learning opportunities to shape their conception about CS. (2) We set our standard to get a one point score very low. The threshold to get a two point score was not much higher: it was essentially sufficient to mention three obvious concepts from the field in the definition, such as *science*, *computers* and *algorithms*. Despite this low standard, about two-thirds of the respondents could not provide such a definition.

We were able to identify five types of students in our qualitative work. The quantitative approach shows that our type system is saturated for our population. We were able to match all responses to a type and all types appeared multiple times. Neither the appearance of the creator nor that of the mathematician are surprising and correspond well with previous research. The creator type aligns well with Hewner's programming view of CS while the mathematician aligns well with the theory view [10]. We can also confirm Hewner's observation that this view is the most academic: for this type it is quite natural to see computer science as a science that is inherently open to change. The idealist subtype of the creators matches Peters' participation model of "contributing to social endeavours" [18]. What is most astonishing is the existence and high proportion of students who can be attributed to the interpreter type that have no direct correspondence in the previous research cited above. Their humanizing idea of how the computer works is, however, strongly reminiscent of ideas of the super bug [16]. Hewner's so called *Broad View* we have taken as the type of the *Differentiated picture*. This leaves the students who have no clear idea about CS that form the majority both in our qualitiatve and in our quantitative approach. They have been hinted at in [6]. What may explain their lack in other studies is the timing of the interviews: Our interviews were conducted in the third and fourth week after the beginning of the semester. Hence, our focus was on the first weeks of studies in order to capture the students' actual prior understanding. We assume that the mathematicians will have it easier in the course of their studies, as this type encounters relatively few problems in the course of their studies because math and theory – typical stumbling blocks for students in the field [1, 4, 24] – is a natural part of computer science for them. Since the creators' ideas often go hand in hand with the expectation of an engineering education in which programming is the focus, the high proportion of mathematics and theory in typical CS curricula may act as a deterrent for this type. Having no clear conception is not initially problematic, as it can also lead to a greater openness for the subject's content. In other words, where there are no expectations, no expectations can be disappointed. Whether the lack of clear ideas really goes hand in hand with openness will have to be investigated further. The translators, however, may end up struggling: Since this idea is based on both a questionable idea of how the computer works and an inadequate idea of what computer science actually is, it is very likely that students who can be attributed to this type will have a hard time in the course of their studies.

Our codings also show traces of another type, the "technician". However, in our database from the interviews and answers to the open-ended question we could not assign this type clearly to any participant. Because of the codings underlying the type, we can describe her as someone for whom computer science is primarily defined by its technical side, i.e. hardware, networks, etc. To some degree, it is astonishing that this type does not occur, since the "hardware tinkerer" is a common cliché of the computer scientist [12]. Perhaps the absence of this type is the downside of the existence of the interpreter: since most beginners nowadays grew up with mobile phones and laptops that hide their inner workings, they are less familiar with technology and develop false ideas about the actual way the computer works.

5.1 Limitations

It is to be questioned whether the use of a free text question is a suitable means to query such a complex topic as the definition of what computer science is. As the results are backed by the insights from our interviews, we nevertheless chose to include the quantitative view that it allows.

Also, the approach of identifying types based on codings of the interviews necessarily disregards some aspects and codes in order to form a coherent idea of the types. As hinted above, ideas of the translators have been identified in more interviews than afterwards were taken to be from "translators" while traces of technicians were identified but did not lead to a description of this type. Based on the results and their overlap with previous research, we assume, however, that the types presented here form a solid basis for future research.

6 FUTURE WORK AND CONCLUSION

In this paper we have summarized the first results of our research on the understanding of computer science and how this understanding changes in the course of studies. With regard to our first research question, we were able to identify five different types (*Creator*, *Mathematician*, *Interpreter*, *No clear picture*, *Differentiated picture*) of understanding of computer science in beginners. The distribution of types shows a high percentage of beginners with *No clear picture*. Noteworthy is the *Interpreter* type that has no correspondence in prior research results.

For the second research question, we could reproduce three types of students – the *Creator*, the *Mathematician* and the *Differentiated picture* – that have correspondences to prior research results and the observation repeatedly mentioned in literature that first-year students have a poor picture of what computer science actually is. This is true even though the students had been in contact with computers since their earliest youth, sometimes had many years of computer science at school and decided to study computer science.

For us, this typification is the starting point for further research, which on the one hand observes the development of the individual types during their studies, but also examines their peculiarities. Particularly in the case of rather questionable views, such as those of the interpreter, we are interested in whether this type encounters particular difficulties how these may be alleviated as early as possible. We have conducted further interviews with students, but also with lecturers and teaching assistants. The first impression regarding the students is that although the students have completed a semester with three lectures – on programming, computer systems and mathematics – it has not led to a substantial change in their conception of computer science. In some cases, to the contrary, they have become more uncertain and the new input cannot be meaningfully integrated into the existing picture. Therefore we are investigating how the topics of the first semester(s) can be taught more coherently.

REFERENCES

[1] Theresa Beaubouef and John Mason. 2005. Why the high attrition rate for computer science students: some thoughts and observations. *ACM SIGCSE Bulletin* 37, 2 (2005), 103–106.

[2] Maureen Biggers, Anne Brauer, and Tuba Yilmaz. 2008. Student perceptions of computer science: a retention study comparing graduating seniors with cs leavers. In *ACM SIGCSE Bulletin*, Vol. 40. ACM, 402–406.

[3] Manfred Broy. 1998. *Informatik Eine grundlegende Einführung: Band 1: Programmierung und Rechnerstrukturen.* Springer-Verlag, Berlin.

[4] DF Butcher and WA Muth. 1985. Predicting performance in an introductory computer science course. *Commun. ACM* 28, 3 (1985), 263–268.

[5] P. J. Denning, D. E. Comer, D. Gries, M. C. Mulder, A. Tucker, A. J. Turner, and P. R. Young. 1989. Computing as a discipline. *Commun. ACM* 22, 2 (1989), 63–70.

[6] Alexandra Funke, Marc Berges, and Peter Hubwieser. 2016. Different perceptions of computer science. In *2016 International Conference on Learning and Teaching in Computing and Engineering (LaTICE).* IEEE, 14–18.

[7] Gesellschaft für Informatik. 2006. Was ist Informatik – Unser Positionspapier. , 40 pages.

[8] Michail Giannakos, Ilias Pappas, Maria Letizia Jaccheri, and Demetrios Sampson. 2017. Understanding student retention in computer science education: The role of environment, gains, barriers and usefulness. *Education and Information Technologies* 22, 5 (9 2017), 2365–2382.

[9] Tony Greening. 1998. Computer Science: Through the Eyes of Potential Students. In *Proceedings of the 3rd Australasian Conference on Computer Science Education (ACSE '98).* ACM, New York, NY, USA, 145–154.

[10] Michael Hewner. 2013. Undergraduate conceptions of the field of computer science. In *Proceedings of the ninth annual international ACM conference on International computing education research.* ACM, San Diego, 107–114.

[11] Andrew Hoegh and Barbara M. Moskal. 2009. Examining science and engineering students' attitudes toward computer science. In *2009 39th IEEE Frontiers in Education Conference.* IEEE, San Antonio, W1G.

[12] Maggie Jaglo. 2013. „Hardwarefreaks und Kellerkinder" – Klischeevorstellungen über Informatik und die Auseinandersetzung der Studierenden damit. In *Sonderheft: Weltbilder in der Informatik: Sichtweisen auf Profession, Studium, Genderaspekte und Verantwortung*, Gesellschaft für Informatik GI (Ed.). Informatik-Spektrum, Vol. 36(3). 274–277.

[13] Maria Knobelsdorf and Carsten Schulte. 2005. Computer biographies-a biographical research perspective on computer usage and attitudes towards informatics. In *Proceedings of the 5th Annual Finnish/Baltic Sea Conference on Computer Science Education, Koli*, Vol. 2006.

[14] Susanne Maaß and Heike Wiesner. 2006. Programmieren, Mathe und ein bisschen Hardware … Wen lockt dies Bild der Informatik? *Informatik-Spektrum* 29, 2 (2006), 125–132.

[15] Philipp Mayring. 2014. *Qualitative content analysis: theoretical foundation, basic procedures and software solution.* AUT.

[16] Roy D Pea. 1986. Language-independent conceptual "bugs" in novice programming. *Journal of educational computing research* 2, 1 (1986), 25–36.

[17] Anne-Kathrin Peters. 2017. *Learning Computing at University: Participation and Identity: A Longitudinal Study.* Ph.D. Dissertation. Acta Universitatis Upsaliensis.

[18] Anne-Kathrin Peters. 2018. Student' Experience of Participation in a Discipline: A Longitudinal Study of Computer Science and IT Engineering Students. *ACM Trans. Comput. Educ.* 19, 1, Article 5 (Sept. 2018), 28 pages.

[19] Rheinisch-Westfälische Technische Hochschule Aachen. 2001. Computer Science B.Sc. - RWTH AACHEN UNIVERSITY - English.

[20] Yannick Schneider and Andreas Mühling. 2017. Das Konzept Nature of Computer Science. In *Informatische Bildung zum Verstehen und Gestalten der digitalen Welt, Proceedings der 17. GI-Fachtagung Informatik und Schule, INFOS 2017, 13.-15. September 2017, Oldenburg.* 123–126. https://dl.gi.de/20.500.12116/4306

[21] Carsten Schulte and Lea Budde. 2018. A Framework for Computing Education: Hybrid Interaction System: The need for a bigger picture in computing education. In *Proceedings of the 18th Koli Calling International Conference on Computing Education Research.* ACM, 12.

[22] Carsten Schulte and Maria Knobelsdorf. 2007. Attitudes towards computer science-computing experiences as a starting point and barrier to computer science. In *Proceedings of the third international workshop on Computing education research.* ACM, 27–38.

[23] Allen Tucker. 2003. *A Model Curriculum for K–12 Computer Science: Final Report of the ACM K–12 Task Force Curriculum Committee.* Technical Report. ACM, New York City.

[24] Brenda Cantwell Wilson and Sharon Shrock. 2001. Contributing to success in an introductory computer science course: a study of twelve factors. In *ACM SIGCSE Bulletin*, Vol. 33. ACM, 184–188.

A.2 Beginning Students' Conceptions of CS: The Effect of the First Semester

Gregor Große-Bölting, Yannick Schneider, and Andreas Mühling. 2020. Beginning Students' Conceptions of Computer Science: The Effect of the First Semester. In Proceedings of the 2020 International Conference on Learning and Teaching in Computing and Engineering (LaTICE). In Publication.

Beginning Students' Conceptions of CS: The Effect of the First Semester

Gregor Große-Bölting
Department of Computer Science
Kiel University
Kiel, Germany
ggb@informatik.uni-kiel.de

Yannick Schneider
Department of Computer Science
Kiel University
Kiel, Germany
ysc@informatik.uni-kiel.de

Andreas Mühling
Department of Computer Science
Kiel University
Kiel, Germany
andreas.muehling@informatik.uni-kiel.de

***Abstract*—Prior research shows that having a correct idea about the chosen field of study is important for motivation and academic success of beginning students. In the same vein, we already know that beginning CS students often lack such a correct idea. We were interested in identifying how this idea develops over the course of the first semester in order to see how it may impact drop-outs, for example, and to identify how the courses of the first semester may be structured in order to help developing a correct idea about CS. We interviewed fourteen students at the beginning and nine at the end of their first semester and examined how their initial ideas about what computer science is developed over the course of the semester. The aim was to find out how their ideas develop, what formative experiences trigger a development, and which of their courses play a pivotal role in this process. As a result, we were able to identify different types of formative events, develop a model to describe the development, and make proposals for action that may lead to an improvement in teaching.**

***Index Terms*—Student conceptions, conceptual change, beginning students, drop out, qualitative analysis**

I. INTRODUCTION

Students who study computer science[1] enter university with very different ideas about the subject. Some of the ideas are simply vague [1], while others differ greatly from professional ideas of the subject [2]. It has already been investigated which computer science images are formed during the course of study [3], [4], but to the best of our knowledge a fine-grained analysis, especially of the initial phase of study, is not yet available. We would like to close this gap with this research.

This detailed investigation of the concepts of computer science beginners is particularly useful because existing research shows that misconceptions prevent students from enrolling in the subject at all and have a negative influence on success as well as motivation [3], [5], [6]. The number of students who drop out of computer science is generally assumed to be relatively high [7]: Recent statistics show that about forty percent of computer science students in Germany do not finish their study. Also, forty percent of engineering, mathematics and science students (which encompasses computer science) drop out during their freshman year [8].

Therefore it makes sense to deal with wrong and unclear conceptions of computer science as one reason for drop-out. Our aim is to find possibilities to pick up the students' ideas in teaching and to enable them to start their studies in a way that is as satisfying for them as possible. We deem it important to survey computer science students' perceptions of the field and investigate how these perceptions evolve during their studies. In this paper we answer the following research questions:

R1 What leads to changes in the concepts, i.e. what are formative experiences?
R2 What influence do the individual lectures (subject areas) have, with which the students are confronted in the first semester?
R3 Are there any changes in the direction of a differentiated picture of computer science?

Regarding the third research question, our research is based on the assumption that it is desirable that students achieve a differentiated picture of computer science. A broad knowledge and an overview of the different subject areas of the field are necessary in order to be prepared for all requirements both in academic and industrial workplaces. For this reason, the course structure at Kiel University is designed in such a way that the students receive an overview of the field of computer science during the Bachelor's program, which prepares them to deepen their knowledge in a self-chosen subject area in a subsequent Master's program.

II. RELATED WORK

A. *Measuring students' perceptions of computer science*

The understanding of the subject and its influence on the success of students has previously been examined by a number of authors. Hewner [3] conducted 37 interviews with students and advisors and analyzed them with Grounded Theory. He interviewed students from three different colleges throughout their studies and was able to identify three different views of the students on computer science, which he calls *Theory View*, *Programming View* and *Broad View*. The *Theory View* is guided by a mathematical or theoretical understanding, while the *Programming View* focuses on the activity of programming. The *Broad View* understands computer science as a science composed of different disciplines (e.g. theory, robotics, programming).

---

[1] Our research was conducted in a German-speaking area, where the term "Informatik" is used. It is sometimes translated as "informatics" but we chose to keep the more frequently used "computer science" (CS) and treat the two terms as equivalent

Knobelsdorf and Schulte used a biographical method to assess perspectives on the use of computers and attitudes towards computer science [9]. In this way, they collected self-images and world-images in addition to ideas on the field [10]. Their result is that, initially, there is a variety of different ideas [9], but these can be divided into three broad areas: Those of *use*, *professional use* and *design*. The authors express the assumption that a transition must happen from *use* to *design* during the course of study, but this step must be taken by the learners themselves: It does not help students to give an explanation of what computer science is, students have to experience it for themselves; world-image, self-image and habits are too closely interwoven [10].

Biggers, Brauer and Yilmaz [11] investigated why computer science students leave the field. They sent questionnaires with open and closed questions to students and graduates of Georgia Tech and divided the answers into students who stayed in computer science for their major ("Stayer") and those who changed ("Leaver"). One of their open-ended questions was: "How do you define computer science/computing?" One result of their analysis was that Leavers often refer passively to the field and often refer negatively to the working conditions and environment. It was particularly striking for the authors that graduates often refer in their definitions to what computer science is *not*. For example, it is neither information technology, nor the repair of hardware, nor just coding. Exactly these definitions, however, were often found among the Leavers. Answers related to the efficient solving of problems or to the fact that computer science is a mixture of topics related to hardware and software were found with about the same frequency in both groups.

Kinnunen, Marttila-Kontio and Pesonen [12] analyzed beginning students' perceptions of computer science with the intention to provide scaffolding for students to increase their success rates. Like KOI, their survey also involved an entire cohort of two Finnish universities. Aside of their interest and expectations, students were also asked to provide a definition of programming, however the evaluations of these definitions have never been published.

B. Measuring students' progression during their first term

Funke, Berges and Hubwieser [13] surveyed 217 students in a biographical study to investigate gender-specific differences in the perception of computer science. Their survey was conducted at the beginning of an introductory course in computer science and contained two items that were answered as free text and evaluated using Mayring's qualitative content analysis [14]. They were able to identify different perceptions of computer science in relation to gender. The authors also note that many of the participants in their study seem to have no idea either of how computers work or of typical computer science topics, even though the majority of the participants had contact with computers since their earliest youth. Both Funke et al. and Hewner refer to Greening [1], who found that the majority of the students he interviewed in his study were not able to give a meaningful definition of computer science.

Sheard et al. [15] analyzed the impact of students' expectations, interest and experiences on their performance. The study found significant correlations between programming experience and success in programming classes and between unclear expectations and high dropout probabilities. However, no statistically significant relationship was found between students' interests and expectations and retention rates.

Peters [4], [16] investigated in a longitudinal survey how students experience participation in computer science and how they identify as computer scientists. First, 120 essays written by students enrolled in an information technology and computer science program were collected. Then, over the following three years, 61 interviews were conducted with these students regarding their choice of study, future career, study experiences, and perception of computer science as a discipline. Peters discovered seven different ways of student participation in computer science: using, learning about technology, creating, problem solving, problem solving for others, creating new knowledge and contributing to social endeavours [4].

III. CONTEXT OF THE RESEARCH

The results presented here are part of a longitudinal research project (Kohortenbefragung Informatik, subsequently referred to as KOI) that investigates how a cohort of students of computer science at Kiel University develops throughout the first three years of their Bachelor's program in order to find out how local study conditions can be improved.

In prior research [2], we examined beginning CS students' perception of the field of their study when entering university. Data was collected in the form of interviews and a free-text question in a survey. We identified five different types of perceptions of CS based on interviews with 14 students and thereby confirmed results of earlier research [3], [4] regarding common perceptions of computer science. We labelled the types *Creator*, *Mathematician*, *Interpreter*, *Differentiated picture* and *No clear picture* (see section VI, fig. 1). The free-text answers prompted students to provide a personal definition of computer science. We mapped these definitions to the five identified types in order to check for different conceptions that did not appear in the interviews and to get an idea about the quantitative distribution of the types. Our examination indicated that all answers could be mapped to one of the five types and the vast majority with 219 out of 311 students who had provided definitions showed no clear picture of computer science. This is in line with earlier research [1].

Three of these types match - maybe coincidentally - different subdisciplines of computer science which are commonly identified in the German-speaking literature [17]: The *Creator* represents "practical computer science" and is mainly interested in solving actual problems by means of programming. The *Mathematician* represents "theoretical computer science" and perceives computer science entirely as a subdiscipline of applied mathematics. The *Interpreter* represents "applied computer science" and recognizes the computer as a mere instrument on the inner workings of which he has no influence.

The fourth type, dubbed *differentiated picture*, unites the first three types into a holistic view of computer science. Finally, there are students who displayed *No clear picture* of computer science at all. Interestingly, while there is an additional subdiscipline of "technical computer science" that focuses more on the hardware and engineering aspects of CS, a corresponding type ("technician") could not be identified in our cohort, even though other studies report on this (stereo-)type [18].

In the first semester at the Kiel University, all students who enroll for the bachelor program as CS-majors attend three courses:

Programming	The course "Introduction to object-oriented programming" covers the basics of Java starting with control flow and then moving on to object-orientation.
Maths	The course "Mathematics for computer science" covers basic mathematic notation and proofs.
CompSys	The course "Computer systems" deals with the design of digital circuits and the inner workings of computers, including programming in assembly language.

Thus, the first semester of the study covers at least the two major areas of computer science, practical and technical computer science, and gives a foretaste of theoretical computer science with the math lecture. All interview participants attended all three courses, although not all of them successfully completed them.

IV. METHODOLOGY

The 14 students mentioned above were selected based on their response to the open-ended text question in the questionnaire, also mentioned above. Of the 311 responses, 125 students indicated a willingness to take part in interviews and we selected from those 125 following the idea of maximum variation sampling. Taking part in the interviews was rewarded with a small gift voucher.

The 14 students were invited for a second round of interviews at the end of their first semester and 9 students followed this invitation. For the research presented here, we are using the interviews for those 9 students at the end of their first term together with their corresponding interviews at the beginning of the term. All interviews were conducted in a semi-structured setting by the first author and recorded. Participants were offered to be interviewed anonymously, using only a pseudonym, however none of the students chose to do so.

The interviews lasted between 14 and 46 minutes, with a mean interview length of 28 minutes (standard deviation: 10 minutes). Five of the participants were female, four were male. The participants were also a mixed group with regard to their background: Three have already successfully completed another course of study, while for two it is the second after an unsuccessful other course of study. Four had just finished school and were beginning their first course of study.

The first interview touched upon three subject areas: existing ideas of computer science, previous learning routine and overall changes in students' lives due to the start of their studies. With regard to the questions on the ideas of computer science, the interviewees were first given their own short definition from the survey. They were then asked to elaborate on this definition. Depending on the answer, a number of follow-up questions were asked. One area that was always taken into account in the questions was their specific understanding of data and information and the interrelationship of these two concepts with computer science.

The second interview also touched upon three subject areas: changes in the ideas of computer science, success and satisfaction with their learning routine and overall satisfaction with their first semester. The students were first asked how their conception of computer science changed in the course of the semester. On the basis of the answers, they were asked various follow up questions. Among other things, they were always asked about special events in the course of their studies that led to an insight or a conscious change in their computer science perception.

For the analysis, one of the authors transcribed the audio recordings and summarized for each participant the passages from both interviews that are important for the computer science perception into one document. Two authors independently coded the first three interviews and then compared their results. Beforehand, a rough framework had been discussed to be used for coding. This framework was then refined and applied to the other interviews. All interviews were coded following Mayring's approach of qualitative content analysis [14]. In particular, the focus was on capturing the concept before and after, as well as the reasons that led to a change in the concept. Finally, both coders compared their results. There was only a disagreement regarding one interview, but this could be resolved by discussion and joint review of the sources.

V. RESULTS

In the following we describe the results of the analysis in detail and according to the research questions. The participants are identified by letters from A to I. Table I summarizes our results by describing the computer science perceptions of the interview participants at the beginning and the end of the semester as well as the most important lectures (in regard to fostering a change of perception) and formative experiences.

A. Formative experiences

The participants were asked about events or experiences that had a lasting impact on their perception of computer science. Of the participants, 7 out of 9 described such an event either upon request or in connection with other questions.

For A and H, working with a model processor in the course "CompSys" was a formative experience. For them, the processor brought together different areas of knowledge, not only within the Computer Systems course, but also beyond

TABLE I
Overview of interview results

	Start (types as described in [2])	**End**	**Important Lecture**	**Formative Experience**
A	Creator (Idealist)	idealism takes a back seat, math becomes more important	CompSys, Math	Exemplary processor
B	No clear picture	remains unclear	–	–
C	Mathematician	development towards programming	Programming	Programming project
D	Interpreter	no settled perception, but development towards Computer Systems	CompSys	Upgrading computer
E	Differentiated picture	math becomes more important	Math, (CompSys, Programming)	Additional, voluntary exercises
F	Creator	computer science as science of information processing	(Math)	Programming project
G	Interpreter	development towards Computer Systems	CompSys	–
H	No clear picture	no settled perception, but development towards math and Computer Systems	CompSys	Exemplary processor
I	Mathematician	supplement, but no integration of Computer Systems	(CompSys, Programming)	Additional, voluntary exercises

that, thus creating a link between the various areas of computer science:

> Otherwise it was always quite superficial. We just learned how to convert a decimal number into binary and all that. And then afterwards I had the feeling, when we came to the topic ALU[2] everything we did before meshed a bit and it was a bit more understandable, why we do all this and, yes. For the first time I had the feeling to understand the lecture in general (laughs), why we were shown so many different things, yes. (H)

For D, a hardware experience was formative: When upgrading her own computer, the participant not only understood the specification of the new component, but was also able to explain it in detail to her fiancé, thus creating a moment of self-efficacy:

> [I] was thinking about giving my PC a new hard drive, just read NAND circuit and thought: Hey, I know what it is (laughs). [...] Yes, I thought it was funny, I told it directly to my fiancé, who didn't understand it at all. But, yes. (D)

For E and I, additional, voluntary exercises in the programming course were a formative experience. These additional exercises are intended for students who want to learn more and are usually intended for students with very good prior programming abilities. Both students describe that they felt like computer scientists when working on these tasks. For I, it was important that the exercises usually had a mathematical background, so that she found her conception of "computer science as applied mathematics" particularly clearly realised here. For E, who already had extensive prior experience in programming, it was important to see that he could do the advanced exercises, so that he could feel like a *real* programmer.

> I think the bonus exercises for programming, where you could at least use something that you have learnt but also somethings that you didn't learn so far, so that you can live out your creativity as a computer scientist, if you like. That you actually have to come up with ideas, find solutions. This game: You had a solution, but maybe there is a better solution and then you think about it and write something down. That's like a puzzle, if you like. Which is also quite nice. Yes. I felt comfortable there and generally also to be surrounded with people who are also interested in it. (E)

E also describes that an important experience for him was meeting people who shared the same interests as he did.

For both C and F, programming projects were formative experiences, albeit with a different background: For C it was the project that was a mandatory part of the programming course. She was able to see what the different concepts presented in lectures and tutorials really mean and what use they have in an application. F, however, started a programming project on his own, not directly related to his studies. It is the result of an everyday situation, a conversation with a family member:

> Uh, I talked to my mother about how we felt that on the radio … there were only three different songs playing. And then my mother said: Yeah, how about you write a program that looks at which songs are actually played during the day. And I'm like: Hey, that's a great idea. (F)

For F, this was an important experience because it made him feel like a computer scientist.

---

[2] i. e. short for *arithmetic logic unit*.

> Yeah, others might have written it all out by hand. . . . [A]nd I just wrote a program that does that and that's what computer science is all about for me.

B. Influence of the individual lectures

Students were asked how the different lectures in the first semester relate to another, i. e. how they are thematically linked. The single most important lecture for the linking of the content of the first semester was Computer Systems (CompSys) for A, D, G and H. For E and I CompSys still plays a vital role for their understanding. The reason for the importance of this particular lecture is that it bridges between Programming and Mathematics: Both are applied in CompSys and the relevance is clear when building a computer. None of the participants had prior knowledge in this field, so it was in some sense the most surprising lecture as well. The role it plays for the students' understanding is well displayed by the following quote by participant G:

> And CompSys is just, it made me understand how big it actually is, like, I never could relate to how big that actually is, how you, how you work with memory, how you work memory-efficient, because I — for me there was just enough power available. I mean, you just can't imagine that there are three billion transistors on a CPU, you can't imagine. Because you have a single machine that only adds two numbers. And then you see: Okay, for this I need so many [transistors] and that's so many more times this and then you can relate to it better, get an idea of how it's structured. And that also helped me with my programming in general [...]. (G)

Most respondents were aware that mathematics at university level is very difficult. However, some respondents were particularly impressed by how different it is from what they knew from school. Especially A, E and F were influenced by the first semester's math lecture and forced to change their conception of computer science, so that mathematics takes a much larger share of it. For F it was a negative experience: he does not see the sense in having this lecture as a mandatory part of computer science:

> I was aware that there was a lot of mathematics involved, but now that you are experiencing it for yourself, I would say that you realize that you are only realizing how much it actually is and how difficult it actually is to get it all together. (F)

It is worth mentioning at this point that G considers mathematics to be a useless part of the study program: He is firmly convinced that the lecture is only intended as a barrier in order to filter out students and thus create an artificial restriction on access to studies.

Only C considered Programming to be the most influential lecture. For I and E it plays a minor, but integral role in the development of their computer science perception. The importance for C is related to low previous knowledge in programming, with at the same time great experiences of success. For I and E, the importance results primarily from the formative experiences described above.

C. Change in perceptions

Directly asked for changes, four of the interviewees stated that they had not noticed any change in their computer science perception, three had noticed a small, two a large change. Nonetheless the analysis of the interviews shows a clearly perceptible (i.e. describable) change in the conception of computer science for eight out of nine participants.

This change consists primarily of the fact that students are trying to assimilate their prior conception in a way to integrate everything that they experienced and learned in their first semester but are struggling to do so. A "textbook example" for this situation is the following excerpt, in which the participant, who is convinced that computer science is essentially applied mathematics, tries to integrate her newly acquired knowledge in CompSys into this idea:

> **Q:** Was there anything else that added to this idea, or is it limited to that?
>
> **A:** No, so there's definitely something additional. Above all, of course, yes, the hardware stuff to some degree, networking. [...]
>
> **Q:** But is that something that can be organically added to the idea of "basically everything is applied mathematics", or is it, is there such a thing as...
>
> **A:** No, that's an extra thing, I'd say.
>
> **Q:** Ok?
>
> **A:** Well it is what processes the mathematical things, like, this outer shell so to speak and that what happens on the inside, the processes, that's very mathematical. But what's on the outside, in these things where the processes run and how the processes are organized, that's definitely an extra thing, something that's been added. (I)

Three groups can be distinguished in their efforts to integrate the newly acquired knowledge, with varying success: A, I and E try to combine all newly acquired knowledge into a uniform picture, but still fail to do so at the present time. C, F, G and D try to integrate individual courses into their image, but reject others as part of computer science or are unsure about their role. B and H, on the other hand, show no awareness of integrating the material acquired during their studies into their image. It can be stated for all respondents that at the time of the second interview, i.e. at the end of the first semester, they do not yet have a consolidated picture of what computer science is.

According to our type system shortly described above, we tried to assign one of the five types to each interviewee at the end of the semester. This was only possible in the case of B because there was *No clear picture*. On the other hand, for the other participants – as shown before – a change from a known type to one or more other known types took place, i.e. mixed or subtypes formed. Therefore, in our existing type system (from previous research, s. III) we cannot assign types to eight out of nine participants.

VI. DISCUSSION

Taking a look back at the initial types identified at the beginning of the course, we can match two types as being somewhat closely related to one of the courses: The type of the *Mathematician* relates to the course in mathematics and the *Creator* relates to the course in programming. By “relate” we mean that a person identifying as this particular type will most probably be intrinsically interested in this particular course and the topics of this particular course will align well with her existing notion of “computer science”. Fig. 1 visualizes this relation and also places other types in relation to the courses: The *Interpreter* and the *No clear picture* types are arranged outside the circles because they do not represent valid computer science images [19]. The *Technician* type is listed because it – as explained above – was reported on in prior studies and because it corresponds well with the contents of the Computer Systems lecture. However, because we were unable to show its existence in our cohort, it is shown in brackets. In the following we will refer to this figure in order to explain the development of the computer science concepts of first-year students.

With regard to the development of computer science concepts, it can be seen that even after one semester of full-time study, it has not yet been possible for beginning students to form a consolidated conception, however a development of the conceptions can be observed for almost all of the students that we interviewed. The slow evolutionary development of concepts as explained here corresponds to the description of conceptual change as presented by representatives of knowledge-as-elements (i. e. “Knowledge in Pieces”, short KiP) perspectives [20]. Also the observed coexistence of competing, not fully integrated conceptions, as documented for example for interviewee I, can be well explained by KiP [21].

The development of conceptions is in every single case a positive one: representatives of the *Interpreter* type and the *No clear picture* have developed in a way that now better corresponds to an accepted or established image of computer science [19], such as that of the *Creator*. They move, so to speak, from the outside into one of the circles of Fig. 1. Students who already had an established picture developed in such a way that they now have a broader idea of computer science (they move into the direction of one of the intersections of Fig. 1). We hope that all students will gradually converge towards the *Differentiated picture* over their course of study.

The first semester shows that the lecture Computer Systems plays an important role. This is probably mainly due to three reasons: Firstly, it is the lecture to which very few students bring previous knowledge. While most beginners have had contact with math or programming, the handling of hardware and the peculiarities of computer systems is something new. This makes it – and this was mentioned several times in the interviews – the most surprising lecture. Secondly, it is the lecture that most likely contradicts the interpreter image of computer science. The deep understanding of computers that the lecture conveys, hardly allows the assumption that computers are anything other than unintelligible machines (like the misconception of a *super bug* [22]). This shakes the *Interpreter* conception. The third reason is the mediating character of the lecture, as it is in some sense located between math and programming: On the one hand it needs mathematical understanding to illustrate many basic concepts and on the other hand it shows how from simple components the basics of programming are followed and finally even something as complex as assembler code.

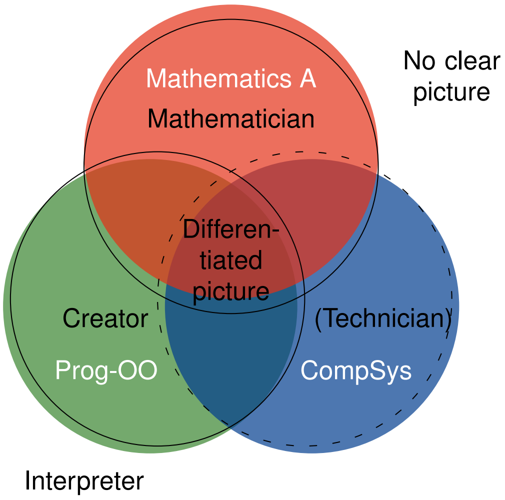


Fig. 1. The first-semester lectures at Kiel University (colored circles, white labels) and the identified student types from [2] (black circles, black labels). Because the existence of the technician could not be shown, this type is shown with a dashed line.

It is therefore not surprising that the formative experiences described by students were sometimes related to hardware. All of the formative experiences can be divided into two categories: Either they are experiences of events that link different themes or they have created a feeling of self-efficacy and identity as computer scientists. This relates well to previous research by Peters [4].

From these insights, proposals for action can be derived:

Since projects increase self-efficacy and thus lead to self-awareness and identification as computer scientists, they should be consciously integrated into the curriculum. Where this is already the case, students should also be given the opportunity to choose the topic of their project themselves. The example of a self-chosen project described by F clearly shows how the motivation of the students – in this case through pride towards the mother – can be promoted.

Networking the individual courses seems to be an important prerequisite for students to have a holistic and motivating view of their studies. The individual lectures should not be regarded as independent and co-existing. Instead, a content-related and structural networking is to be recommended. This can, for example – as in the example of the formative experiences of

E and I – be achieved by tasks that treat the material of another with the means of that lecture.

The lecture Computer Systems played a special role for the interviewees. On the one hand, this is related to the previously mentioned linking of contents from other subject areas (math and programming). On the other hand, it seems to have something to do with the fact that it tackles a new and unexpected topic for the students. Accordingly, when designing curricula, consideration should not only be given to offering introductory courses that are in some kind familiar to all beginners: Courses that are not yet known to anyone have their place, as long as they are not too difficult and can be easily combined with the other courses.

With regard to the type of change in computer science concepts, it can be stated that students who can be assigned to the *Interpreter* or the *No clear picture* developed in the first semester in such a way that they can now be found within the Venn diagram (fig. 1). Students who have been *Mathematicians* and *Creators* before, developed in a way that they have moved to one of the interfaces with between two conceptions. E, who was the only one who could be assigned to the *Differentiated picture* type, remains differentiated, but deepens his knowledge within.

A. Limitations

An obvious limitation of this research is that the interviews were conducted at only one university. This makes it questionable whether the results can be easily transferred. On the other hand, this makes it possible to examine the influence of individual courses more closely, which would be difficult or even impossible with several universities with different curricula.

VII. CONCLUSION AND FUTURE WORK

In this paper we showed how beginning students' conception of their chosen field of study develop over the first semester. Starting with five different types we observe a change in the students that is fostered partly by the content of the lectures of their first semester and partly by moments of self-efficacy. The students do not yet possess a consolidated picture of computer science but all have moved more towards a differentiated, integrative picture of what encompasses CS.

Based on the findings, a set of suggestions has been made that may be able to trigger a process of accommodation more quickly in order to improve students entry phase into their CS studies. This may partly alleviate the phenomenon of high drop-out rates.

For our further research we are particularly interested in three questions: First, we now have a good insight into the ideas of first-semester students and how they develop. But what about the ideas of the teachers who meet the students in the first semester? Do their ideas differ and what influence does this have? To further explore this question, we have already conducted a series of interviews, which we plan to evaluate together with the material presented here and in our earlier research.

Second, in order to further support the validity of our data, a large-scale quantitative study would be necessary. Such a study would make it possible to say with greater certainty whether the results presented here can be generalised. In this context, the transfer and study at other universities would also be of interest.

Third, the presented results show that at the end of the first semester there is no fixed idea of what computer science is. However, as the research literature presented above suggests, this will happen at some point in the course of the study. Accordingly, it is of great interest to us which experiences and events in our studies contribute to the formation of the image of computer science and, so to speak, allow it to coagulate. Therefore, we will continue to accompany the students in their studies with interviews and questionnaires.

REFERENCES

[1] T. Greening, "Computer science: Through the eyes of potential students," in *Proceedings of the 3rd Australasian Conference on Computer Science Education*, ser. ACSE '98. New York, NY, USA: ACM, 1998, pp. 145–154.

[2] G. Große-Bölting, Y. Schneider, and A. Mühling, "'It's like computers speak a different language': Beginning students' conceptions of computer science," in *Proceedings of the 19th Koli Calling International Conference on Computing Education Research – Koli Calling '19*. ACM Press, 2019.

[3] M. Hewner, "Undergraduate conceptions of the field of computer science," in *Proceedings of the ninth annual international ACM conference on International computing education research*. San Diego: ACM, 2013, pp. 107–114.

[4] A.-K. Peters, "Student' experience of participation in a discipline: A longitudinal study of computer science and it engineering students," *ACM Trans. Comput. Educ.*, vol. 19, no. 1, pp. 5:1–5:28, Sep. 2018.

[5] A. Hoegh and B. M. Moskal, "Examining science and engineering students' attitudes toward computer science," in *2009 39th IEEE Frontiers in Education Conference*. San Antonio: IEEE, 2009, p. W1G.

[6] S. Maaß and H. Wiesner, "Programmieren, Mathe und ein bisschen Hardware … Wen lockt dies Bild der Informatik?" *Informatik-Spektrum*, vol. 29, no. 2, pp. 125–132, 2006.

[7] M. Giannakos, I. Pappas, M. L. Jaccheri, and D. Sampson, "Understanding student retention in computer science education: The role of environment, gains, barriers and usefulness," *Education and Information Technologies*, vol. 22, no. 5, pp. 2365–2382, 9 2017.

[8] Autorengruppe Bildungsberichterstattung, *Bildung in Deutschland 2018*. wbv Media GmbH, 2018. [Online]. Available: https://www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Bildung-Forschung-Kultur/Bildungsstand/Publikationen/Downloads-Bildungsstand/bildung-deutschland-5210001189004.html

[9] M. Knobelsdorf and C. Schulte, "Computer biographies-a biographical research perspective on computer usage and attitudes towards informatics," in *Proceedings of the 5th Annual Finnish/Baltic Sea Conference on Computer Science Education, Koli*, vol. 2006, 2005.

[10] C. Schulte and M. Knobelsdorf, "Attitudes towards computer science-computing experiences as a starting point and barrier to computer science," in *Proceedings of the third international workshop on Computing education research*. ACM, 2007, pp. 27–38.

[11] M. Biggers, A. Brauer, and T. Yilmaz, "Student perceptions of computer science: a retention study comparing graduating seniors with cs leavers," in *ACM SIGCSE Bulletin*, vol. 40, no. 1. ACM, 2008, pp. 402–406.

[12] P. Kinnunen, M. Marttila-Kontio, and E. Pesonen, "Getting to know computer science freshmen," in *Proceedings of the 13th Koli Calling International Conference on Computing Education Research – Koli Calling '13*. ACM Press, 2013.

[13] A. Funke, M. Berges, and P. Hubwieser, "Different perceptions of computer science," in *2016 International Conference on Learning and Teaching in Computing and Engineering (LaTiCE)*. IEEE, 2016, pp. 14–18.

[14] P. Mayring, *Qualitative content analysis: theoretical foundation, basic procedures and software solution*. AUT, 2014.

[15] J. Sheard, A. Carbone, S. Markham, A. J. Hurst, D. Casey, and C. Avram, “Performance and progression of first year ict students,” in *Proceedings of the Tenth Conference on Australasian Computing Education - Volume 78*, ser. ACE ’08. Darlinghurst, Australia, Australia: Australian Computer Society, Inc., 2008, pp. 119–127. [Online]. Available: http://dl.acm.org/citation.cfm?id=1379249.1379261

[16] A.-K. Peters, “Learning computing at university: Participation and identity: A longitudinal study,” Ph.D. dissertation, Acta Universitatis Upsaliensis, 2017.

[17] H. Herold, B. Lurz, J. Wohlrab, and M. Hopf, *Grundlagen der Informatik*. Hallbergmoos: Pearson, 2017.

[18] M. Jaglo, “„Hardwarefreaks und Kellerkinder“ – Klischeevorstellungen über Informatik und die Auseinandersetzung der Studierenden damit,” in *Sonderheft: Weltbilder in der Informatik: Sichtweisen auf Profession, Studium, Genderaspekte und Verantwortung*, ser. Informatik-Spektrum, G. GI, Ed., 2013, vol. 36(3), pp. 274–277.

[19] Y. Schneider and A. Mühling, “Das Konzept Nature of Computer Science,” in *Informatische Bildung zum Verstehen und Gestalten der digitalen Welt, Proceedings der 17. GI-Fachtagung Informatik und Schule, INFOS 2017, 13.-15. September 2017, Oldenburg*, 2017, pp. 123–126. [Online]. Available: https://dl.gi.de/20.500.12116/4306

[20] G. Özdemir and D. B. Clark, “An overview of conceptual change theories.” *Eurasia Journal of Mathematics, Science & Technology Education*, vol. 3, no. 4, 2007.

[21] A. A. DiSessa, “Toward an epistemology of physics,” *Cognition and instruction*, vol. 10, no. 2-3, pp. 105–225, 1993.

[22] R. D. Pea, “Language-independent conceptual “bugs” in novice programming,” *Journal of educational computing research*, vol. 2, no. 1, pp. 25–36, 1986.

A.3 Identity in K-12 Computer Education Research: A Systematic Literature Review

Gregor Große-Bölting, Dietrich Gerstenberger, Lara Gildehaus, Andreas Mühling, and Carsten Schulte. 2021. Identity in K-12 Computer Education Research: A Systematic Literature Review. In Proceedings of the 17th ACM Conference on International Computing Education Research (ICER 2021). Association for Computing Machinery, New York, NY, USA, 169–183. `https://doi.org/10.1145/3446871.3469757`

Note: The page numbering of the following pages is taken from the original publication and does not correspond to the numbering of this thesis. The page numbering returns to the numbering of the thesis after this publication.

Identity in K-12 Computer Education Research: A Systematic Literature Review

Gregor Große-Bölting
Kiel University
Kiel, Germany
ggb@informatik.uni-kiel.de

Dietrich Gerstenberger
Paderborn University
Paderborn, Germany
dietrich.gerstenberger@uni-paderborn.de

Lara Gildehaus
Paderborn University
Paderborn, Germany
lara.gildehaus@math.uni-paderborn.de

Andreas Mühling
Kiel University
Kiel, Germany
andreas.muehling@informatik.uni-kiel.de

Carsten Schulte
Paderborn University
Paderborn, Germany
carsten.schulte@uni-paderborn.de

ABSTRACT

The identity or self-concept of computer scientists has received increasing attention in the computing education research (CER) literature in recent years. Identity is often considered relevant both for initially choosing a path of study and subsequent retention. It is therefore also considered highly relevant for the questions of how to reduce drop-out rates and broadening participation of currently underrepresented groups in computing in higher education. However, as more and more students have eligible or mandatory computing education in their K-12 years, identity may become relevant in this area as well.

In this article, we analyze the use and development of identity in the CER literature with a focus on K-12 education. To do so, we undertook a systematic literature review that identified appropriate publications through both a traditional database search (ACM DL, IEEE Xplore, SpringerLink, ScienceDirect, DBLP, and Google Scholar) as well as an additional forward and backward snowballing process. In total, 31 papers from the years 1997-2020 were identified that address identity in the K-12 CS context.

We summarize key research findings from these articles and develop a category system that demonstrate how and why identity is used in CER in the K-12 context. Our findings suggest that the use of identity in K-12 research needs to be thought of in fundamentally different ways than for higher education. Alongside, we provide evidence that the underlying theory is less fragmented than often claimed and highlight potentials arising from greater networking and discussion of identity research in (K-12) CER.

CCS CONCEPTS

• **Social and professional topics → K-12 education**; **Computational science and engineering education**.

KEYWORDS

computer science education, k-12, theory, identity, self-concept, sense of belonging, systematic literature review

1 INTRODUCTION

Identity has been popular in educational sciences for well over a decade as a means to analyze learning not only from a cognitive, but also from a cultural and social viewpoint [62, 104]. As Flum and Kaplan state in their introduction to a special issue on the topic of identity of *Contemporary Educational Psychology*: “The concept of identity is very widely used. Some may say, overused. It is a key concept in the social sciences in general and a term that captures a variety of nuanced meanings. Identity is described as a ‘heavily burdened’, ‘elusive’ and ‘deeply ambiguous’ term which is, nevertheless, viewed as being ‘indispensable’.” [40]

Unsurprisingly, discipline-based educational research has soon also followed the trend and adapted and extended the theories from the perspective of a specific subject (for mathematics, e.g. [59, 104]). For the two subjects arguably closest to computer science, engineering education [84] and mathematics [28, 99] the large body of work has even been compiled in reviews already. The continuing interest is perhaps unsurprising, as the concept may be the “missing link” [104] that brings together cognitive, affective, social, and cultural characteristics of learners under one theoretical concept and can serve as an explanatory model. In particular, it is also a very flexible theoretical construct that allows looking at problems at many different levels. As Darragh puts it: “Identity is a lens that is adjustable; one can zoom in to the level of interactions between individuals or zoom out to look at the wider socio-political context. We can look at the big picture, that is, at issues of mathematics learning in general. We can look at the experiences of specific groups of people and issues of equity. Or we can look at the individual level and try to understand learners’ relationships with mathematics.” [28]

For computer science education (CSE), the interest has been growing in the same way, as indicated for example in the works of Knobelsdorf [66], Peters [95], or most recently the developed instruments of Mahadeo et al. [73] and Kong & Wang [67]. As an integrative concept, identity has the potential to explain classic problems of CSE, such as diversity and retention [101].

However, for computer science education, a systematic review is still missing even though is would be particularly helpful because of the vagueness of the concept that both Darragh and Radovic point out, and take it as the starting point for their own research [28, 99]. So, while there is a broad discussion in e.g. mathematics about how the concept of identity is applied (again: [28, 99]), this is not yet to be found for computer science.

To this end, we conducted a systematic literature review (SLR), focusing on the overarching question of how the concept of identity is used in CSE and what computer science-specific observations and developments can be observed. The review is conducted separately both for K-12 and for higher education. This paper presents the K-12 part. The reasons for dividing the review into a K-12 and a higher education sub-corpus are explained in section 2.2. In order to provide a comprehensive overview of the use and development of identity in both CSE and computer science education research (CER) we will answer the following research questions:

R1: What theoretical backgrounds are used to define and conceptualize identity?
R2: How is identity used in research and what are the main findings from this research?
R3: Why is the concept of identity used in literature, i.e. what are the supposed effects?

In addition to presenting a category system that has been developed to answer the questions, we will also shortly review the main findings from the articles. This is followed by a discussion regarding the importance of the concept for computer science within the K-12 context.

2 THEORETICAL BACKGROUND

In this section, we will briefly review the theories around the concept of *identity* in general and in specific subjects, as we will refer back to them in our methods and results and describe the special role of identity in the K-12 context.

The concept of identity exists under different terms. For the professional identity as a computer scientist, Rodriguez and Lehman [101] use the term *computing identity*, for the identity of the discipline itself the term *computing disciplinary identity*; this already indicates the subjective and social dimension of the phenomenon. Furthermore, the authors state that the concept *sense of belonging* has a close connection to identity, since it describes the subjective affiliation of an individual to a social group. In addition, there is the term self-concept, which is used in particular in psychological literature [94]. Other concepts similar to identity are referred to as self-image or self-perception. In order to do justice to this conceptual diversity, we have oriented ourselves to the definition of Gee [45]: Identity describes "being recognized as a certain 'kind of person,' in a given context" We attempted to represent everything that applies within the context of this definition in our literature review. We will return to these difficulties of definitions in sections 3 and 5.1.

Not only are there different terms to describe the same or related concepts, but there is also additional jargon provided by different theories in the field: Lave and Wenger [69] introduced the term *community of practice* (CoP) to describe a discipline as a group of people that share practices. Those practices are socially situated, handed down from one generation to another, and part of a living society of like-minded people. A newcomer into a CoP at first only takes part in *legitimate peripheral participation*: "By this we mean to draw attention to the point that learners inevitably participate in communities of practitioners and that the mastery of knowledge and skill requires newcomers to move toward full participation in the sociocultural practices of a community." [69, p. 29] The movement from periphery to center *is* learning and *is* identity development: "[L]earning is not merely a condition of membership, but is itself an evolving form of membership. We conceive of identities as long-term, living relations between persons and their place and participation in communities of practice. Thus identity, knowing, and social membership entail one another." [69, p. 53]

Sfard and Prusak [104] operationalize identity as *narratives*, as "stories about a person": "In concert with the vision of identifying as a discursive activity, we suggest that identities may be defined as collections of stories about persons or, more specifically, as those narratives about individuals that are *reifying*, *endorsable*, and *significant*." By analyzing the narratives that someone tells about themselves to others or that others tell about them and extracting family resemblances from similar narratives, statements about identity are subsequently possible, but they do not neglect the discursive, socially negotiated aspect of the concept.

The high number of different identity theories has led to a grouping of different theories into *theory strands*. For example, the mentioned theories of Lave and Wenger [69], Gee [45], Sfard and Prusak [104] as well as Holland et al. [59], are often grouped under the umbrella term of *sociocultural theories*, since they each focus in particular on the social situatedness of an individual in a (cultural) community. Theories, on the other hand, that conceive of identity as a predominantly subjective character trait are particularly popular in psychology (for example [33, 72, 92, 100]) and are therefore referred to as *psychological theories* [28, 113]. Using Radovic's [99] terms, the psychological conceptualizations can be said to tend to be more *subjective*, *representational*, and *stable*, while the sociocultural ones tend to be more *social*, *enacted*, and *change*-oriented. Accordingly, in the former, identity is mainly operationalized as an *individual attribute*, whereas in the sociocultural theories we find operationalizations of *identity as specific practices* or *narratives*. There are also other theories, like sociological and sociopolitical theories, e.g. Foucault [41] or Butler [18] that focus on power structures and how groups secure their status through the (implicit and explicit) inclusion and exclusion of individual members [113] (also termed identity politics [42]).

Computing disciplinary identities can be considered as a possible conceptualization of computer science. In order to clarify what characterizes this conceptualization, a demarcation from two other possible conceptualizations follows. The consideration is thus not conclusive, but nevertheless illustrates the conceptual peculiarities: In recent years, there have been a number of publications that use the term *computer science (CS) conceptions* [9, 49, 50, 56]. CS conceptions attempt to capture the image that novices have of the discipline and to identify various categories of such images. For example, Große-Bölting et al. [49] identifies five categories: The *Creator*, the *Mathematician*, the *Interpreter*, the one with *no clear picture*, and the one with a *differentiated picture*. The distinction to the concept of identity here is subtle and consists in a stronger worldview orientation: While CS conceptions focus more on the discipline, under the concept of computing identity one would ask about one's understanding of one's role and position, that is, one's self-image, within the subject.

From a more normative point of view, constructs like "computing attitudes" [30] that measure students' attitudes about computer science as a science, might be considered a form of conceptualization of computer science, as they define common traits of what a computer scientist is. This is also true for something like computational thinking that, by the very definition [116], is some form of practice that computer scientists share. In contrast to the terms of identity considered here, both however share a normative and cognitive focus and as such are out of context for the subjective. For a relation between computational thinking and computing identity, however see [67], that also appears in our corpus.

2.1 Identity in related disciplines

Clark and Kajfez [24] review literature on K-12 engineering education. Their analysis groups the studies into formal (more specifically into types of school) and informal settings. The authors summarize that in the K-12 years, hands-on engineering activities contribute to the development of an engineering identity, independent of gender. As a particularly effective intervention, the authors cite addressing students as engineers and applying real engineering processes. They indicate that an early introduction to engineering, especially among underrepresented minorities, will improve the development of an engineering identity.

Morelock [84] also reviews literature on the concept of engineering identity. He reports that a quarter of the papers do not explicitly define the term identity. Two-thirds of the studies report on factors that might have an influence on how a person perceives engineering. Morelock suggests beginning quantitative research to test the relevance of these factors. Additionally, he summarizes specific interventions designed to introduce participants to the field of engineering and gives an overview of the methods used, most of which were based on qualitative research.

Darragh [28] gives an overview of identity research in mathematics education with a specific focus on how the various authors understand the term identity, which theories they reference, and which methods they apply or develop themselves. She discusses two main developments of identity theory based on Erikson [34] and Mead [81], which she refers to as mainly psychological and social perspectives. Following the social perspective she identifies those theories as the most commonly used frameworks on identity: Lave and Wenger [69] understand identity in the social context of communities of practices within their social learning theory. Similarly, Holland et al. [59] refer to identity in so called figured worlds. With a sociocultural approach they conceptualize both positioning by others and finding oneself in the social context. Gee's four types of identities [45], focuses on a natural, institutional, discursive and affinity identity. While those approaches stem from sociocultural backgrounds, the following authors directly developed their theories within mathematics education: Sfard's [104] approach explicitly refers to narratives, in the form of stories that individuals tell about themselves. Boaler and Greeno [10] adapted Holland et al. and Lave and Wengers framework and specified it to mathematics education. Martin [77, 78] focuses on diversity and identity development in mathematics classrooms.

Radovic et al. [99] undertake a systematic literature review on mathematics learners' identity, where the authors aim to identify commonalities in the consideration of identity in their field. They present three dimensions that characterize how identity is conceptually defined in literature: Subjective/social, representational/enacted and change/stability. Furthermore, they identify five "main categories that describe how the literature has implemented these dimensions operationally" [99]. The authors argue that their model can be used prospectively to employ these dimensions in the planning of research projects. It could also be used retrospectively to discuss research results. The authors mention great differences in studies when differentiating between K-12 and higher education. While studies "tended to emphasize practiced and enacted views of identity" at school age, they "tended to focus on representational and subjective forms of identity" in higher education [99]. When summarizing the state of research, instead of analyzing the use and operationalization of theoretical concepts (see [28]), the authors suggest looking at the social/subjective, representative/actual, and change/stability dimensions instead.

2.2 Identity in the K-12 context

In the same vein as reported by Radovic et al. [99], we also expect to find different results for K-12 and higher education, as K-12 education differs from higher education in several fundamental ways:

First, in K-12 there is always a multitude of subjects in parallel, with computer science usually only playing a minor role. Clearly, developing an identity that is specific to computer science will happen differently, if at all. For example, Lave and Wenger [69] assume that the school itself represents a community of practice in which students learn to persist in the school system instead of really becoming part of the community of a discipline.

Second, even if developing a subject-specific identity might not happen in the K-12 years, experiences in school still can serve as the foundation for later identity development [62]. For example, during adolescence and thus during the school years, the ability to reflect is developing, which goes hand in hand with the development of identity [40, 55]: According to Erikson, the formation of an identity is at the same time the result of reflection (as an internal process), as well as of observation (as an external, social-relational process) [34]. Nevertheless, research for the special role of the school context with regard to subsequent identify formation is still a desideratum [68, p. 87].

Third, there still is a connection between K-12 education and higher education in terms of the typical research focii of identity in higher education (as reported below): Retention, dropout prevention and diversity: Even the choice not to take one "feeder course" in school can effectively prevent students from entering STEM studies [85]. Accordingly, positive and identity-forming experiences with STEM in general and with CS in particular are important in school, thus making it a worthwhile research topic on its own.

A systematic literature review by Verhoeven et al. [113] confirms this specific role of schools in identity development. In their analysis of 111 articles on identity development in school (35 of which were from the STEM field), the authors were able to identify sociocultural identity theories (defined here as influenced primarily by Gee, Holland, and Lave) as the most dominant theoretical tradition (55, versus psychosocial (8), psychological (4), and sociological (4); 26 articles without theory). They also distinguish the papers into three broad groups: those that unintentionally influence students' identity development (48) – the authors also refer to this as "hidden curriculum" –, those that intentionally attempt to influence identity development, and those that create the preconditions for indented identity development. We will return to this distinction in the discussion.

Taking all of this together, we decided to split the review into two separate parts. Also, within the K-12 focus, we decided to only focus on learners' identity, rather then teachers' identity as well. Not only are there are too few articles for meaningful evaluation in this area: After the database search with the identity term (see section 3), only three articles [52, 64, 89] could be identified in this area. They also mainly differ in their purpose, focusing on professional identity rather than students identity and using different conceptualizations [99].

3 METHODS

Our systematic literature review (SLR) follows the method introduced by Kitchenham [65], with some adjustments: The SLR was carried out in the above described knowledge that identity is a vague concept that may be understood under various terms. Therefore, even before the search was started, a process for *forward and backward snowballing* [117] was established – as a first adjustment to the classic [65] process. The details can be found in section 3.2. In the course of the search and after evaluation of the first papers, as described in section 3.1 and 3.2, it became clear that there are large intersections between identity and self-concept, self-perception and self-image. Although these are by no means congruent concepts and the relation between the concepts is hard to grasp [5, 6], the search was subsequently extended to include these terms. Since the social understanding of identity emerged as one dominant theory, the concept of *sense of belonging* was also eventually included.

3.1 Database Search Strategy and Selection Process

Since most of the literature in the field of computer science education is published digitally, an automatic search strategy was pursued. For this purpose the databases ACM Digital Library[1], IEEE Xplore[2], SpringerLink[3], ScienceDirect[4], DBLP[5], and Google Scholar[6] were used. The search was carried out on both title and abstract. No restriction was placed on the years of publication to determine if there was indeed a trend in the use of the concept. The following search query was initially used:

```
identity AND (education OR k12 OR students OR cs1 OR
learn) AND ("computer science" OR "computer engineering"
OR informatics OR computing)
```

After the original query proved to be too narrow, the search was extended by the results of the following query:

```
("self-concept" OR "self-image" OR "self-perception" OR
"sense of belonging") AND (education OR k12 OR students
OR cs1 OR learn) AND ("computer science" OR "computer
engineering" OR informatics OR computing)
```

The search results were then exported for further processing. For the SLR we were only interested in full publications from conference proceedings and journals. In order to exclude posters, pure abstracts and panel papers as easily as possible, the list of publications were filtered by number of pages, excluding publications with less than 4 pages. The remaining list was independently evaluated by two raters (first and second author) on the following scale: -1 (reject), 0 (discuss or investigate further), 1 (accept). The evaluations were then merged and compared to define the corpus of articles for the SLR. Conflicts were resolved through discussion. Further exclusion of articles appeared – again through joint discussion – during closer examination and coding of the articles. The reason for exclusion was usually that the papers did not meet the search criteria despite a fitting title or abstract.

For ACM DL, IEEE Xplore as well as DBLP, the last part of the query ("computer science" OR ...) was omitted. Publications were excluded if they dealt with a specific scientific discipline other than computer science. However, if the search terms produced results related to science or engineering in general, they were included into the corpus. Finally, the list of results was examined to see whether the title or abstract suggested a K-12 context. These articles were selected for the present study.

3.2 Data Extraction and Coding

All articles and papers determined in the previous step were imported into MAXQDA 2020. One rater coded the publications, the other reviewed the codings and – where necessary – made extensions and corrections. In the process, additional papers were excluded that after reading them fully turned out to be inappropriate (usually because of a non K-12 focus) or duplicates (hits in more than one database).

Because of the relatively small corpus (s. sec. 4), we analyzed the data qualitatively, following the method of content analysis according to Mayring [80]. The coding was both inductive and deductive, depending on the purpose and research question: It was performed deductively, especially with regard to basic information, such as type of publication, method used, etc. The methods used, as reported below, were coded following the categories of Radovic et al. [99].

With regard to RQ1, a deductive category system based on Darragh's [28] (see section 2.1) identified theoretical frameworks, was used and inductively extended. The deductive categories referred to her distinction of mainly psychological and mainly social oriented theory perspectives that we named *psychological constructs* and *sociocultural theories*. The latter are then further divided into sub-categories for the specific theoretical frameworks of Lave &

---

[1] https://dl.acm.org/

[2] https://ieeexplore.ieee.org

[3] https://link.springer.com/

[4] https://www.sciencedirect.com/

[5] https://dblp.uni-trier.de/

[6] The search was not carried out using the search mask of Google Scholar (https://scholar.google.com/), but by using the tool *Publish or Perish* (see https://harzing.com/resources/publish-or-perish).

Wenger [69], Sfard & Prusak [104], Gee [45], Holland et al. [59], Boaler & Greeno [10] and Martin [77, 78] respectively. Only the *dominant* theory, i. e. the one actually used in the main work of each publication, was coded, as some papers reference a variety of theories in their theoretical background or related work sections. Theories that could not be assigned to any of the categories were coded inductively and given a code memo to record this fact together with the context.

For RQ2 and RQ3 inductive coding was used to form categories from observations in the data. The system of categories reported below was developed by two of the authors. Therefore, they divided the articles among themselves for this purpose and performed coding. The results of the other coder were re-coded and checked for plausibility. Conflicts were resolved through joint discussion. The category system was presented to the other authors and evaluated for plausibility and pragmatic utility as a measures of validity [114].

3.3 Forward and Backward Snowballing

In addition to our initial search, the references in each of our publications were reviewed in light of our search criteria and, as *backward snowballing*, included in our corpus if they fit. A *forward snowballing* was then carried out by querying GoogleScholar for the articles in our corpus and then inspecting the list of referencing sources. All papers from this list, with ten or more references on their own, have been put on the forward snowballing result-list for further investigation.

Both backward and forward snowballing was repeated until saturation was reached, i.e. until no more new papers could be found as a result. In both cases, this happened after two iterations.

For both types of snowballing, stricter criteria have been applied for inclusion in the corpus: A clear connection to computer science or programming and identity, self-concept, or sense of belonging had to emerge directly from the title and abstract. This restriction was made deliberately in order to keep the corpus as close to our subject matter as possible.

A methodological overview of forward and backward snowballing can be found in Wohlin [117]. A discussion of the advantages and disadvantages of a hybrid approach, i.e., a mixture of database search and snowballing, for systematic literature reviews has been given by Mourão et al. [86].

4 RESULTS

The initial search focusing on *identity* resulted in 3881 hits. The evaluation process, which took into account title and abstract, left a selection of 144 papers. The narrower focus on K-12 first reduced the articles to 22 to which forward and backward snowballing added another 19 papers. A closer look at the articles, finally led again to exclusion, if e.g. the context was not clearly computer science, or the topic identity was not really treated, which left 11 publications in the K-12 corpus, initially.

The extended search for articles on the topics self-concept, self-image, self-perception and sense of belonging resulted in 556 new hits, which were reduced to 50 papers after rating title and abstract. For the K-12 context, 20 articles remained with 17 articles being added by snowballing. Reading the articles again led to exclusions so that in total 20 new papers could be added to the corpus that in the end contained *31 publications* that form the data set for the following review [1, 2, 8, 13, 15, 17, 20, 21, 29, 31, 32, 36–38, 58, 61, 67, 70, 71, 79, 83, 96, 97, 102, 105, 107, 108, 111, 112, 118, 119].

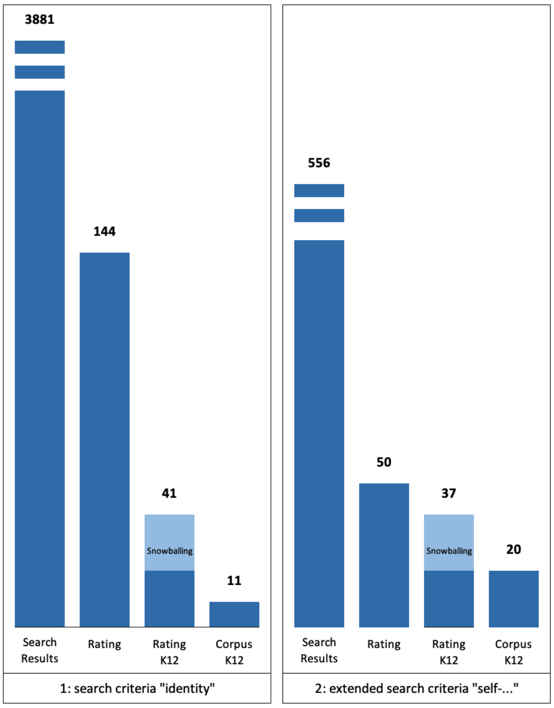


Figure 1: Initial search results and selection process

4.1 Overview

Twelve of the publications appeared as conference contributions, nineteen as journal articles. The most frequently represented conferences are the *Workshop in Primary and Secondary Computing Education* (WiPSCE, 3) and *IEEE Frontiers in Education* (FIE, 3), the most frequently represented journal is *Research in Science Education* (2). All articles were published between 1997 and 2020, but most (26) were published in the last ten (between 2010 and 2020) years (see Fig. 2).

19 of the articles focus on secondary, eight on primary education. For the remaining four, no focal point can be identified. 13 of the articles are qualitative, 11 a quantitative in nature, while five can be classified as mixed method. Two of the articles are theoretically oriented or unclear in regards to the methodology. The most commonly used research tools were surveys (18), interviews (15), social interaction observation (6), and artifact analysis (5). Of the studies, seven have fewer than 10 study participants, 13 have between 10 and 100, 10 have between 100 and 1000, and only one has more than 1000.

Most (22) of the articles in our corpus use the concept of identity, while some (8) refer to self-concept (multiple counts possible). Others discuss identity along the term *sense of belonging* (3) or *stereotype* (3). According to our expectations (s. sec. 3), there are

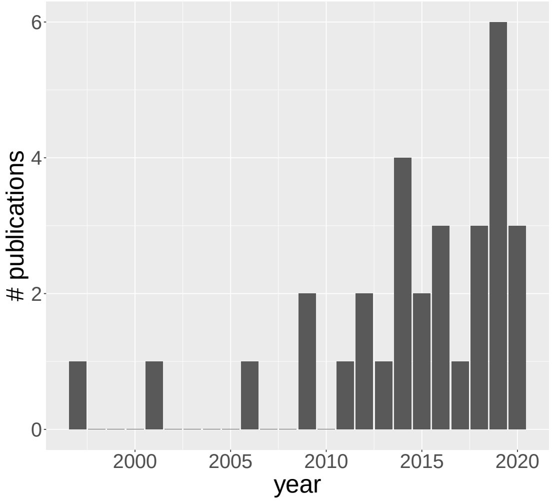


Figure 2: Publications on computer science identity in K-12 per year

some articles that utilize mixed forms or use the terms interchangeably. There are also some articles that employ even more elaborate hybrids, like "self-perception and belonging" [2] or engineering identity as in part "self-perception of belonging in the engineering community" [36]. In contrast, our search did not yield any hits that explicitly emphasize *self-perception* or *self-image*.

4.2 RQ1: What theoretical backgrounds are used to define and conceptualize identity?

The distribution of theories across articles is shown in table 1. There are slightly more works using one of the sociocultural frameworks. Most frequently used was the theory by Lave and Wenger [69]. Two other articles focused on narratives, adapting Sfard's [104] theoretical framework and one used Gee's four kind of identities [45]. The other sociocultural frameworks identified by Darragh [28] were not used in our corpus.

Articles that use identity from a psychological perspective reference a wide range of different frameworks and authors. Eccles et al. [33] were most popular here. In line with the psychological perspective on identity as a personal characteristic, the articles usually focus more on individuals identity development rather than identity dimensions (e.g. specific practices, ways of acting, positioning, relationships) [113] and are treated quantitatively more often.

In addition, three theories not previously included emerged from the open coding: Socio-political or sociological theories, referring e.g. to Butler [18] and Nasir [87], Godwin et al.'s *engineering identity* [47], which is a further development of Carlone and Johnson's *science identity* [22], and Flum and Kaplan's *identity exploration* [39].

A cross-comparison between the theoretical framing used and the methods shows that the psychological theories use surveys (8) as a tool frequently, while the sociocultural ones use interviews (7) and social interactions (5) most often.

4.3 RQ2: How is identity used in research and what are the main findings from this research?

We identified three main categories of how identity theory is used in K-12 research: For *deepening understanding*, for *developing instruments*, and for *evaluating interventions*. The first category (*deepening understanding*) can be further subdivided into two subcategories: First, articles where the *relation to other concepts* is elaborated, and second, articles that explore the disciplinary identity for *specific groups*.

While the categories are not necessarily mutually exclusive in the sense that research might be addressing multiple goals in multiple categories, we tried to identify the predominant category for each article in our corpus.

In the following the main research findings of the articles in our corpus are presented for each of these categories. Please note that the grouping of papers below the categories is not a result of coding, but a summary for ease of understanding that emerged while the results were being put together.

4.3.1 Deepen Understanding: Relation to Other Concepts.

Identitiy in a Larger Social Context. Three (qualitative) studies elaborate the dynamics of positioning the self in social relations and illustrate the reciprocal relationships of influence on disciplinary identity. In line with the presentation of complex social and power relations, the authors draw on identity as a sociocultural (for [108]) or sociopolitical (for [112]) concept (e. g. [45, 69]) with Fields and Enyedy [38] having a stronger emphasis on narratives [104] in their theoretical framing.

Suzuki and Kato [108] describe the formation of identity through the educational software AlgoArena in two cases by analyzing observations and conversations. They conclude that identity formation is not a clear-cut, linear process, but rather a navigation between conflicting identities driven by self-attributions and experiences (as programmers), which may be very situated.

Fields and Enyedy [38] examine how expert and novice identities can co-exist within a programming class and observe that the acceptance of an expert-role by classmates depends on the social setting, may require further reinforcement from the outside, and is not realized through narratives alone, but in practice, through knowledge sharing and working together.

Vakil [112] explores the connection between political identity (with a focus on equity) and learning in computer science education. He concludes that identity development needs (1) a reflection of narratives about the discipline itself and its relationship to the outside world, (2) an opportunity to participate and to work constantly on one's own disciplinary identity and (3), an opportunity to identify with the lived values within computer science. He points out, that these requirements present particular challenges for minorities.

Identitiy and Stereotypes. Two articles investigate the perception of stereotypes by students and their influence on identity development.

Mercier et al. [83] use among other methods a modified *draw a scientist* [23, 48, 53] task and a list of *fluency building activities* developed by the authors (e.g., "Hand-coded a web page using HTML")

Theory/Category	Count	Source papers
Psychological constructs	8	[17, 32, 58, 61, 70, 71, 79, 97]
Sociocultural theories	13	
– Communities of Practice (Lave & Wenger)	10	[1, 13, 20, 37, 67, 96, 102, 107, 108, 111]
– Narratives (Sfard & Prusak)	2	[38, 105]
– Four kinds (Gee)	1	[21]
Other	6	
– Sociopolitical/sociological	3	[112, 118, 119]
– Eng./Science Identity	2	[31, 36]
– Identity exploration	1	[2]
No explicit framework	4	[8, 15, 29, 83]

Table 1: Theoretical framings used by the articles.

and present a statistically significant correlation between the number of corresponding activities and identification as a computer person. Regarding their *draw a computer scientist* task, the drawings differed between boys and girls both for 6th and 8th grade, but the 6th grade drawings contained fewer stereotypical characteristics showing how those characteristics are learned and internalized over time. 75% of students say that *computer persons* exist and associate with it mainly character (not external) traits: Interest in, knowledge of, and great time commitment to computers. Overall, the variance within gender is greater than between gender; however, there is clear evidence of the male stereotype of the computer user.

Master et al. [79] examined gender differences from a "sense of belonging" perspective, to examine the influence of different classrooms designs on boys and girls. A stereotypical classroom was presented to students in photographs or narratives and featured Star Wars/Star Trek items, science fiction books, video games, technology magazines, etc. The authors find that girls report higher interest in participating in a CS course in a non-stereotypical classroom, with a medium to high effect size of $d = 0.61$. Interestingly, there is no difference in interest among boys, indicating that stereotypical classroom design shows a negative cost-benefit balance in any case. Based on a mediation analysis, the authors suspect that differences in students' interest depend on how much they feel they belong in the environment.

Identity and Self-Concept. Finally, the relationship between interest and self-concept is discussed by two quantitative, longitudinal studies from Canada and Finland.

Potvin et al. [97] investigate the development of interest and self-concept in a 2-year study in grades 7 and 8 (N = 540). They report that over the period of four surveys conducted by the authors, interest and self-concept decreases (measured as agreement to predefined items), while the intention to pursue classes in science & technology remains relatively stable. In addition, the study shows that the intention to pursue depends solely on the perceived easiness. A student's interest does not depend on his or her self-concept or gender, but mainly on the novelty of the material. The self-concept in turn depends mostly on the grades and again on the perceived easiness.

Kang et al. [61] observe a stable self-concept of students aged between 13 and 16 (N = 472) and an increasing interest (albeit on a low level), which they attribute to recently introduced curricular changes, namely inquiry based instruction, alignment of teaching with students' lifeworld, and increased student involvement. As a further result, they find that interest and self-concept are distinct and independent facets, at least for their cohort of secondary school students. Interest, but not self-concept, is strongly linked to science aspiration. The authors themselves note that (in line with Potvin et al. [97]), this is in contradiction to other research literature (namely [33, 51, 72]), where self-concept is considered as the most important factor in the development of scientific interest and and the aspirations to choose a scientific career.

4.3.2 Deepen Understanding: Relation to Specific Groups.

Identity and Gender. The relationship between gender and CS identity in the K-12 years is the focus of four studies in our corpus.

Abbiss [1] researches the impact on students' experience of information and communication technology (ICT) in 10th and 12th grade (N = 22) The author finds strong, socially constructed role conceptions in relation to ICT, which go hand in hand with the idea that men are more capable in dealing with computers. At the same time, there is a devaluation of practices that are considered feminine, such as the use of computers primarily for communication purposes. The students rationalize these conceptions in different ways, for example, by resorting to the explanation that it is due to different natural predispositions. The study also identifies three types of computer users considered typical by the students: the Expert Controller, the Aspiring Controller and the Competent Controller, each of which is also gender coded, so that the Expert Controller is perceived as more male, the others as mixed. According to the author, this illustrates a gendered hierarchy of knowledge and power that students already find but performatively actualize with their own attributions.

Sullivan and Bers [107] investigate how kindergarten and school children up to the second grade learn robotics and programming, and in particular which differences can be identified between boys and girls. While both boys and girls were able to learn the basic concepts of computer programming equally well, boys performed significantly better on more advanced concepts. The authors suggest that this can be attributed to different problem-solving strategies. The children's attitude towards the robots, on the other hand, was unrelated to gender, in contrast to, for example, Lego toys, which were identified by the study participants as toys for boys.

This might present a *stereotype threat* [106], using such a system may have a deterrent effect on girls and can prevent their active engagement with technology.

Wong [118] investigates prevailing stereotypes and their impact on the development of a disciplinary identity with a focus on gender. He finds students' conceptions of identity to be clearly shaped by stereotypes. The characteristics he has identified from interviews with adolescents are that individuals with a CS identity are highly committed, often exhibit cleverness or geekyness, are individualistic or even anti-social, but are also curious and creative. The reasons given by young people for not wanting to identify with this image are that the subject appears to be too difficult and "not for me", is too masculine in its character, or does not have attractive career perspectives (the respondents cannot imagine spending all day in front of the computer). In a follow-up study by Wong and Kemp [119] using the same interview data, the aspect of creativity in particular is critically discussed further: It could make pursuing CS in tertiary education more attractive for women, but at the same time also deepen existing prejudices. In addition, the authors discuss curricular interventions to help breaking up stereotypical ideas about the subject, such as broadening the scope of CS towards a more general education or integrating programming topics into other subjects.

Identity and Girls. Two studies specifically focus on women's identity in CS.

The long-term study by Bieri et al. [17] investigates the choice of study paths after K-12 and shows that the students' intentions for the choice of studying a STEM-subject remained stable in the period before their graduation and entering university. The intention was mostly shaped by an early fascination for scientific topics, as well as the possibility to shape and actively influence the world. The father as a role model could not be quantitatively confirmed as an important factor for career choice, while an ambivalent picture emerged from interviews conducted by the authors. Math proficiency was shown to be a good predictor of STEM choice, although the study participants interviewed did not describe themselves as 'good' at math, but only as 'not bad'. The authors suspect a kind of coping strategy behind this observation.

DuBow et al. [31, 32] interviewed applicants for and winners of the NCWIT Aspirations Award. The award is given to women and an application expresses a strong interest in computer science. Over a period of three years, interviews were conducted with 64 women, most of whom were no longer students at the time of the interviews, but who, in retrospect, reflected on the conditions that steered them towards computer science. The authors create profiles and identify the essential characteristics leading to students remaining in computer science. These are: sufficient exposure to computer science both in and out of school, support from the community, i.e. teachers, parents and friends, and respect and encouragement from others. It turned out that it was much more difficult for women who lacked even one of these aspects to remain in the discipline.

Identity and Minorities. The relationship of CS identity to a minority identity is examined by three studies.

DiSalvo et al. [29] describe how their study subjects (African American males) continually renegotiate and justify their engagement as game testers and CS-interested people to different groups in different ways. For the youth, not identifying themselves with CS means not learning, even though they could succeed in the subject. The authors therefore recommend offering not just one way of identifying with CS, but different approaches; including those that allow otherwise underserved groups to save their face while acting *geeky*. This as well implies a necessity to develop novel interventions that have a greater connection to the reality of young people's lives.

Tupou and Loveridge [111] explore the nature of engagement with CS for six year 7 and 8 Pasifka students. Their qualitative analysis condensed four aspects of engagement with CS: participation, reification, imagination and alignment. The study indicates that participants used computers frequently and rated themselves as confident. But the computer use of the students was *consuming* and not *creating*. Additionally it was hard for the students to tell about the nature of CS and about the relevance of CS for their own life. The authors conclude a lack of understanding about the students' role in CS. They cite the correlation to a missing future engagement in CS.

As one of several examples for intersectional identity research, in a long-term study by Brickhouse and Porter [13] the computer science identity development of two black high school girls – Ruby and Chrystal – was observed. While Ruby is less successful in school overall, she is successful in her CS course. This is probably because she values social opportunities (chatting on AOL) and has a strong and inclusive role model in her father. She sees CS as a means to an end and is able to successfully unify it with her other hobbies and interests, especially those that are more feminine coded. Chrystal, while very successful in school overall, is less so in CS. She takes a rather reserved position in school – presumably out of *stereotype threat* – as a good, quiet girl. She does not have a CS role model, nor does she see the social opportunities Ruby recognizes for herself in technology. The detailed case studies thus point to the importance of anchoring CS content in the social and lifeworld, the importance of role models, and the difficulty of negotiating one's identity in a subject with the rest of the framework.

4.3.3 Developing instruments. Capobianco et al. [20, 21] describe the development of an an instrument for measuring young learners' engineering identity development and elaborate the Engineering Identity Development Scale [EIDS] to measure children's academic, school, and occupational identity and engineering aspirations. Based on the Utrecht-Groningen Identity Development Scale (GIDS) and the Harter's Self-Perceptions Profile for Children (SPPC) [54, 82] they combine both with Capobianco's engineering identity dimensions framework [19] and identify (and empirically confirm) four factors for their own scale for measuring engineering identity and created at least six item for each factor [20]: (1) academic identity (self-beliefs or self-images in who children think they are as students) (2) school identity (children's affiliation or attachment to their school) (3) occupational identity (children's self-understandings of an occupation) (4) engineering aspirations (children's self-goals, aims, or objectives of becoming an engineer). The authors summarize this as follows: "EIDS results indicate that girls developed a new sense of identification with what engineers do (i.e. design, work in teams, use science and math, are creative) and who they want to become relative to engineering (i.e. solve problems that help people, design different things, and work on a team with engineers)" [20].

The "Self-Concept and Attitude toward Programming Assessment" (SCAPA) by Leifheit et al. [70, 71] is a seven scale questionnaire, focusing on self-reported prior experience and understanding of programming, self-concept with regard to programming, intrinsic value belief about programming, attainment value belief about programming, utility value belief about programming, cost belief about programming and compliance and persistence with regard to programming. The target group of the questionnaire are primary school children, so the authors placed particular emphasis on simple and understandable language during creation. The instrument was inspired by similar questionnaires in mathematics [14, 43, 109] that have also already been successfully transferred to other subject areas (biology, physics, English) [44]. SCAPA was initially tested with 31 students [70] and validated in a subsequent study with 197 students [71]. The authors were able to demonstrate both a good model fit of the subscales as well as a high consistency of the individual scales. Nevertheless, in their summary the authors recommend reviewing the instrument with further studies.

In their recent study, Kong and Wang [67] not only develop an instrument for determining computational identity, but also attempt to fill the gap of the connection between computational identity (CI) and computational thinking (CT). To do so, they develop items together with experts for both concepts, starting from a sociocultural understanding of identity. The four subconstructs of their identity conceptualization – Programming Affiliation, Programming Goal Setting, Programming Engagement, and Programming Actualization – can be located on the axes Personal/Social Identity and Present/Future Identity (see the discussion in sections 5.1 and 5.2). Programming Affiliation corresponds to what is called *sense of belonging* in other studies, but is clearly distinguished here as a subconstruct of identity. A pilot study to validate the instruments for CI and CT was able to demonstrate (by means of confirmatory factor analysis) both their model fit and their discriminatory power with respect to each other. A subsequent main study examined the relationship between CI and CT and found that two of the subconstructs of CT significantly positively influenced CI (the third positively but not significantly).

4.3.4 *Evaluating interventions.*

Çakır et al. [2] develop a game design workshop for young girls to support their self-perception and identity in computer science. To evaluate the effectiveness the authors used a mixed method approach [27] including pre- and post-workshop surveys as well as interviews. 21 girls in grades five through eight took part in the workshop. The authors summarize their results as highlighting the importance of positive hands-on experiences in shaping the attitudes of girls towards CS. From a theoretical point of view, the authors attribute the positive results to the application of the identity exploration framework in combination with a motivating team task from the area of game development. The learning environment included psychosocial and sociocultural elements that contributed positively to the students' identity development. The participants experienced themselves in the different roles of software developers and learned how to work together in real projects. Both contributed to being able to identify themselves and the others as a "computer person".

Shaw et al. [105] present a unit for the Computer Science Curriculum that makes use of reflective portfolios to enable students to shape their perceptions on CS. The authors were able to show that by preparing their portfolios, students developed their own narratives about computing and their place in the field. They not only articulated who they were in connection with computer science, but also identified numerous resources, skills and personal characteristics that helped them construct artifacts. In addition, the portfolios allowed them to deepen their understanding of computer science and develop new approaches to the subject. They allowed students to narrate who they could be in the future in the field of computer science and to express their interest in computer science outside the classroom.

Other articles report the development of interventions [8, 15, 36, 37, 58, 96, 102] but have no, insufficient, anticipated, or preliminary results, so we do not detail their findings here.

4.4 RQ3: Why is the Concept of Identity Used in Literature?

The authors of the articles in our corpus cite a variety of reasons to justify the discussion of – or the theoretical foundation in – identity. As our inductive coding shows, these reasons can essentially be divided into five, non-mutually exclusive categories, shortly described in the following sections. Table 2 shows the distribution of the reasons over our corpus. A cross comparison between the reasons and the theoretical framing used shows that sociocultural theories are used most frequently in the context of diversity (7), and next most frequently in the context of retention (3). For psychological theories, the picture is less clear: the two most frequent appearances are in connection with diversity (4) and the promotion of young talent (4).

4.4.1 *Diversity.*

The motivation most often cited by authors for using identity as a theoretical framework is diversity: identity is thus particularly well suited for capturing existing inequalities between groups of learners and analyzing how these can be avoided in the future. In this context, results from research on stereotypes are often referred to, e. g. [63, 90, 103].

4.4.2 *Retention.*

The next most frequent reason given for dealing with identity was the positive influence on retention or persistence that a pronounced identity has: Students who have developed this pronounced identity are more likely to cope with the demands of the subject. Literature and empirical results, to which reference is made in this context, are e. g. [4, 16, 19, 26, 115].

4.4.3 *Promotion of Young Talent.*

A pronounced identity is also seen as a positive influence on a subsequent choice of electives, courses of study, or jobs [47, 60, 72]. In particular, as, e. g. [31] argue, developing a CS identity is particularly important for counteracting the often reported lack of young talent in STEM subjects [120]. This is especially true for groups that are currently underrepresented in computer science, such as women [7].

4.4.4 *Improved Learning/Better Performance.*

Another positive aspect of a strong identity is seen in improving overall learning and performance and promoting academic performance in the long term. This aspect in particular is difficult to separate from motivational aspects [39] and, both are often mentioned in unison (e. g. [2]).

4.4.5 Higher Motivation. The role of increased motivation resulting from identity is discussed in literature from various points of view: (1) Motivation arises as a by-product, so to speak, which serves to maintain identity [57, 93]. (2) Motivation follows from the examination of new identities, since the exploration of possible new identities is a dynamic process and thus holds interesting new discoveries in store [39]. (3) Motivation is more or less equated with perseverance, for example by reference to [25]. (4) Furthermore, there is the abstract statement that self-concept and motivation go hand in hand. Here [33] and [91] are mentioned as evidence.

5 DISCUSSION

We analyzed our corpus along three research questions. We will now discuss the findings presented above for each question but also show how they interrelate with one another.

5.1 RQ1: What theoretical backgrounds are used to define and conceptualize identity?

The sociocultural theories appear predominantly in our corpus, followed by psychological theories. This reproduces not only similar findings by Darragh [28], but also by Pozzer and Jackson [98] who, in a literature review on identity in science education research, have found the same result regarding theory distribution. It is noteworthy that for psychological theories, we find a much broader range of authors and ideas, than for sociocultural theories, which mainly focus on Lave and Wenger's [69] Communities of Practice. Also, works based on the psychological strand tend to operate quantitatively while sociocultural works tends to be qualitative. Obviously, our two strands characterize further differences, besides the shown relations to the methods used. For example, sociocultural theories are more often used to address diversity questions, while psychological constructs focus on interest development or retention.

Also noteworthy is that the specific theory of Carlone and Johnson [22] with their concept of *science identity*, which was then extended by Godwin [47] to an *engineering identity* is now appearing in computer science as *computing identity* [73]. There are some interesting differences between this and Kong and Wang's [67] conceptualization of *computing identity* (CI) present in our corpus: While CI by Mahadeo et al. [73] assume more of a psychological, *subjective* [99] notion of identity, Kong and Wang's framing is explicitly sociocultural, but with a simultaneous investigation and exploration of the connection to computational thinking. Thus, both instruments have their place in the appropriate research setting. However, Kong and Wang's instrument seems more explanatory and – in the sense of Nelson and Ko [88] – more like a theoretical and practical development out of the specific needs of computer science.

In summary, the division into the presented strands of theory makes the concept of identity seem much less fuzzy than the plethora of concepts does. However, there are still some problems with the theory that need attention:

First, identity theory is used but rarely critiqued. Others have pointed out that identity is difficult to operationalize in various popular conceptualizations, earning the term the suspicion of unscientificness. Also, some conceptualizations are susceptible to an implicit essentialism that obscurely conveys the very thing it actually seeks to resolve: a normative, non-socially negotiated image of what a discipline is [104].

Second, a conceptual clarification and differentiation from other theories is necessary: It is unclear, for example, whether *sense of belonging* is a subconstruct of identity (in a sociocultural framing) or can be meaningfully considered as a concept in its own right. Regarding the distinction between *self-concept* and *self-efficacy*, on the other hand, there is a long and ongoing debate in psychology and educational science that deserves attention [12, 76]. This particularly affects psychological theories, which, as noted above, often operate under the notion of self-concept.

Third, taking into account the results of RQ2, especially in the use of sociocultural conceptualizations of identity, there is a general lack of the specific, e.g.: What practices make up the community of practice of computer scientists? Is there only one community of practice or how are the many, small communities of practice interconnected? What is the lowest common denominator of values in computer science? How are shared values and practices negotiated? This lack of concrete and operationalization has been noted before [104]. Conversely, in the use of self-concept, i. e., psychological theories, there is often a lack of an overarching theoretical framework that embeds the subjective view of identity in the disciplinary context with all its social and discursive practices. Again, Kong and Wang [67] are the exception because they take both domains seriously, thus creating a link between the different strands of theory.

5.2 RQ2: How is identity used in research and what are the main findings from this research?

Overall, the results regarding RQ2 paint a rather clear picture: Identity is used in particular with regard to diversity in computer science. This includes both positive aspects of increasing identification for specific groups as well as negative aspects, such as stereotypes as negative templates or means of disidentification. Most surprising regarding our review is the apparent lack of interventions developed. Specific recommendations are mostly limited to broad, general, or even superficial hints. For example Master et al. [79] find that geek-friendly designed classrooms have significant negative effects on girls while neutral classrooms do not affect any group particularly negatively. Perhaps the lack of interventions is due to a lack of instruments development to reliably and validly evaluate them, so that it is all the more welcome that there is now at least a small range of instruments available in the K-12 context.

Computer science is still perceived as male-dominated and white, despite the continuing effort of broadening participation. The articles by Wong [118, 119] and DiSalvo et al. [29] provide a perspective on how to mitigate this problem: Through more openly designed participation opportunities that allow for mediation between different identities. It is particularly this concept of intersectionality that could also prove to be beneficial as a focal point for the analytical

Why?	Count	Source papers
Diversity	19	[1, 2, 13, 15, 17, 20, 21, 29, 31, 32, 36, 61, 79, 83, 107, 111, 112, 118, 119]
Retention	7	[20, 31, 36, 70, 71, 96, 102]
Promotion of Young Talent	7	[21, 32, 36, 58, 61, 96, 97]
Improved Learning	7	[2, 8, 29, 37, 70, 71, 111]
Higher Motivation	5	[2, 36, 61, 71, 105]
Without, i.e. academic curiosity	3	[38, 67, 108]

Table 2: Reasons given for using identity.

lens when focusing on the specific needs of particular groups, as it is precisely these points of friction between different identities and how they come together that can lead to difficulties in one identity or another. In this context, the reported demarcation phenomena are interesting – that students just do not want to be identified with computer science – because this would be perceived as disturbing within another identity. Overall, however, it would be beneficial for such a consideration if a sharper picture (or approximation) of the disciplinary identity as a computer scientist was developed beforehand.

5.3 RQ3: Why is the concept of identity used in literature?

According to the analysis presented above, the question why identity is important to consider in the K-12 area is essentially answered by literature in five different ways (see 4.4). These five reasons are elaborated on to varying degrees, though: As mentioned before, the by far most often cited reason to employ identity as theoretical framework is some kind of investigation into *diversity*. This is unsurprising given the expectations associated with the concept as a connecting link between subjective learning success and social circumstances.

The link between *better performance* and identity remains unclear. Theory postulates that "learning and a sense of identity are inseparable: They are aspects of the same phenomenon" [69, p. 115] yet empirical evidence from our corpus so far only shows half of this interdependence [97]: Good grades lead to an enhanced identity. In the same way, additional literature points out that it is the recognition of teachers or role models that strengthens the self-concept [46]. Meanwhile, DiSalvo et al. show that learning can be prevented when one's identity (e.g. as African American male) cannot be aligned with one's disciplinary identity, that is, when the complex, social negotiation process mentioned in many of the articles cited in Sec. 4.3 fails.

A frequently voiced reason from our corpus for the importance of developing a disciplinary identity is to get more young people into computer science. A pronounced self-concept is mentioned as important for the choice of studies and career, however, the role of self-concept remains unclear [61, 97]: It seems to play a role, but is less important than interest. Interest, on the other hand, is best stimulated when students come into contact with novel teaching methods and objects. Kang attributes this misperception of self-concept to the fact that the concepts of interest and self-concept are often not sharply separated in measurement.

The evidence cited in our corpus suggest a strong, albeit unclear connection between identity development and motivation [11]: Four different kinds of relationship between the two concepts were identified. A clear separation between identity and motivation seems to be difficult and a more precise, empirical investigation of the dependency relationships and a delimitation of the terms would be desirable.

The importance of identity for retention is undisputed in our corpus. Thus it has an interrelation with interest: While interest ensures that students pursue a computer science degree or career [61], identity ensures that they persevere and keep on track (see section 4.4.2). Both should be sparked and developed – if one follows these results – already at school, maybe even as early as possible [74]. Retention is also often associated with the Big Five characteristic of conscientiousness [3, 110] or "grit" [35]. Further research would be interesting in this respect to clarify whether there is a connection between identity and these concepts.

5.4 Limitations

Due to the nature of systematic literature reviews as well as the definitional boundaries of identity (s. sec. 2), the need to define a precise search term that makes the search reproducible also makes it possible, even probable, that certain articles will not appear in the corpus. We have taken steps – such as forward and backward snowballing – to minimize this risk and detect articles relevant to our focus but not retrieved by our initial search. Yet there may still be publications that are not detected in this manner and, in particular, there may also be articles where the search terms used do not appear prominently enough to be detected by our search.

As a search criterion, it was determined in advance that papers are included into the corpus that originate from engineering or science, as long as they have a CS connection (e.g., through study participants from computer science classes). While this broad search criterion does justice to the – internationally – large variety of terms (computer science, computer engineering, information and technology studies, etc.), it also means a loss of definitional sharpness. This finding as well points to the need for further research to ensure that the results presented hold for (pure) computer science classes.

6 CONCLUSION

In this article we reviewed the literature regarding the concept of *identity* in computer science education (research) with a specific focus on K-12. In summary, our findings show that currently, identity theory is only used (in the sense of consumed) in CSER, there currently is no indication of a subject-specific theory building that not only applies theoretical concepts, but discusses and develops them in relation to specifics of computer science education. This presents a clear opportunity for further research in our community. Also, there is still potential to use other (sociocultural-) theoretical frameworks, such as *figured worlds* by Holland et al. [59].

The most prominent reasons for research in our corpus is identity exploration, learning about and developing an interest in computer science and preventing that students are "scared away" from CS right at the beginning. Interestingly, there are numerous articles in our corpus that cite retention as the reason for referring to theories of identity, yet there are no empirical interventions that would explore the usefulness of this in the K-12 context. The empirical interventions that we identified are mostly related to diversity and interest. This is in line with our expectations (see section 2.2) regarding the role that the K-12 years can play with regard to identity development. Accordingly, a sensible further development would be to steer research in this direction in order to fully realize the potential of identity theory as a "missing link" [104].

The "hidden curriculum" [113] that we referred to in this context, is apparent in the form of social perceptions and stereotypes. As detailed in the findings on RQ2, numerous papers deal with the consequences of stereotypes and stereotype threat. In particular, underserved groups in computer science view existing stereotypes as negative templates for their identification. This urge for differentiation then ensures that a negative feedback loop develops. These problems are not new [75], so it is all the more surprising that they are no less acute than they were twenty years ago. Teachers and educators have to actively work against those to provide equal access for all kinds of students and thus broaden participation in computer science.

The literature review presented here can hopefully serve as a basis for making informed decisions about designing interventions and research desiderata. In particular, evaluating the success of interventions needs more research about operationalizing or measuring (in a quantitative or qualitative sense) identity.

REFERENCES

[1] Jane Abbiss. 2011. Boys and machines: gendered computer identities, regulation and resistance. *Gender and Education* 23, 5 (2011), 601–617.

[2] Nur Akkus Çakir, Arianna Gass, Aroutis Foster, and Frank J. Lee. 2017. Development of a game-design workshop to promote young girls' interest towards computing through identity exploration. *Computers & Education* 108 (May 2017), 115–130. https://doi.org/10.1016/j.compedu.2017.02.002

[3] Gene Alarcon and Jean Edwards. 2013. Ability and Motivation: Assessing Individual Factors That Contribute to University Retention. *Journal of Educational Psychology* 105 (Feb. 2013), 129. https://doi.org/10.1037/a0028496

[4] Jane M. Armstrong and Richard A. Price. 1982. Correlates and Predictors of Women's Mathematics Participation. *Journal for Research in Mathematics Education* 13, 2 (1982), 99–109. https://doi.org/10.2307/748357

[5] Joseph A. Bailey. 2003. Self-image, self-concept, and self-identity revisited. *Journal of the National Medical Association* 95 5 (2003), 383–6.

[6] Roy F. Baumeister. 1997. Chapter 26 - Identity, Self-Concept, and Self-Esteem: The Self Lost and Found. Academic Press, San Diego, 681–710. https://doi.org/10.1016/B978-012134645-4/50027-5

[7] David N Beede, Tiffany A Julian, David Langdon, George McKittrick, Beethika Khan, and Mark E Doms. 2011. Women in STEM: A gender gap to innovation. *Economics and Statistics Administration Issue Brief* 04-11 (2011).

[8] Kera Bell-Watkins, Tiffany Barnes, and Nathan Thomas. 2009. Developing computing identity as a model for prioritizing dynamic K-12 computing curricular standards. *Journal of Computing Sciences in Colleges* 24, 3 (Jan. 2009), 125–131.

[9] Maureen Biggers, Anne Brauer, and Tuba Yilmaz. 2008. Student Perceptions of Computer Science: A Retention Study Comparing Graduating Seniors with Cs Leavers. In *Proceedings of the 39th SIGCSE Technical Symposium on Computer Science Education* (Portland, OR, USA) *(SIGCSE '08)*. Association for Computing Machinery, New York, NY, USA, 402–406. https://doi.org/10.1145/1352135.1352274

[10] Jo Boaler. 2000. *Multiple perspectives on mathematics teaching and learning.* Vol. 1. Greenwood Publishing Group.

[11] Mimi Bong and Richard E. Clark. 1999. Comparison between self-concept and self-efficacy in academic motivation research. *Educational Psychologist* 34, 3 (June 1999), 139–153. https://doi.org/10.1207/s15326985ep3403_1

[12] Mimi Bong and Einar M. Skaalvik. 2003. Academic Self-Concept and Self-Efficacy: How Different Are They Really? *Educational Psychology Review* 15, 1 (2003), 1–40. https://doi.org/10.1023/A:1021302408382

[13] Nancy W. Brickhouse and Jennifer T. Potter. 2001. Young women's scientific identity formation in an urban context. *Journal of Research in Science Teaching* 38, 8 (2001), 965–980. https://doi.org/10.1002/tea.1041 arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1002/tea.1041

[14] Brigitte Maria Brisson, Anna-Lena Dicke, Hanna Gaspard, Isabelle Häfner, Barbara Flunger, Benjamin Nagengast, and Ulrich Trautwein. 2017. Short Intervention, Sustained Effects: Promoting Students' Math Competence Beliefs, Effort, and Achievement. *American Educational Research Journal* 54, 6 (2017), 1048–1078. https://doi.org/10.3102/0002831217716084

[15] Monica F. Bugallo and Angela M. Kelly. 2014. A pre-college recruitment strategy for electrical and computer engineering study. In *2014 IEEE Integrated STEM Education Conference*. 1–4. https://doi.org/10.1109/ISECon.2014.6891010

[16] Jennifer Burg, V. Paúl Pauca, William Turkett, Errin Fulp, Samuel S. Cho, Peter Santago, Daniel Cañas, and H. Donald Gage. 2015. Engaging Non-Traditional Students in Computer Science through Socially-Inspired Learning and Sustained Mentoring. In *Proceedings of the 46th ACM Technical Symposium on Computer Science Education (SIGCSE '15)*. Association for Computing Machinery, New York, NY, USA, 639–644. https://doi.org/10.1145/2676723.2677266

[17] Christine Bieri Buschor, Simone Berweger, Andrea Keck Frei, and Christa Kappler. 2013. Majoring in STEM—What Accounts for Women's Career Decision Making? A Mixed Methods Study. *The Journal of Educational Research* 107, 3 (dec 2013), 167–176. https://doi.org/10.1080/00220671.2013.788989

[18] Judith Butler. 1990. Gender trouble, feminist theory, and psychoanalytic discourse. *Feminism/postmodernism* 327 (1990), x.

[19] Brenda M Capobianco. 2006. Undergraduate women engineering their professional identities. *Journal of Women and minorities in Science and Engineering* 12, 2-3 (2006), 95–117.

[20] Brenda M. Capobianco, Heidi A. Diefes-Dux, and Meara M. Habashi. 2009. Generating measures of Engineering Identity Development among young learners. In *2009 39th IEEE Frontiers in Education Conference*. 1–6. https://doi.org/10.1109/FIE.2009.5350817 ISSN: 2377-634X.

[21] Brenda M. Capobianco, Brian F. French, and Heidi A. Diefes-Du. 2012. Engineering Identity Development Among Pre-Adolescent Learners. *Journal of Engineering Education* 101, 4 (2012), 698–716. https://doi.org/10.1002/j.2168-9830.2012.tb01125.x

[22] Heidi B. Carlone and Angela Johnson. 2007. Understanding the science experiences of successful women of color: Science identity as an analytic lens. *Journal of Research in Science Teaching* 44, 8 (2007), 1187–1218. https://doi.org/10.1002/tea.20237

[23] David Wade Chambers. 1983. Stereotypic images of the scientist: The draw-a-scientist test. *Science education* 67, 2 (1983), 255–265.

[24] Abigail Clark and Rachel Louis Kajfez. 2019. Engineering Identity in Pre-College Students: A Literature Review. In *ASEE 2019 Annual Conference*.

[25] Joanne Cohoon and William Aspray. 2008. Just Get Over It or Just Get On with It: Retaining Women in Undergraduate Computing. (2008).

[26] J McGrath Cohoon and William Aspray. 2006. *Women and information technology: Research on underrepresentation.* Vol. 1. The MIT Press.

[27] John W Creswell and J David Creswell. 2017. *Research design: Qualitative, quantitative, and mixed methods approaches.* Sage publications.

[28] Lisa Darragh. 2016. Identity research in mathematics education. *Educational Studies in Mathematics* 93, 1 (Sept. 2016), 19–33. https://doi.org/10.1007/s10649-016-9696-5

[29] Betsy DiSalvo, Mark Guzdial, Amy Bruckman, and Tom McKlin. 2014. Saving Face While Geeking Out: Video Game Testing as a Justification for Learning Computer Science. *Journal of the Learning Sciences* 23, 3 (2014), 272–315. https://doi.org/10.1080/10508406.2014.893434 arXiv:https://doi.org/10.1080/10508406.2014.893434

[30] Brian Dorn and Allison Elliott Tew. 2015. Empirical validation and application of the computing attitudes survey. *Computer Science Education* 25, 1 (2015), 1–36. https://doi.org/10.1080/08993408.2015.1014142

[31] Wendy DuBow, Joanna Weidler-Lewis, and Alexis Kaminsky. 2016. Multiple factors converge to influence women's persistence in computing: A qualitative analysis of persisters and nonpersisters. In *2016 Research on Equity and Sustained Participation in Engineering, Computing, and Technology (RESPECT)*. 1–7. https://doi.org/10.1109/RESPECT.2016.7836161

[32] Wendy M. DuBow and Laurie James-Hawkins. 2015. What influences female interest and persistence in computing? Preliminary findings from a multi-year study. In *2015 Research in Equity and Sustained Participation in Engineering, Computing, and Technology (RESPECT)*. 1–5. https://doi.org/10.1109/RESPECT.2015.7296498

[33] J. Eccles. 1983. Expectancies, values and academic behaviors. *Achievement and achievement motives* (1983).

[34] E.H. Erikson. 1968. *Identity: youth and crisis.* Norton & Co., Oxford, England.

[35] Lauren Eskreis-Winkler, Elizabeth P. Shulman, Scott A. Beal, and Angela L. Duckworth. 2014. The grit effect: predicting retention in the military, the workplace, school and marriage. *Frontiers in Psychology* 5 (Feb. 2014). https://doi.org/10.3389/fpsyg.2014.00036

[36] Aamir Fidai, Luciana R. Barroso, Mary M. Capraro, and Robert M. Capraro. 2018. Can Building an Electric Bicycle Build an Engineering Identity?. In *2018 IEEE Frontiers in Education Conference (FIE)*. 1–5. https://doi.org/10.1109/FIE.2018.8659037 ISSN: 2377-634X.

[37] Aamir Fidai, Cynthia Jarvis, Madison Benzor, Saaransh Verma, Mary Margaret Capraro, and Robert M. Capraro. 2019. Motivating Future Engineers: Building Situation Sensing Mars Rover with Elementary School Students. In *2019 IEEE Frontiers in Education Conference (FIE)*. 1–7. https://doi.org/10.1109/FIE43999.2019.9028368 ISSN: 2377-634X.

[38] Deborah Fields and Noel Enyedy. 2013. Picking up the mantle of "expert": Assigned roles, assertion of identity, and peer recognition within a programming class. *Mind, Culture, and Activity* 20, 2 (2013), 113–131.

[39] Hanoch Flum and Avi Kaplan. 2006. Exploratory Orientation as an Educational Goal. *Educational Psychologist* 41, 2 (June 2006), 99–110. https://doi.org/10.1207/s15326985ep4102_3

[40] Hanoch Flum and Avi Kaplan. 2012. Identity formation in educational settings: A contextualized view of theory and research in practice. *Contemporary Educational Psychology* 37, 3 (2012), 240 – 245. https://doi.org/10.1016/j.cedpsych.2012.01.003 Identity Formation in Educational Settings.

[41] Michel Foucault. 1980. *Power/knowledge: Selected interviews and other writings, 1972-1977.* Vintage.

[42] Francis Fukuyama. 2018. *Identity: The demand for dignity and the politics of resentment.* Farrar, Straus and Giroux.

[43] Hanna Gaspard. 2015. *Promoting Value Beliefs in Mathematics : A Multidimensional Perspective and the Role of Gender.* Ph.D. Dissertation. Universität Tübigen. https://doi.org/10.15496/publikation-5241

[44] Hanna Gaspard, Isabelle Häfner, Cora Parrisius, Ulrich Trautwein, and Benjamin Nagengast. 2016. Assessing Task Values in Five Subjects During Secondary School: Measurement Structure and Mean Level Differences Across Grade Level, Gender, and Academic Subject. *Contemporary Educational Psychology* 48 (Oct. 2016). https://doi.org/10.1016/j.cedpsych.2016.09.003

[45] James Paul Gee. 2000. Chapter 3: Identity as an analytic lens for research in education. *Review of research in education* 25, 1 (2000), 99–125.

[46] Donald E. Gibson. 2003. Developing the Professional Self-Concept: Role Model Construals in Early, Middle, and Late Career Stages. *Organization Science* 14, 5 (2003), 591–610.

[47] Allison Godwin, Geoff Potvin, Zahra Hazari, and Robynne Lock. 2016. Identity, Critical Agency, and Engineering: An Affective Model for Predicting Engineering as a Career Choice. *Journal of Engineering Education* 105, 2 (2016), 312–340. https://doi.org/10.1002/jee.20118

[48] Florence Laura Goodenough. 1926. Measurement of intelligence by drawings. (1926).

[49] Gregor Große-Bölting, Yannick Schneider, and Andreas Mühling. 2019. It's like Computers Speak a Different Language: Beginning Students' Conceptions of Computer Science. In *Proceedings of the 19th Koli Calling International Conference on Computing Education Research* (Koli, Finland) *(Koli Calling '19)*. Association for Computing Machinery, New York, NY, USA, Article 2, 5 pages. https://doi.org/10.1145/3364510.3364527

[50] Gregor Große-Bölting, Yannick Schneider, and Andreas Mühling. 2020. Beginning Students' Conceptions of Computer Science: The Effect of the First Semester. In *2020 International Conference on Learning and Teaching in Computing and Engineering (LaTICE)*. Ho-Chi-Minh-City; Vietnam.

[51] Jiesi Guo, Herbert W. Marsh, Philip D. Parker, Alexandre J. S. Morin, and Theresa Dicke. 2017. Extending expectancy-value theory predictions of achievement and aspirations in science: Dimensional comparison processes and expectancy-by-value interactions. *Learning and Instruction* 49 (June 2017), 81–91. https://doi.org/10.1016/j.learninstruc.2016.12.007

[52] Lelia Hampton, Robert Cummings, and Kinnis Gosha. 2019. Improving Computer Science Instruction and Computer Use for African American Secondary School Students: A Focus Group Exploration of Computer Science Identity of African American Teachers. In *Proceedings of the 2019 on Computers and People Research Conference (SIGMIS-CPR '19)*. Association for Computing Machinery, New York, NY, USA, 78–84. https://doi.org/10.1145/3322385.3322399

[53] Dale B Harris. 1965. Children's drawings as measures of intellectual maturity. (1965).

[54] Susan Harter. 1985. *Manual for the self-perception profile for children:(revision of the perceived competence scale for children).* University of Denver.

[55] Susan Harter. 1990. *Self and identity development.* Harvard University Press, Cambridge, MA, US, 352–387.

[56] Michael Hewner. 2013. Undergraduate Conceptions of the Field of Computer Science. In *Proceedings of the Ninth Annual International ACM Conference on International Computing Education Research* (San Diego, San California, USA) *(ICER '13)*. Association for Computing Machinery, New York, NY, USA, 107–114. https://doi.org/10.1145/2493394.2493414

[57] Michael Hewner and Maria Knobelsdorf. 2008. Understanding computing stereotypes with self-categorization theory. In *Proceedings of the 8th international Conference on Computing Education Research.* 72–75.

[58] Claudia Hildebrandt and Ira Diethelm. 2012. The school experiment InTech: how to influence interest, self-concept of ability in informatics and vocational orientation. In *Proceedings of the 7th Workshop in Primary and Secondary Computing Education (WiPSCE '12)*. Association for Computing Machinery, New York, NY, USA, 30–39. https://doi.org/10.1145/2481449.2481458

[59] Dorothy C Holland, William Lachicotte Jr, Debra Skinner, and Carole Cain. 2001. *Identity and agency in cultural worlds.* Harvard University Press.

[60] Henriette Tolstrup Holmegaard, Lars M. Ulriksen, and Lene Møller Madsen. 2014. The Process of Choosing What to Study: A Longitudinal Study of Upper Secondary Students' Identity Work When Choosing Higher Education. *Scandinavian Journal of Educational Research* 58, 1 (Jan. 2014), 21–40. https://doi.org/10.1080/00313831.2012.696212

[61] Jingoo Kang, Tuula Keinonen, and Anssi Salonen. 2019. Role of Interest and Self-Concept in Predicting Science Aspirations: Gender Study. *Research in Science Education* (Nov. 2019). https://doi.org/10.1007/s11165-019-09905-w

[62] Avi Kaplan and Hanoch Flum. 2012. Identity formation in educational settings: A critical focus for education in the 21st century. *Contemporary Educational Psychology* 37, 3 (2012), 171 – 175. https://doi.org/10.1016/j.cedpsych.2012.01.005 Identity Formation in Educational Settings.

[63] Amy K. Kiefer and Denise Sekaquaptewa. 2007. Implicit stereotypes and women's math performance: How implicit gender-math stereotypes influence women's susceptibility to stereotype threat. *Journal of Experimental Social Psychology* 43, 5 (2007), 825–832. https://doi.org/10.1016/j.jesp.2006.08.004

[64] Royce Kimmons and George Veletsianos. 2014. The fragmented educator 2.0: Social networking sites, acceptable identity fragments, and the identity constellation. *Computers & Education* 72 (March 2014), 292–301. https://doi.org/10.1016/j.compedu.2013.12.001

[65] Barbara Kitchenham. 2004. Procedures for Performing Systematic Reviews. *Keele, UK, Keele Univ.* 33 (Aug. 2004).

[66] Maria Knobelsdorf. 2011. *Biographische Lern- und Bildungsprozesse im Handlungskontext der Computernutzung.* Ph.D. Dissertation. http://dx.doi.org/10.17169/refubium-8680

[67] Siu Cheung Kong and Yi Qing Wang. 2020. Formation of computational identity through computational thinking perspectives development in programming learning: A mediation analysis among primary school students. *Computers in Human Behavior* 106 (2020), 106230. https://doi.org/10.1016/j.chb.2019.106230

[68] Lyda Lannegrand-Willems and Harke A Bosma. 2006. Identity development-in-context: The school as an important context for identity development. *Identity* 6, 1 (2006), 85–113.

[69] Jean Lave and Etienne Wenger. 1991. *Situated Learning: Legitimate Peripheral Participation.* Cambridge University Press, Cambridge. https://doi.org/10.1017/CBO9780511815355

[70] Luzia Leifheit, Katerina Tsarava, Korbinian Moeller, Klaus Ostermann, Jessika Golle, Ulrich Trautwein, and Manuel Ninaus. 2019. Development of a Questionnaire on Self-concept, Motivational Beliefs, and Attitude Towards Programming. In *Proceedings of the 14th Workshop in Primary and Secondary Computing Education (WiPSCE'19)*. Association for Computing Machinery, New York, NY, USA, 1–9. https://doi.org/10.1145/3361721.3361730

[71] Luzia Leifheit, Katerina Tsarava, Manuel Ninaus, Klaus Ostermann, Jessika Golle, Ulrich Trautwein, and Korbinian Moeller. 2020. SCAPA: Development of a Questionnaire Assessing Self-Concept and Attitudes Toward Programming. In *Proceedings of the 2020 ACM Conference on Innovation and Technology in Computer Science Education (ITiCSE '20)*. Association for Computing Machinery, New York, NY, USA, 138–144. https://doi.org/10.1145/3341525.3387415

[72] Robert Lent, Steven Brown, and Gail Hackett. 1994. Toward a Unifying Social Cognitive Theory of Career and Academic Interest, Choice, and Performance. *Journal of Vocational Behavior* 45 (Aug. 1994), 79–122. https://doi.org/10.1006/jvbe.1994.1027

[73] Jonathan Mahadeo, Zahra Hazari, and Geoff Potvin. 2020. Developing a Computing Identity Framework: Understanding Computer Science and Information Technology Career Choice. *ACM Trans. Comput. Educ.* 20, 1, Article 7 (Jan. 2020), 14 pages. https://doi.org/10.1145/3365571

[74] Adam V. Maltese and Robert H. Tai. 2010. Eyeballs in the Fridge: Sources of early interest in science. *International Journal of Science Education* 32, 5 (2010), 669–685. https://doi.org/10.1080/09500690902792385

[75] Jane Margolis and Allan Fisher. 2002. *Unlocking the clubhouse: Women in computing.* MIT press.

[76] Herbert W. Marsh, Reinhard Pekrun, Philip D. Parker, Kou Murayama, Jiesi Guo, Theresa Dicke, and A. Katrin Arens. 2019. The murky distinction between self-concept and self-efficacy: Beware of lurking jingle-jangle fallacies. *Journal of Educational Psychology* 111, 2 (feb 2019), 331–353. https://doi.org/10.1037/edu0000281

[77] Danny Bernard Martin. 2000. *Mathematics success and failure among African-American youth: The roles of sociohistorical context, community forces, school influence, and individual agency.* Routledge.

[78] Danny Bernard Martin. 2012. Learning mathematics while Black. *Educational Foundations* 26 (2012), 47–66.

[79] Allison Master, Sapna Cheryan, and Andrew N Meltzoff. 2016. Computing whether she belongs: Stereotypes undermine girls' interest and sense of belonging in computer science. *Journal of educational psychology* 108, 3 (2016), 424.

[80] Philipp Mayring. 2014. *Qualitative content analysis: theoretical foundation, basic procedures and software solution.* AUT.

[81] George H Mead. 1913. The social self. *The Journal of Philosophy, Psychology and Scientific Methods* 10, 14 (1913), 374–380.

[82] Wim Meeus, Jurjen Iedema, Marianne Helsen, and Wilma Vollebergh. 1999. Patterns of adolescent identity development: Review of literature and longitudinal analysis. *Developmental review* 19, 4 (1999), 419–461.

[83] E. M. Mercier, B. Barron, and K. M. O'Connor. 2006. Images of self and others as computer users: the role of gender and experience. *Journal of Computer Assisted Learning* 22, 5 (2006), 335–348. https://doi.org/10.1111/j.1365-2729.2006.00182.x arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1365-2729.2006.00182.x

[84] John R. Morelock. 2017. A systematic literature review of engineering identity: definitions, factors, and interventions affecting development, and means of measurement. *European Journal of Engineering Education* 42, 6 (Nov. 2017), 1240–1262. https://doi.org/10.1080/03043797.2017.1287664

[85] Michele S. Moses, Kenneth R. Howe, and Tricia Niesz. 1999. The Pipeline and Student Perceptions of Schooling: Good News and Bad News. *Educational Policy* 13, 4 (Sept. 1999), 573–591. https://doi.org/10.1177/0895904899013004005

[86] Erica Mourão, João Felipe Pimentel, Leonardo Murta, Marcos Kalinowski, Emilia Mendes, and Claes Wohlin. 2020. On the performance of hybrid search strategies for systematic literature reviews in software engineering. *Information and Software Technology* 123 (2020), 106294.

[87] Na'ilah Suad Nasir and Jamal Cooks. 2009. Becoming a Hurdler: How Learning Settings Afford Identities. *Anthropology & Education Quarterly* 40, 1 (2009), 41–61. https://doi.org/10.1111/j.1548-1492.2009.01027.x

[88] Greg L. Nelson and Andrew J. Ko. 2018. On Use of Theory in Computing Education Research. In *Proceedings of the 2018 ACM Conference on International Computing Education Research* (Espoo, Finland) *(ICER '18)*. Association for Computing Machinery, New York, NY, USA, 31–39. https://doi.org/10.1145/3230977.3230992

[89] Lijun Ni and Mark Guzdial. 2012. Who AM I? understanding high school computer science teachers' professional identity. In *Proceedings of the 43rd ACM technical symposium on Computer Science Education (SIGCSE '12)*. Association for Computing Machinery, New York, NY, USA, 499–504. https://doi.org/10.1145/2157136.2157283

[90] Brian A Nosek, Mahzarin R Banaji, and Anthony G Greenwald. 2002. Math = male, me = female, therefore math≠ me. *Journal of personality and social psychology* 83, 1 (2002), 44.

[91] J. Steve Oliver and Ronald D. Simpson. 1988. Influences of attitude toward science, achievement motivation, and science self concept on achievement in science: A longitudinal study. *Science Education* 72, 2 (1988), 143–155. https://doi.org/10.1002/sce.3730720204

[92] Daphna Oyserman, Deborah Bybee, and Kathy Terry. 2006. Possible selves and academic outcomes: How and when possible selves impel action. *Journal of Personality and Social Psychology* 91, 1 (July 2006), 188–204. https://doi.org/10.1037/0022-3514.91.1.188

[93] Daphna Oyserman, Deborah Bybee, Kathy Terry, and Tamera Hart-Johnson. 2004. Possible selves as roadmaps. *Journal of Research in personality* 38, 2 (2004), 130–149.

[94] Daphna Oyserman, Kristen Elmore, and Smith. 2012. Self, self-concept, and identity. In *Handbook of self and identity.* 69–104.

[95] Anne-Kathrin Peters. 2017. Learning Computing at University: Participation and Identity : A Longitudinal Study. (2017).

[96] Anne-Kathrin Peters and Detlef Rick. 2014. Identity development in computing education: theoretical perspectives and an implementation in the classroom. In *Proceedings of the 9th Workshop in Primary and Secondary Computing Education (WiPSCE '14)*. Association for Computing Machinery, New York, NY, USA, 70–79. https://doi.org/10.1145/2670757.2670787

[97] Patrice Potvin, Abdelkrim Hasni, Ousmane Sy, and Martin Riopel. 2018. Two Crucial Years of Science and Technology Schooling: A Longitudinal Study of the Major Influences on and Interactions Between Self-Concept, Interest, and the Intention to Pursue S&T. *Research in Science Education* (Aug. 2018). https://doi.org/10.1007/s11165-018-9751-6

[98] Lilian L. Pozzer and Phoebe A. Jackson. 2015. *Conceptualizing Identity in Science Education Research: Theoretical and Methodological Issues.* Springer Netherlands, Dordrecht, 213–230. https://doi.org/10.1007/978-94-007-4240-6_11

[99] Darinka Radovic, Laura Black, Julian Williams, and Christian E. Salas. 2018. Towards conceptual coherence in the research on mathematics learner identity: a systematic review of the literature. *Educational Studies in Mathematics* 99, 1 (Sept. 2018), 21–42. https://doi.org/10.1007/s10649-018-9819-2

[100] Laurie Hart Reyes. 1984. Affective Variables and Mathematics Education. *The Elementary School Journal* 84, 5 (1984), 558–581. https://doi.org/10.1086/461384

[101] Sarah L Rodriguez and Kathleen Lehman. 2017. Developing the next generation of diverse computer scientists: the need for enhanced, intersectional computing identity theory. *Computer Science Education* 27, 3-4 (2017), 229–247.

[102] Mark Santolucito and Ruzica Piskac. 2020. Formal Methods and Computing Identity-based Mentorship for Early Stage Researchers. In *Proceedings of the 51st ACM Technical Symposium on Computer Science Education (SIGCSE '20)*. Association for Computing Machinery, New York, NY, USA, 135–141. https://doi.org/10.1145/3328778.3366957

[103] Toni Schmader. 2002. Gender Identification Moderates Stereotype Threat Effects on Women's Math Performance. *Journal of Experimental Social Psychology* 38, 2 (2002), 194–201. https://doi.org/10.1006/jesp.2001.1500

[104] Anna Sfard and Anna Prusak. 2005. Telling Identities: In Search of an Analytic Tool for Investigating Learning as a Culturally Shaped Activity. *Educational Researcher* 34 (May 2005), 14–22. https://doi.org/10.3102/0013189X034004014

[105] Mia S. Shaw, Deborah A. Fields, and Yasmin B. Kafai. 2019. Connecting with Computer Science: Electronic Textile Portfolios as Ideational Identity Resources for High School Students. *International Journal of Multicultural Education* 21, 1 (March 2019), 22–41. https://doi.org/10.18251/ijme.v21i1.1740

[106] Claude M Steele. 1997. A threat in the air: How stereotypes shape intellectual identity and performance. *American psychologist* 52, 6 (1997), 613.

[107] Amanda Sullivan and Marina Umashi Bers. 2016. Girls, boys, and bots: Gender differences in young children's performance on robotics and programming tasks. *Journal of Information Technology Education: Innovations in Practice* 15, 1 (2016), 145–165.

[108] Hideyuki Suzuki and Hiroshi Kato. 1997. Identity formation/transformation as the process of collaborative learning through AlgoArena. In *Proceedings of the 2nd international conference on Computer support for collaborative learning (CSCL '97)*. International Society of the Learning Sciences, Toronto, Ontario, Canada, 283–291.

[109] Ulrich Trautwein and Olaf Köller. 2003. Was lange währt, wird nicht immer gut. *Zeitschrift für Pädagogische Psychologie* 17, 3/4 (2003), 199–209. https://doi.org/10.1024//1010-0652.17.34.199

[110] Stuart A. Tross, Jeffrey P. Harper, Lewis W. Osher, and Linda M. Kneidinger. 2000. Not Just the Usual Cast of Characteristics: Using Personality To Predict College Performance and Retention. *Journal of College Student Development* 41, 3 (2000), 323–34.

[111] Jessica Tupou and Judith Loveridge. 2019. Understanding Pasifika Students' Engagement with Computer Science: A New Zealand-Based Case Study. *New Zealand Journal of Educational Studies* 54, 2 (Nov. 2019), 277–296. https://doi.org/10.1007/s40841-019-00147-y

[112] Sepehr Vakil. 2020. "I've Always Been Scared That Someday I'm Going to Sell Out": Exploring the relationship between Political Identity and Learning in Computer Science Education. *Cognition and Instruction* 38, 2 (2020), 87–115. https://doi.org/10.1080/07370008.2020.1730374 arXiv:https://doi.org/10.1080/07370008.2020.1730374

[113] Monique Verhoeven, Astrid M. G. Poorthuis, and Monique Volman. 2019. The Role of School in Adolescents' Identity Development. A Literature Review. *Educational Psychology Review* 31, 1 (2019), 35–63. https://doi.org/10.1007/s10648-018-9457-3

[114] Robin Whittemore, Susan K Chase, and Carol Lynn Mandle. 2001. Validity in qualitative research. *Qualitative health research* 11, 4 (2001), 522–537.

[115] Jesse L. M. Wilkins. 2004. Mathematics and Science Self-Concept: An International Investigation. *The Journal of Experimental Education* 72, 4 (July 2004), 331–346. https://doi.org/10.3200/JEXE.72.4.331-346

[116] Jeannette M. Wing. 2006. Computational thinking. *Commun. ACM* 49, 3 (2006), 33. https://doi.org/10.1145/1118178.1118215

[117] Claes Wohlin. 2014. Guidelines for Snowballing in Systematic Literature Studies and a Replication in Software Engineering. In *Proceedings of the 18th International Conference on Evaluation and Assessment in Software Engineering* (London, England, United Kingdom) *(EASE '14)*. Association for Computing Machinery, New York, NY, USA, Article 38, 10 pages. https://doi.org/10.1145/2601248.2601268

[118] Billy Wong. 2016. 'I'm good, but not that good': digitally-skilled young people's identity in computing. *Computer Science Education* 26, 4 (2016), 299–317. https://doi.org/10.1080/08993408.2017.1292604 arXiv:https://doi.org/10.1080/08993408.2017.1292604

[119] Billy Wong and Peter EJ Kemp. 2018. Technical boys and creative girls: the career aspirations of digitally skilled youths. *Cambridge Journal of Education* 48, 3 (2018), 301–316.

[120] Yi Xue and Richard C Larson. 2015. STEM crisis or STEM surplus? Yes and yes. *Monthly labor review* 2015 (2015).

A.4 Identity in Higher Computer Education Research: A Systematic Literature Review

Gregor Große-bölting, Dietrich Gerstenberger, Lara Gildehaus, Andreas Mühling, and Carsten Schulte. 2023. Identity in Higher Computer Education Research: A Systematic Literature Review. ACM Trans. Comput. Educ. 23, 3, Article 35 (September 2023), 39 pages. `https://doi.org/10.1145/3606707`

Note: The page numbering of the following pages is taken from the original publication and does not correspond to the numbering of this thesis. The page numbering returns to the numbering of the thesis after this publication.

Identity in Higher Computer Education Research: A Systematic Literature Review

GREGOR GROSSE-BÖLTING, Kiel University, Germany
DIETRICH GERSTENBERGER, Paderborn University, Germany
LARA GILDEHAUS, Paderborn University, Germany
ANDREAS MÜHLING, Leibniz Institute for Science and Mathematics Education, Germany and Kiel University, Germany
CARSTEN SCHULTE, Paderborn University, Germany

The disciplinary identity as a computer science student has recently received increasing attention as a well-developed subject identity can help with increasing retention, interest and motivation. Besides, identity theory can serve as an analytical lens for issues around diversity. However, identity is also often perceived as a vague, overused concept with a variety of theories to build upon. In addition, connections to other topics, such as computer science conceptions, remain unclear and there seems to be little intra-disciplinary exchange about the concept. This article therefore attempts to provide a starting point by presenting a so far missing systematic literature review of identity in Computing Education Research (CER). We analyzed a corpus of 41 papers published since 2005 with a focus on the variety of identity theories that are used, the reasons for using them and the overall theoretical framing of the concept in the CER literature up to this point. We use content analysis with both inductive and deductive coding to derive categories from the corpus to answer our research questions. The results show that there is less variety in the theories than originally expected, most publications refer to the theory of "Communities of Practice". The reasons for employing identity theory are also rather canonical, in particular, there is only little theoretical development of the theories within CER and also only little empirical work. Finally, we also present an extended version of a computing identity that can be theoretically derived from the work in our corpus.

CCS Concepts: • **Social and professional topics → Computer science education**.

Additional Key Words and Phrases: Computer Science Education, Identity, Theory, Systematic Literature Review

1 INTRODUCTION

The construct of computer science students' identity has received increasing attention in recent years, reflected in the number of publications around the role and function of identity. Indeed, it has been demonstrated many times that a strong subject identity increases retention and thus minimizes the risk of dropout [122, 130], just as a link between a profound self-concept and interest in the subject has been shown [32, 59, 73, 98] (see section 2.3). Regarding diversity, identity theory frameworks prove their worth through an analytical approach that seeks to understand and explain the individual in their social context, thereby opening up new perspectives. As Darragh puts it (with respect to mathematics education): "Identity is a lens that is adjustable; one can zoom in to the level of interactions between individuals or zoom out to look at the wider socio-political context. We can look at the big picture, that is, at issues of mathematics learning in general. We can look at the experiences of specific groups of people and issues of equity. Or we can look at the individual level and try to understand learners' relationships with mathematics." [27]

Authors' addresses: Gregor Große-Bölting, Kiel University, Kiel, Germany, ggb@informatik.uni-kiel.de; Dietrich Gerstenberger, Paderborn University, Paderborn, Germany, dietrich.gerstenberger@uni-paderborn.de; Lara Gildehaus, Paderborn University, Warburger Str. 100, Paderborn, Germany, gildehaus@khdm.de; Andreas Mühling, Leibniz Institute for Science and Mathematics Education, Olshausenstraße 62, Kiel, Germany, 24118, Kiel University, Kiel, Germany, muehling@leibniz-ipn.de; Carsten Schulte, Paderborn University, Paderborn, Germany, carsten.schulte@uni-paderborn.de.

Identity occasionally appears in this context like a “silver bullet” that can solve all problems currently preoccupying computing and computer science education: Student interest, dropout, diversity, motivation, skill development, and more. However – in keeping with the old adage, especially popular among programmers, that “there are no silver bullets” – it has been noted just the same that identity is a vague, overused concept, of which a variety of theories and even labels exist [38]. Its connection to other topics, such as computer science conceptions, which have been addressed by various authors in recent years [7, 50, 54], is unclear, although there is some clear overlap. Although identity is – as our research shows – widely used as a theoretical framework, there seems to be only little intra-disciplinary exchange about the concept. This may be due to the fact that, unlike other subjects related to computer science, such as mathematics [27, 99] or engineering [20, 82], there has been no compilation and synthesis of the existing literature in the field.

We hereby attempt to provide a starting point for such an intra-disciplinary examination by presenting a systematic literature review (SLR) on identity in computing education research (CER). Following a hybrid approach, we have both undertaken a database search in the most popular databases and – to do justice to the perceived vagueness of the concept – extended it to include a snowballing process. As a result, we were able to identify 41 articles dealing with identity for higher education. Further evaluation of these articles is the subject of the rest of this paper. We already have taken a first step towards providing this missing synthesis with a systematic literature review of publications on computer science identity specifically in the K-12 context [51], as these two domains differ enough to warrant a separate analysis: As a university student, the development of a subject identity may play an even more important role than in school; after all, one has voluntarily chosen to study the subject intensively and is planning a possibly lifelong career in computer science. It seems unlikely that this is possible without identifying with the community of computer scientists and the subject. But what exactly does that even mean?

To answer this overarching question, we approach the literature from a theoretical or theorizing standpoint, examining ways in which different identity theories are used within research and how (subject) identity is understood within them. This provides a point of reference for the adaptation of different theories, their further development within the discipline, and the specifics in relation to university teaching. The research questions are:

- **RQ1** How is the concept of identity used and developed?
- **RQ2** Why is an identity theory used as a theoretical framing?
- **RQ3** How is the identity concept understood across different theories in computing education research?
  - **RQ3.1** Which theoretical identity frameworks are used?
  - **RQ3.2** How is identity understood and conceptualized?
  - **RQ3.3** What specific psychological attributes and socio-cultural practices are associated with a computer science identity?

As we conceived the systematic literature review as an exploration of ways of understanding identity within computing education research, the research questions were initially broad. The third question was differentiated in the course of coding and discussion of interim findings and as a result of our iterative inquiry into the research field.

2 THEORETICAL BACKGROUND

In order to keep the rest of the presentation as concise and clear as possible, we will first define some terms that serve as the basis for the description of our results:

A **concept** serves to describe the meaning of an identifier or an idea and forms a semantic unit. In other words: “Concepts are the building blocks of thoughts.” [76] However, for the definition of a concept, almost inevitably (cf.

"Münchhausen trilemma") further concepts are used, so that a single definition of a concept can be regarded as a structure of references. Concepts serve the description of phenomena and explanation of facts of the real world, but must neither represent these adequately, nor even offer the possibility to be verifiable by measurement. The process of describing a concept by specifying defining properties is described as **conceptualization** [76].

A **construct** differs from this in that it also represents a description of the real world, but does so in a measurable, albeit latent, way: "A construct is some postulated attribute of people, assumed to be reflected in test performance." [23] Thus, the construct can directly serve as the basis of the development of an instrument in which it maps the construct as a whole or subconstructs in an objectively ascertainable way. The process of inferring a construct is referred to as **operationalization** [23].

Understood like this, not all concepts are also constructs, but every construct describes or relates to one (or more) concept(s). Therefore, the difference mainly relevant in the following is in the usage: Constructs represent a necessary abstraction for the use in instruments, e.g., for psychometric studies. Concepts are more general and do not necessarily require empirical evidence. They are the subject of theoretical discussion and qualitative categorization.

As a starting point for our exploration of identity, we have oriented ourselves to the definition of Gee [42]: Identity describes "being recognized as a certain 'kind of person,' in a given context". In order to do justice to the diversity of identity theories, we have tried to include all papers that deal with identity in a broader sense (and even going beyond Gee's definition) in our literature review and address further methodological considerations and limitations in section 3. The following two sections develop a more complex portrait of identity: 2.1 traces the historical development of identity in the educational sciences, while 2.2 takes a more abstract viewpoint and explores conceptual differences among theories (independent of historical dependencies). The theoretical background is completed by a look into related disciplines, as well as the results of the systematic literature review we have already provided in the K-12 field.

2.1 Historical Overview

The concept of identity has a long history. Fukuyama [40] suggests that this history begins with Plato's Politea, while other authors consider Aristotle as the origin [24]. In educational studies, the roots of the discussion of identity are the works of Vygotsky and Mead [79], and to a lesser extent Piaget, whose developmental model was greatly expanded in an identity theory direction by Erikson. Erikson's "Identity: Youth and Crisis" [35], in turn, is something of a *big bang* for the psychological discussion of identity.

In the early 1990s, Lave & Wenger's "Situated Learning" [67] was published, the writing that led to an increased, socio-cultural examination of the concept of identity and subsequently inspired numerous authors to further explore the topic, but also to delimit and criticize it. For example, Sfard & Prusak criticize the lack of a clear definition of identity by Lave & Wenger, although this is one of their core concepts. As a consequence, this makes operationalization and application difficult [110]. Lave & Wenger clearly state their theoretical roots: their theory goes back to the activity theory of Engeström [33] and the social constructivism of Vygotsky, respectively.

Facing these varying theories with different roots, several divisions and placements were discussed lately to gain orientation within the field: For example, Verhoeven et al. [125] investigated the role of school on adolescents identity development with a comprehensive literature review. The authors explored the theoretical perspectives by distinguishing between socio-cultural, psychosocial, social psychological and sociological theories as well as combinations thereof. Darragh [27] discussed different theoretical frameworks and identity definitions in the field of mathematics learner's identity. She identified mainly socio-cultural frameworks: Wenger [128], Holland and colleagues [56] as well as Gee [42] as the largest influences on identity research from outside of the discipline. Within these theoretical frameworks, she

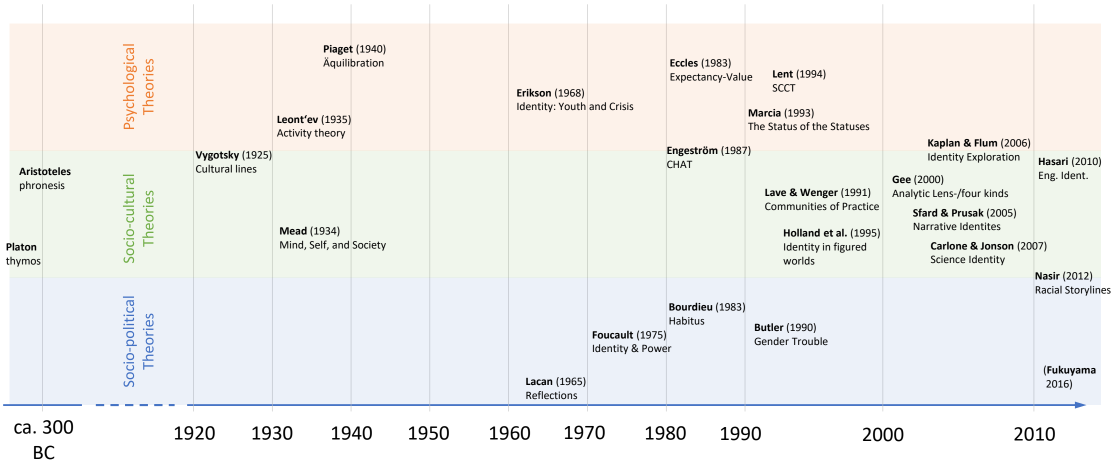


Fig. 1. Historical overview of the development of identity theories.

provided the summarized definition of identity as participatory, narrative, discursive, psychoanalytic or performative, bringing socio-cultural and psychological frameworks together.

For our overview shown in Figure 1 we decided to situate identity theories as follows: There are psychological, socio-cultural, and socio-political theories, whose main representatives (in the context of educational science) are shown in the figure. Psychological theories are characterized by the fact that they conceive of identity primarily as individual attributes of a person, which often have the advantage of being observable and measurable. In contrast, socio-cultural theories view identity as shaped by a social community and culture. Accordingly, identity is based on the interrelationship between the individual and their social context (e.g. community), which consists not only of a set of individuals, but also of the traditions, behaviors, etc. lived by the individuals. An individual can have different identities according to his or her (non-)belonging to different social groups, which can be in harmony or conflict with each other. Socio-political theories, in turn, look at power and dependency relations to describe the formation of identity. In many cases, it is about describing processes of exclusion or inclusion of an individual in a group to which identification is linked. The perspective thus shifts to group and power dynamics, so that Verhoeven et al. [125] describe this form of theory as "sociological perspectives".

This division into three major strands is, of course, a simplification chosen for this review and the placement of theories in Figure 1 follows from a discussion of the authors. All these perspectives are – qua their status as *perspectives* – not without overlap: the psychological factors are just as tied to cultural factors as socio-cultural theories can be about power. The purpose of this classification is to provide orientation based on the *main* foci of the respective theories.

2.2 Conceptual Overview

In addition to the historical roots and structures of identity, we focus on different conceptualizations of identity: Naturally, different conceptualizations are usually found for different theoretical approaches, but even within one theoretical approach, certain foci and aspects can be elaborated and conceptualized differently. For this purpose, we mainly adapt the proposed categorization of Radovic et al. [99] which they developed for the conceptualization of mathematics learners (see section 3). We use it to situate the various identity theories and thus discuss in more detail here:

As a first step, Radovic et al. [99] distinguish different dimensions of how identity can be understood, before characterizing different conceptualizations along their position within these dimensions. Those dimensions are subjective-social, representational-enacted and stability-change. These should not be understood as separable categories with fixed poles, but as a possibility of orientation along focal points: “Subjective” describes identity as a private experience of who one is, e.g. one’s sense of a place in the world or self-descriptions. On the other hand, “social” describes identities as social products, e.g. something constituted by social discourses or performed and recognized in social practice. “Representational” describes identity as mediated by discourse or language, e.g. (self-)concepts, discourses, narratives or stories about oneself, while “enacted” describes identity as engagement in action, e.g. ways of being in actions, forms of participation, or roles performed during activities. “Stability” and “change” describe identity as relatively stable personal factors versus identity being constructed through a process, learned, and open to change.

Along those dimensions, different conceptualizations are distinguished [99]: Identity as constituted by individual attributes refers to a rather subjective-social, representational and stable view of identity and is often operationalized as level of agreement with different statements of oneself, for example in surveys. Social worlds are taken into account, but separately from their context. This conceptualization often describes “having/agreeing with/or expressing” a certain level of a given identity. Identity as a relationship with a specific practice is based on a representational, social-subjective, changeable identity conceptualization. Identities are operationalized as a sense of belonging or membership to and within the discipline. Identity development is described as a negotiation of meanings within shared and collective practices. The relationships of the students with such particular practices becomes the center of this conceptualization. Surveys and mainly interviews are therefore a commonly used method. Identity as narratives refers to a representational and subjective-social conceptualization of identity which is operationalized in so called narratives. Identity is here seen to unfold in time, hence these conceptualizations give high emphasis on the change dimension. Furthermore, Radovic et al. also categorize identity as ways of acting (enacted/social-subjective, fluid) and identities as afforded and constrained by local practices (representational/social). However, both in transition (between K-12 and higher education and in the context of mathematics education) and in higher education, individual attributes, relationships with a specific practice and narratives are found to be the most common conceptualizations, while the other conceptualizations are more frequently used within K-12 contexts [99].

Those dimensions and conceptualizations stem from mathematics education, but can be identified in related work as well: For example, Rodriguez and Lehman [103] use the term *computing identity* to describe a professional identity as a computer scientist and for the identity of the discipline itself the term *computing disciplinary identity*, which indicates the subjective and social dimension of the phenomenon. On the other hand, conceptualizations, close to the concept of identity, can also be found under different names: Rodriguez and Lehman state that the concept *sense of belonging* has a close connection to identity, since it describes the subjective affiliation of an individual to a social group. In addition, there is the term self-concept, which is used in particular in psychological literature [87]. Other concepts similar to identity are referred to as self-image or self-perception.

2.3 Empirical Overview

Kapoor and Gardner-Mccune [62] examine 55 articles on identity in Undergraduate Computing Identity. They consider not only identity itself but also related concepts as inclusion criteria for their corpus. The authors note a trend in research of identity over the past five years. Their study primarily examines the thematic orientation of the articles considered and finds two main themes: Identity-centered studies and non-identity-centered studies, each of which breaks down into subgroups. The largest subgroup (36) represents a number of studies within the identity-centered studies that attempt to explore the connection of identity with other factors. Another large group (23) of articles looks at different descriptions and conceptualizations of identity in computing. The authors conclude that more methodological and conceptual uniformity, reuse of developed tools and further development of a common vocabulary for talking about identity would be desirable.

The potentials, as well as the problems, of the concept of identity have been discussed in other, related disciplines over time, with different empirical foci. As discussed in the previous historical and conceptual overviews identity is often conceptualized as individual attribute within psychology and used as a framework to explain retention and motivation in the context of values and self-perception. [31, 32, 44]. Operationalizing attainment value within the expectancy-value framework as identity based, may predict student's dropout intention from science, mathematics and engineering [101, 102, 107].

Besides the mentioned literature reviews on identity in mathematics education [27, 99] recent research often conceptualizes identity as narratives or ways of acting. Individuals' relationships and identity development with mathematics and related consequences are focused, e.g. identity in rejection, re-fusion or disinterest of mathematics [19, 45] as well as consequences for gender-based identities and mathematics learning in schools [39] or ethnicity and identity and how it is dealt with in the mathematics classroom [55]. In addition, there is also theoretical work that goes beyond an empirical contribution and discusses the role of narratives as identity-creating elements within the discipline [110].

Similar to mathematics, identity in science has been heavily discussed, facing similar challenges around the diversity of the concept and a its coherent conceptualization [4, 71]. The work of Carlone [15, 17] as well as Carlone and Johnson [16] showed a significant influence in this field. Their identity conceptualization is mainly based on an anthropological approach that we would frame as social-cultural, mainly focused on narratives and local practices. The main empirical works building on this theory use identity as an analytical lens and investigate equity in science learning [71].

A recent review by Danielsson et al. [26] look at the identity turn in science education research by examining 198 articles in the field. These articles were divided by the authors into three groups: (A) Macro-studies within a psychological tradition, (B) Macro-studies within a sociological tradition and (C) Micro-studies within an interpretative tradition, with the last group being by far the most extensive (146) and breaking down into two subgroups. The difference between macro- and micro-studies is that macro-studies try to make universal and generalizable statements, while micro-studies aim at a small scale investigation or intervention. For the most part, group C consists therefore of qualitative studies that undertake detailed investigations of the identity of a small group of students. Danielsson et al. also note that the study of identity has increased in recent years. The authors criticize a strong methodological homogeneity found in the articles, which is often based on case studies with a small group of people. In addition, the conceptionalization of science identity is often unclear or allows for little continuity with existing research. The authors conclude with three practical recommendations for researchers working on identity theories.

In addition to science, there is also work on engineering identity:

Morelock published a systematic literature review on engineering identity [82] in which he analyzed 46 studies. He evaluates which definitions of identity are used in the field of engineering and assigns them to different categories. In the largest category, the authors make use of an identity framework based on how the construct has been conceptualized in previous research. A smaller number of papers define engineering identity "in the context of larger collectives, including nations and cultures" [82]. Further, Morelock summarizes studies that interacted with students to examine the characteristics they associate with engineers. Morelock ends his review with recommendations for further research in the area of engineer identity. He would like to see studies “that bridge the gaps between the professional, collective, and developmental psychology perspectives on engineering identity” [82].

Patrick and Borrego also provide a review of the literature on identity relevant to the field of engineering [90]. Many of the studies refer to the Multiple Identity Framework, which is due to Gee [42], Tate [118], and Capobianco [14], for example. Here, identity does not describe how a person perceives self, but “individuals project different parts of their identity as dependent on the environment and context” [90]. They also reiterate that definitions of engineering identity vary widely, even when studies cite the same sources and found studies that refer to less cited theories, e.g. Dutton’s Organizational Identity [29] or the Identity Stage Theory of Erikson [35] and Arnett [3], and Matusovich.

Clark and Kajfez [20] summarize the literature on K-12 engineering education. Their analysis groups studies into formal (more specifically, school types) and informal settings. The authors summarize that in the K-12, hands-on engineering activities contribute to the development of an engineering identity, regardless of gender. The authors cite addressing students as engineers and applying real-world engineering processes as particularly effective interventions. They suggest that early introduction to engineering, especially among historically marginalized, improves the development of an engineering identity.

Rodriguez et al. summarize the state of research on engineering identity development in higher education in their literature review [104]. They emphasize that many studies discuss engineering identity in the context of other, more established theories. Among others, they classify studies that focus on the study of gender, race, and the influence of intersectional identities in engineering disciplines. About the underlying theories they write: “These critical frameworks focused on culture, discourse, agency, and engagement as key aspects to the identity development process” [104].

2.4 Previous Work: Identity in K-12 CER

In a previous review [51], we summarized the use of the concept of identity in computing education research as it relates to the K-12 field. The results of this study are reproduced here in short for the sake of completeness and because we will return to them later.

The papers from the present review and the K-12 review were identified in the same search process. The division of the overall corpus into two sub-corpora was done because K-12 and university education differ significantly in terms of identity formation processes: While in school one is confronted with a variety of subjects, university education presents a much greater degree of choice in what one studies and how much time to devote to a specific subject. Moreover, the K-12 period coincides with the stage of life when identity formation occurs in the first place through increasing self-reflection [38, 109]. Finally, the research places different foci on K-12, which is about sparking interest in a subject, and university education, which is often about preventing dropping out of a subject.

With these differences in mind, similar research questions were addressed, but different emphases were used. Methodologically, the evaluations are also very similar (see Sec. 3), with the biggest differences being in terms of coding. The subcorpus for K-12 includes 31 articles published between 1997 and 2020, but the majority were published in the last ten years (26 between 2010 and 2020). The most represented theoretical approaches are sociocultural (13, including 10x Lave and Wenger) and psychological (8) theories. Three papers use sociopolitical or sociological approaches, and four do not use explicit theoretical framing. Three main categories of how identity theory is used in the papers were identified: Deepening understanding, developing instruments and evaluating interventions. The first category was further divided into two subcategories: First, articles where the relation to other concepts is elaborated, and second, articles that explore the disciplinary identity for specific groups. The motivation of authors for using identity was also coded, and five rationales were identified: Diversity, Retention, Promotion of Young Talent, Improved Learning/Better Performance, and Higher Motivation. Of these, Diversity was cited by far the most (19), ahead of Retention (7), Promotion of Young Talent (7), and Improved Learning (7). Regarding the theoretical frameworks used, a low diversity in the use of theories is striking, as is the lack of custom developments and discipline-specific adaptations. Furthermore, the strengths and weaknesses of the theories employed are rarely discussed, i.e. a theory is taken for granted without critically discussing it. Regarding the use of the concept of identity, much focus has been on issues of diversity, that is, how identity as a computer scientist can be brought together with other identities and affect specific groups. In this context, the concept shows its strengths as a "missing link" [110] between individual and social development because it shifts the focus from the individual and their possibly deficient performance to the social environment. Although three instruments for measuring identity were found in the corpus [13, 65, 68], there is a great lack of reliable and empirically evaluated interventions.

Overall, what seems to be specific about the concept of identity in schools, and especially in relation to computer science, is that it opens up the analysis of a "hidden curriculum" [125]. Demarcation phenomena and stereotypes about computer science and computer scientists set in motion developments that continue in higher education in the form of a negative feedback loop. In the end, while computer science presents a way of identification for certain groups of people, it explicitly does not for an even larger group of people, who in turn don't develop interest or even are actively deterred because they precisely do not (want to) identify themselves as computer scientists. This has a major impact on who studies computer science and is successful in their studies. And this, in turn, might reinforce and perpetuate the existing image and identity of computer science.

3 METHODS

Our systematic literature review (SLR) follows a hybrid approach, using both a classic database search process and *snowballing*. The database search follows the method introduced by Kitchenham [64]. It is accompanied by both *forward* and *backward snowballing* [131]: These start from a corpus of carefully selected literature – the result from the database search – and then search for further literature via its references; see Sec. 3.3 for details. It was decided to choose this work-intense process to do justice to the identified multifacetedness of the concept of identity.

For the same reason the search was extended after a first run: In the course of the initial search and after evaluation of the first papers (as described in section 3.1 and 3.2) it became clear again that there are large intersections between identity and self-concept, self-perception and self-image (s. Sec. 2.2). Although these are by no means congruent concepts and the relation between the concepts is hard to grasp [5, 6], the search was subsequently extended to include these terms. Since the social understanding of identity emerged as one dominant theory, the concept of *sense of belonging* was also eventually included.

3.1 Database Search Strategy and Selection Process

The database search was automated, which is natural given the almost exclusively digital publication practice in computer science. For this purpose, the following databases were considered: ACM Digital Library[1], IEEE Xplore[2],

---

[1] https://dl.acm.org/

[2] https://ieeexplore.ieee.org

SpringerLink[3], ScienceDirect[4], DBLP[5], and Google Scholar[6]. The search was conducted without limiting the search to specific years in order to be able to identify trends in the development of publication on the topic of identity. However, the search was limited to title and abstract. The following search query was initially conducted in March 2020:

```
identity AND (education OR k12 OR students OR cs1 OR learn) AND (”computer science” OR ”computer
engineering” OR informatics OR computing)
```

After the original query proved to be too narrow, the search was extended by the results of the following query in May 2020:

```
(”self-concept” OR ”self-image” OR ”self-perception” OR ”sense of belonging”) AND (education OR
k12 OR students OR cs1 OR learn) AND (”computer science” OR ”computer engineering” OR informatics
OR computing)
```

Search results were exported from the databases and articles with less than four pages were removed to easily exclude abstracts only, panel papers, and posters. The resulting list of hits was independently coded by two of the authors: -1 (reject), 0 (discuss or investigate further), 1 (accept). The results of the coding were merged and conflicts, as well as uncertainties (papers coded 0), were resolved through discussion and, if necessary, a review of the article. In the further course of the coding, papers were also removed if it turned out that the content did not correspond to the topic despite an allegedly suitable title and abstract. Such decisions were again made collectively and not by one author alone.

Because of their focus on computer science and engineering, the last component of the search phrase (“computer science OR ...”) was omitted from the ACM, IEEE, and DBLP searches. However, papers were excluded if they explicitly did not deal with computer science or had no clear relation to the discipline. Finally, the list of results was examined to see whether the title or abstract suggested a K-12 context. These articles were not considered in this review but, as mentioned, were analyzed in a separate literature review [51].

3.2 Data Extraction and Coding

The papers that initially resulted from the database search were imported into MAXQDA 2020. Three papers were selected and coded by three coders (first three of the authors), then the codings were compared and discussed to develop a common understanding. Subsequently, the entire corpus was divided into three parts and each part was analyzed by one coder. Regular meetings were held to clarify questions, conflicts, and disputed cases, and to further develop understanding of the codings. In addition, further papers were removed from the corpus that turned out not to fit after all. Mostly this was due to the fact that they either did not have a computer science focus or were only marginally concerned with identity.

The analysis of the articles follows Mayring’s [77] method of qualitative content analysis. Coding was both deductive and inductive, as shown in Tab 1; in case of deductive coding, the source of the category system is indicated. In the case

---

[3] https://link.springer.com/

[4] https://www.sciencedirect.com/

[5] https://dblp.uni-trier.de/

[6] The search was not carried out using the search mask of Google Scholar (https://scholar.google.com/), but by using the tool *Publish or Perish* (see https://harzing.com/resources/publish-or-perish).

of inductive coding, as in the case of inductive expansion of already existing category systems, the relevant passages in the articles were first paraphrased and then discussed between the authors to subsequently combine the paraphrased codings into categories.

Coding
Deductive - Basic Information / Year - Basic Information / Publication - Motivation: K-12 SLR on identity in CER [51] - Identity Dimension: Radovic et al. [99]
Deductive, inductively extended - Theoretical Framework: inspired by Darragh [27], extended by own research (section 2.1, figure 1) - Identity Conceptualization: inspired by Radovic et al. [99], adjusted for Computer Science Education
Inductive - Theory Usage - Theory Development - Individual Attributes - Specific Practices

Table 1. Coding

Regarding the theoretical frameworks, we initially started from Darragh [27], but eventually extended the overview of frameworks considerably through our own research – as shown in section 2.1. Accordingly, the overview there, visualized in Figure 1, forms the background for deductive coding of the frameworks. A similar approach was taken for the identity conceptualizations: Following Radovic et al. [99], whose conceptualization was presented in section 2.2, we used a reduced category system for coding, consisting of only three codes: *As Individual Attributes*, *As Specific Practices*, *Other*. Further insights on the inductively extended and inductive codings are given in the results 4.2 and discussion section 5).

The results of the coding were presented to different groups of researchers in two workshops and discussed with respect to their plausibility and pragmatic usefulness – as a measure of their validity [129]. In both cases, plausibility and utility were confirmed.

3.3 Forward and Backward Snowballing

Based on the seed corpus formed by the database search, *backward* and *forward snowballing* [131] were performed. For the *backward snowballing*, the references of the articles were searched: Further relevant articles were identified based on the title and their use within the paper and included in the corpus. For *forward snowballing*, Google Scholar was used and all papers were considered that had the considered paper as a reference and in turn had ten or more citations. The choice of "ten or more" was made arbitrarily by the authors in order to limit an otherwise unmanageable result set. The snowballing was conducted in January and February 2021. For a discussion of different search strategies and their respective performance, see Mourão et al. [83].

Backward and forward snowballing was repeated until saturation was reached, i.e. until no more new papers could be found as a result. This happened in both cases after two iterations. Further analysis and coding of the papers was done as described above.

4 RESULTS

The following section describes the results of our two search phases.

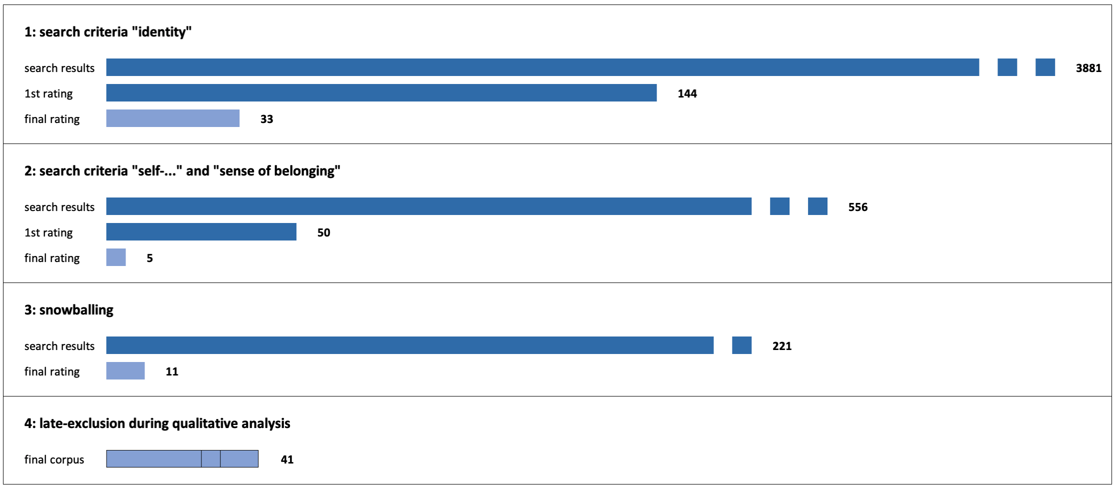


Fig. 2. Initial search results and selection process

The second search incorporating related concepts resulted in 556 articles. The first rating reduced the number of hits to 50 papers. This was followed again by a more detailed review of the previous selection in full text. Here it was a closer look at the concept *Sense of Belonging* after which we decided not to include these papers into our corpus because this area is theorizing on its own, with little in common with the strands of theory otherwise considered in our review. Furthermore, we found many articles from the context of engineering, which we only included in our corpus if they were clearly related to computer engineering or computer science. At the end of this rating phase five new papers were added to our corpus.

The results from snowballing were combined from both searches. Snowballing initially resulted in 221 additional hits, but these were reduced to 11 after applying the same rating process. Thus, more than a quarter of the papers from our corpus came from the additional snowballing process. In total, the corpus consists of the 41 following articles: [9, 18, 22, 25, 28, 41, 52, 57, 60, 61, 63, 66, 70, 72–74, 78, 80, 81, 85, 86, 88, 89, 93–97, 103, 105, 108, 111, 112, 117–121, 123, 126, 127]

Although no restrictions were placed on the year of publication during the search, the oldest articles in our corpus are from 2005. Furthermore, it can be seen that more than half of the articles from the corpus were published in the last five years (Fig. 3). A general overview of the specific papers in the corpus, including a very short summary on their overall concerns and results can be found in the Appendix.

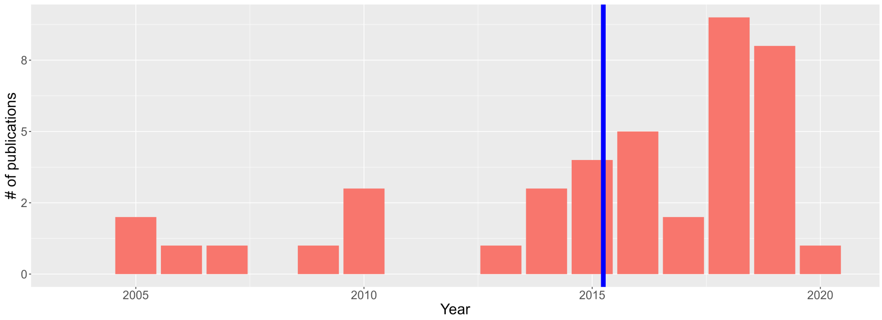


Fig. 3. Publications on computer science identity in higher education per year. The blue line indicates the median.

4.1 RQ1: How is the concept of identity used and developed?

Table 2 shows which papers form the basis for the design of the codes. Four subcodes each could be identified for theory use and theory development. By theory use, we mean that a theory provides the background of an empirical work. In this context, the concept of identity is used as a theoretical framing, but is not developed further in its own right. Theory development, on the other hand, is necessarily accompanied by an expansion or deepening of the concept of identity in one of the ways described below. Only the primary purpose of a paper was considered for coding, so that mixed forms of the following codes can also exist, but were neglected for the analysis.

For theory use the four subcodes are:

As Theoretical Lens: An existing theory (or part of it) is used as a theoretical background to further investigate a phenomenon not directly related to the theory. The concept or construct is not significantly developed, but the understanding of the group or the differences between different groups are developed. Ex: Theory X is used to study different groups, e.g. a minority vs. majority, and to explain observed differences.

As Basis for Interventions: An existing theory (or part) is used as a theoretical background to develop an intervention. Ex: Article describes the development of a lesson or activity, theoretical background is an identity theory. In some cases, an existing instrument is used to measure identity development.

As Explanation for Other Concepts: Identity is used as a theoretical background to explain another concept. This deepens the understanding of the other concept without further differentiating or deepening the used concept of identity itself. Ex: Theory X is used to sharpen concept Y. For example, an identity theory could be used to explain a group's declining motivation due to socio-cultural factors.

Transfer to other Target Group: The article describes the transfer of a theory to a new, not yet investigated target group. Ex: For theory X, it is known that it has been used many times profitably for first-year students (in whatever form). This article shows that the theory also works in the context of graduate students.

For theory development the four codes are:

Concept Refinement: An already existing concept of identity is empirically or theoretically differentiated, i.e., a (partial) aspect of the conceptual definition is examined more closely and something new is found out about it. The conceptualization is refined as a whole. Ex: Theory X is known to have partial aspect Y. The article looks at how subaspect Y can be conceptualized more precisely. Or: Theory X is more precisely defined overall by comparing it with theory A and working out significant differences.

Category	**Count**	**Source papers**
Theory Use		
- As Theoretical Lens	9	[22, 25, 41, 70, 78, 94, 95, 97, 111]
- As Basis for Interventions	7	[61, 85, 88, 96, 108, 118, 123]
- As Explanation for Other Concepts	3	[9, 63, 66]
- Transfer to other Target Group	1	[18]
Theory Development		
- Concept Refinement	7	[28, 52, 57, 60, 89, 93, 112]
- Concept Expansion	8	[52, 72, 81, 103, 105, 118, 120, 123]
- Concept Location	1	[121]
- Construct Definition, etc.	3	[73, 103, 117]

Table 2. Use and development of identity theory.

Concept Expansion: An existing concept of identity is empirically or theoretically expanded to include a new aspect that was not previously part of the concept of identity. An already existing conceptualization is thus expanded. A "concept expansion" is usually also a "concept refinement", but not vice versa. Ex: Theory X previously consisted of subconcepts Y and Z. The article argues that W should also be included.

Concept Location: Identity as a whole (albeit presumably represented by a specific theory) is located within one or more other concepts that have a similar *epistemic scope*. Locating means: it is clarified which fields of application or specifics characterize the terms and which position they thereby occupy in an overarching, theoretical framework. Ex: The relationship between sense-of-belonging and "personal fit" is discussed within identity-theoretical perspectives, thereby determining how identity is understood.

Construct Definition, Refinement or Expansion: An operationalization of an identity concept is presented, improved or extended by new aspects. In contrast to the concept related codes above, this is mainly about the operationalization and measurable construct of the identity theory, s. Sec. 2. Ex: Based on identity concept X, operationalization Y is presented, which has been validated in studies with many participants and CFA etc.

4.1.1 Psychometric Instrument Development. We took note of articles that have developed psychometric instruments and report them here as they may provide an accessible and practical way for some researchers to develop interventions:

The Computer Science Cultural Attitude and Identity Survey (CSAIS) by Washington et al. [127] was developed to measure five constructs that describe the attitude and identity of undergraduate students of color. Parker [89] describes the preliminary results of a survey instrument on professional identity that encourages students to self-reflect from the perspective of their peers and supervisors and conceptualize their future selves based on job aspirations.

The work of Garcia et al. [41] and Mahadeo et al. [73] each build on the tradition of Carlone and Johnson [16] (see sections 2.1, 4.3, and 5.2) and accordingly consider the three subconstructs of *recognition*, *interest*, and *competence*. In Garcia et al. this is done explicitly from the point of view of underserved students.

These same subconstructs are also included in the instrument of Choe and Borrego [18], but extended by the dimension *interpersonal skills competence*, which is newly added for the group of graduate students studied by them (in contrast to undergraduate students, for whom this dimension is not found). Finally, Scott and Ghinea [108] present an instrument that identifies the self-beliefs of undergraduate students using five technical-oriented subconstructs.

A general overview of results to RQ1 is included in the summary of the corpus in the Appendix.

Why?	Count	Source papers
Retention	17	[9, 18, 52, 60, 63, 70, 73, 74, 89, 93, 103, 105, 111, 117, 119, 123, 126]
Diversity	13	[22, 25, 28, 41, 60, 70, 73, 85, 93, 103, 111, 118, 127]
Recruitment	5	[18, 28, 86, 97, 127]
Motivation	4	[88, 95, 112, 126]
Better Performance	3	[9, 86, 108]
Without, i.e. not specified	5	[72, 78, 80, 81, 121]

Table 3. Reasons given for using identity.

4.2 RQ2: Why is an identity theory used as a theoretical framing?

To answer this question, we used the same category system that was developed in the K-12 SLR for coding the motivation for using identity. The coding results are presented in Table 3, a brief description of the codes can be found in Sec. 2.4.

Retention here means that the authors indicated identity theory is used because it is associated with a positive effect on retention in the discipline, while *diversity* is aimed at either a more diverse student body or the entry and retention of groups historically disadvantaged in computer science. *Recruitment* (referred to as *Promotion of Young Talent* in [51]) was assigned as a code if the authors associated with a distinct identity the more likely entry into an academic career as a computer scientist, *motivation* if identity was assumed to mean an increased willingness to engage with topics within the discipline, and *better performance* if the authors expressed hope for improved student performance with an increased identity.

Some articles did not give an explicit or implicit reason for why they chose to frame with an identity theory. In this case, we assumed it was *not specified.* Remarkably, those articles that were coded with diversity, mainly focused either gender (that was usually operationalized as sex) or race as main focus. Only three out of the 13 papers who gave diversity as their motivation actually focused intersections of gender and race or social class and race [103] [41] [118].

A general overview of results to RQ2 is also included in the summary of the corpus in the Appendix.

4.3 RQ3: How is the identity concept understood?

We answer the question of understanding identity by first outlining which theoretical framings have made their way into the CSE literature, then present the results of dimensions and conceptualizations of identity (s. Sec. 2.2). Out of the categorized conceptualizations we further developed the inductive categories of CSE specific attributes and practices.

4.3.1 RQ3.1: Which theoretical identity frameworks are used? Regarding RQ3.1 we found a wide range of different theoretical frameworks at first sight. All three strands we introduced in Sec. 2.1 were represented at least once. However, our corpus included theoretical papers as well, that discussed a wide range of different theoretical frameworks. We therefore restricted our analysis of the theoretical frameworks on the 36 empirical papers (papers that include the analysis of empirical data collected from students) in our corpus. More than one theoretical framework was only coded if the paper directly assessed such a double theoretical background (which was the case for five papers here). Within these empirical papers most either focused on socio-cultural frameworks (27 overall) or psychological backgrounds (11 overall), while there were no socio-political backgrounds found. Seven papers did not mention or use a theoretical framework at all.

Most popular within the socio-cultural theories was Lave and Wengers [67] *Communities of Practice* theory (12 times overall), followed by Carlone and Johnsons [16] (and based on that Hazari [53], Godwin [46] and Mahadeo et al. [73])

science and engineering identity theory (6 times overall) and Gee's [42] four kind of identities (4 times overall). Holland et al. [56] as well as Sfard & Prusak [110] were coded at least once. We identified and coded the use of three more socio-cultural theories, that did not refer to any of the theories mentioned in Sec. 2, namely [37, 115, 116]. Within the psychological strand most theories referred to Eccles et al. [30] or related (e.g. Pekrun [91] (4 overall). Some papers also refer to Erikson [35] (2 overall) or Marcia [75] (1 overall). The remaining four papers used different psychological backgrounds, that did not refer to any of the theories mentioned (e.g. Entwistles learning theory [34], Spencer's et al. PVEST [113] and professional identity [24, 47, 58, 106]).

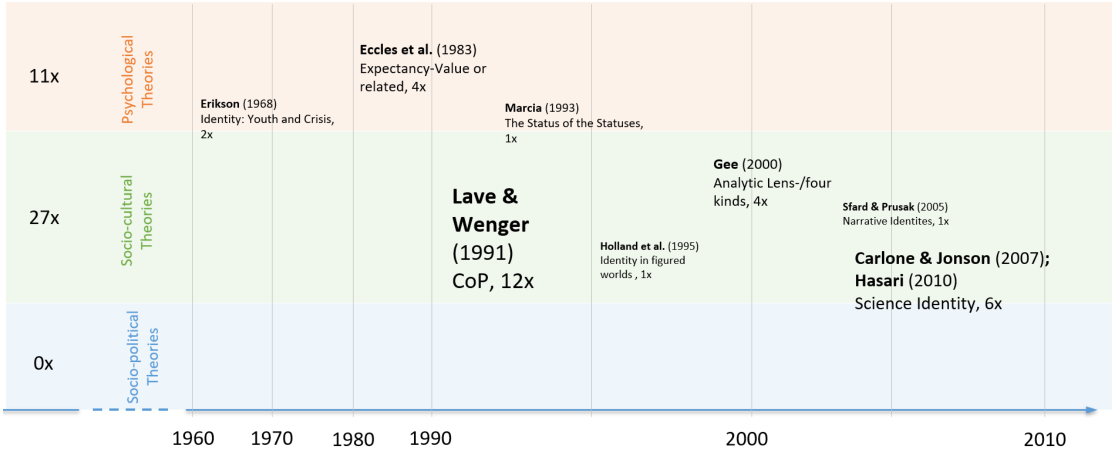


Fig. 4. Overview of the identity theories used.

A cross-comparison of the motivations with the theoretical framework shows that retention is mentioned particularly frequently in connection with the theory of Lave and Wenger [67] (6) and other, predominantly sociocultural theories (8). In contrast, psychological theories appear particularly frequently with the motivation of *recruitment* (5) and *diversity* (5). A general overview of results to RQ3.1 is also included in the summary of the corpus in the Appendix.

4.3.2 RQ3.2: How is identity understood and conceptualized? To answer this question, we coded the different dimension as well as conceptualizations, mentioned by Radovic et al. [99] (s. Sec. 2.2). It turned out, that the two categories *As Individual Attributes* (coded in 19 papers) and *As Specific Practices* (coded in 19 papers) appeared most often, so we restricted our further analysis on these two. The other categories were used rarely, as can be seen in Table 4. The category *other* therefore consists of few mentions of "Identity as Narratives" and "Constrained by local Practices", as well as papers that were unclear about their conceptualization and/or could not be categorized.

While Radovic et al.'s [99] description of identity *As Individual Attributes* fitted our corpus relatively well, we had to adjust the conceptualization of *As a Relationship with a specific Practice*, which we therefore named *As Specific Practices*. While Radovic et al. [99], focus on different types of belonging and the role of students' relationship in relation to different, fluid, and shared practices, we also extend this conceptualization in relation to acquiring or appropriating certain practices. The notion of practices here is thus more normative. Identity formation can be reconstructed not only from the context of different practices, but also be understood as the appropriation and acquisition of certain practices.

Category	Count	Source papers
As Individual Attributes	19	[18, 22, 41, 52, 61, 63, 72, 73, 78, 81, 86, 88, 103, 105, 108, 112, 117, 118, 127]
As Specific Practices	19	[9, 22, 57, 74, 85, 86, 88, 89, 93–97, 103, 105, 118–120, 123]
Other	12	[22, 22, 25, 28, 60, 66, 70, 80, 86, 103, 121, 123]

Table 4. Conceptualizations of identity.

Please note that in total there are more than 41 code assignments, because in some cases several conceptualizations were coded.

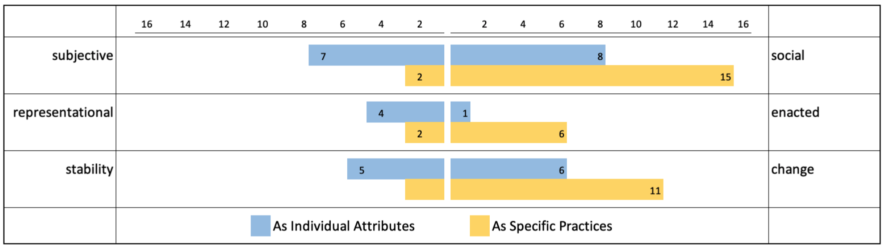


Fig. 5. Distribution of conceptualizations according to Radovic et al. [99] per identity category

Figure 5 shows a comparison of the coding of identity conceptualizations and the dimensions according to Radovic et al. [99]. In studies where the category *As Individual Attributes* was coded, the *social* and *subjective* dimensions were assigned in nearly equal proportions. The same applies to the dimensions *stability* and *change*, while the dimension *representational* was predominantly chosen. This is in line with the conceptualizations and dimensions of Radovic et al. [99]. Studies in the *As Specific Practices* category show a different picture. Here, the dimensions *social*, *enacted*, and *change* were predominantly assigned, which differs from the original dimensions and conceptualizations and hence made us adjusting and renaming the category, as described above.

4.3.3 RQ3.3: What specific attributes and practices are associated with a CS identity? In addition to the general theoretical framing and conceptualization of identity, there is the question of the subject-specific differentiation of identity. For this purpose, we inductively summarized the attributes and practices.

Four categories of attributes emerged as shown in Tab. 5: Recognition, Interest/Motivation, Performance/Competence and Values/Responsibility. The categories are sorted by the number of papers supporting their formation. The Recognition category consists of attributes that describe a persons belonging to the group of computer scientists. The subcode most often found is in fact *recognition* itself as many papers describe perceived recognition as a fundamental characteristic of a profound identity (formation). This is true as well for the Interest/Motivation category, where *interest* is the subcode most often assigned. Besides, the category contains also attributes regarding career goals, as career perspectives are cited as a reason to be (and stay) motivated and interested in the field. The third category Performance/Competence consists of codes such as *academic* or *domain knowledge*, *technical and interpersonal skills* etc. Again, the terms performance and competence themselves were mentioned several times. Under the category Values/Responsibility, evidence was collected for attributes describing qualities such as moral and values, but also the assumption of a *professional role* and *responsibility*.

Category	Count	Source papers
Attributes		
- Recognition	13	[18, 22, 41, 63, 72, 73, 86, 103, 105, 112, 117, 118, 127]
- Interest/Motivation	9	[18, 41, 61, 63, 73, 78, 88, 105, 108]
- Performance/Competence	6	[18, 41, 73, 81, 105, 117]
- Values/Responsibility	3	[52, 72, 86]
Practices		
- Reflection	4	[22, 94, 103, 123]
- Participation	3	[85, 94, 97]
- Project work	3	[88, 89, 119]
- Real-world experiences	3	[96, 105, 120]
- Problem solving	1	[95]

Table 5. Clustering of conceptualizations *As Individual Attributes* and *As Specific Practices*.

The results for practices are less clear, as the understanding of what counts as a practice was almost never specified by authors or grounded in literature (e.g. Wenger [128]). A notable exception is the work of Peters [93], Peters et al. [96] and Peters and Pears [97], who look in detail at participation and practices as an expression of participation in a community. Furthermore, it can be observed that the descriptions and designations of practices by the authors themselves are often not very specific and do not refer to a certain action, i.e. an activity that can be found in the lifeworld. Therefore, the categories formed are at a high level of abstraction. Reflection describes practices such as the negotiation of meaning within the discipline and self-assurance about, for example, gender-specific practices. Participation combines codes that describe practices that characterize technology as participatory. The categories project work and real-world experiences are partly supported by the same papers because project work often takes place in internship or work contexts. On the other hand, there are both projects that do not take place in a real-world context and real-world experiences that are not linked to a project context. Problem solving as a specific computer science practice is discussed primarily in the work of Peters et al. [93–95].

In conclusion, despite the large number of articles that draw on Lave and Wenger's [67] community of practice (CoP) approach, only a small number of authors provide any information at all about what should be understood by practices within the community of computer scientists.

5 DISCUSSION

As the results show, there is indeed an increasing involvement with identity in computing education research. The category systems created for the research questions allow a statement about the thematic variety in which identity is used: In accordance with the hopes expressed in the introduction associated with the concept of identity (increased motivation, interest, retention etc.), identity is being used and developed in a variety of contexts, and a variety of different theories are being employed. The following discussion takes up the main findings and addresses them first according to the research questions, before concluding with general recommendations on how a value-creating engagement with identity within CER that is appropriate to the discipline might be undertaken in the future.

5.1 RQ1: Identity is used as a theoretical lens, but there is little (disciplinary) theoretical development

The results (see Sec. 4.1) for the first research question show that the code most frequently assigned in the area of theory use is *As Theoretical Lens*, which is unsurprisingly, given the natural context of identity frameworks. However, there currently is no *Concept Definition* or *Generation* within CSE. Although existing concepts are further developed and differentiated, as the strong expression of *Concept Refinement* shows, there are few new developments originating from the discipline itself. In this context, Mahadeo et al. [73] are to be mentioned, whose definition of a computing identity builds strongly on the work of Carlone and Johnson [16]. Another example, however, coming from the K-12 field, is Kong and Wang [65], who actually established a disciplinary notion of computing identity in relation to computational thinking. That this is a desideratum in CER to date becomes clear with a look to related disciplines illustrated in Section 2.3. There, a multifaceted examination and theoretical development of identity with respect to the discipline is taking place, providing specific analytical lenses as well as theoretical and empirical development for further insights and interventions planed.

There is also little integration in and localization of identity within other concepts at present, the only exception being Tonso [121]. However, it is precisely in this way that the perceived vagueness of identity could be clarified, such as the distinction from the notion of *Sense of Belonging*, or the extensive discussion of CS conceptions in the last years [49, 50, 54, 92]. Increased efforts in this areas might make clearer what the analytical strength and epistemic scope of each concept is. Further, interventions in this area are lacking. A connection between retention and a distinct disciplinary identity is regularly stated [2, 11, 12, 21, 130]. However, there is a lack of long-term studies that specifically evaluate interventions aimed at strengthening identity. No qualitative assessment of studies that included interventions was conducted as part of this systematic review. That said, it had already been noticed in the K-12 study that papers presenting interventions sometimes lack methodological resources to provide clear evaluation and measure effects.

The lack of high-quality studies presenting meaningful interventions also means that there is a lack of specific guidance for practitioners: for example, what does one say to a university instructor who asks how to strengthen the professional identity of their students? This question can hardly be answered satisfactorily even after reviewing the papers for this study; we are left with superficial hints.

5.2 RQ2: Retention and diversity are the main motivation for using identity theory

Among the reasons given for using identity, it is striking how wide the gap is between the first two – retention (17) and diversity (13) – and the other mentions, recruitment (5), motivation (4) and better performance (3), see Sec. 4.3). Since it is already known from other fields, or at least the subject of research there, that a distinct disciplinary identity can have a positive effect there as well (see section 2.3), this points to a view of identity within CER that is still too limited: certainly, identity presents itself as a theoretical framing for examining retention and diversity, but it is capable of more beyond that.

Moreover, retention and diversity can be seen as "chronic problems" of CER that have shaped numerous discussions for decades now [10, 36, 43]. Especially socio-cultural and socio-political theories of identity, which guide the view away from the individual towards the circumstances shaping the individual, may be profitable here. However, the scope of findings in these areas should not be overestimated: Parts of what has been found from an identity-theoretic perspective had already been researched and known in other circumstances (e. g. from motivational research). That may also be related to the fact, that diversity is not seen within a broader sense, but mainly reduced on focusing either race or gender, instead of intersectionality [103]. Specifically, identity theories seem to offer various potentials to frame this and put intersectional discrimination into the focus of research. As mentioned in the discussion of RQ1 already, the connection of the empirical results to the theoretical perspective is sometimes missing so that problems are named and identified, but no practical implications for changing circumstances can be made. The lack of effective interventions already noted becomes even more pressing from this point of view.

Within the categorization of different motivations for using identity as a framework, the question arises to what extent diversity and retention can be considered fundamentally independent of each other. Diversity studies usually focus on the interests, needs and perspectives of historically marginalized groups, while retention studies focus on why students cannot be retained in computer science. From an identity-theoretical point of view, it could be stated for both areas that a lack of opportunities for identification and participation, of recognition, perception and a sense of belonging are the reason for the problems. The question of why many students leave the subject and why certain student groups do not feel welcome within the subject cannot and should not be considered independently of each other. Addressing these problems would, conversely, mean that all students are served.

5.3 RQ3: Little theoretical diversity, but a new dimension for *Computing Identity*

With respect to the different theoretical frameworks, a clear result emerged in these evaluations: A majority of the categorized papers worked with either psychological constructs or Lave and Wenger's *Community of Practice* perspective, see Sec. 4.3. This focus on just two perspectives thereby undercuts the elaborated broad potentials of identity theories: For example, positioning in the context of diversity [8, 56], has hardly been used.

The connection of the categories from research questions 2 and 3 is remarkable: While we mentioned that socio-cultural perspectives show high potential within diversity perspectives, papers that stated diversity as a motivation were particularly likely to draw on psychological frameworks. On the other hand, papers that stated retention as a motivation used socio-cultural frameworks more often.

Similarly to this, when comparing the results with existing reviews such as the one in mathematics [27, 99], it is noticeable that the discourse around identity in CSE is more normative. Not only in terms of *Individual Attributes* aspects of how computer scientists "should be" are repeatedly brought up, but also in terms of *Practices*: the focus is sometimes on what practices *true* computer scientists perform and how students can acquire them. Thus, there is less room for alternative practices and participation opportunities that would be of great importance, especially from a diversity perspective.

Focusing on specific practices and attributes within the identified frameworks (see Sec. 4.3.3), the lack of subject-specific elaboration of theories, as mentioned in RQ1, is clearly evident in the vagueness that forms the outcome of the analysis of practices: Although CoP is the most popular theory, the papers studied often do not state how practices that make up the community could be seen or described. The practices found are on a very general level, difficult to operationalize, and give only a superficial and in some cases not even domain-specific impression of what constitutes computer scientists as a community.

The situation is different with regard to attributes, for which four clusters with broad meanings could be identified: Performance/Competence, Recognition, Interest/Motivation and Values/Responsibility.

What is interesting regarding this grouping of the attributes is that it largely corresponds to the *Science Identity* concept of Carlone and Johnson [16] and thus also – as a further development – to that of Mahadeo [73]: Figure 6 shows the evolution of this line of tradition, starting from Carlone and Johnson [16], through Hazari [53] and Godwin [46], and finally culminating in Mahadeo's et al. [73] concept of *Computing Identity*. What all conceptualizations have in common is that they include performance, competence, and recognition and later also interest. Therefore, the

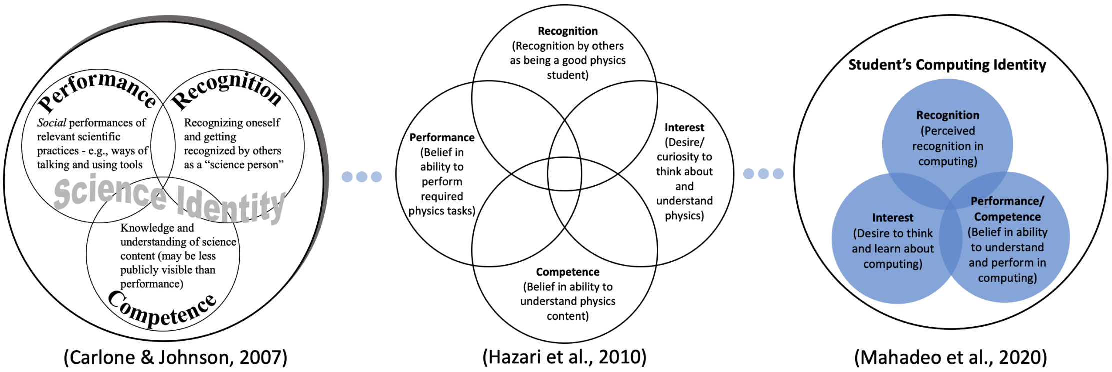


Fig. 6. Development of the identity theory initiated by Carlone and Johnson [16]

sub-concept of Values/Responsibility (see section 4.3.3) is new and not yet present in the conceptualizations mentioned. The identification of the Values/Responsibility subconcept leads to a possible extension of the Computing Identity and results in a further development of the model, which can be seen in Figure 7.

The identification of this new sub-concept of Values/Responsibility deserves further investigation by looking at the papers from our corpus that constitute this new subdimension: Harrington et al. [52] cite the importance of values and taking up professional roles when considering identity from Ibarra, who investigate the identity transformation of professionals when transitioning into senior roles [58]. In contrast, Loui [72] looks in great detail at the impact of an ethics course on the development of computer engineering students' professional identity. His interviews show that students describe honesty, integrity, responsibility, and moral standards as essential qualities of an engineer. Peters and Pears [97] emphasize the importance of a broader focus on technology, especially in relation to society. Their respondents saw it as part of a computer scientist's role to work on projects that allow one to contribute to society.

Values/responsibility thus can be seen to represent a new subdimension in the model similar to Interest/Motivation, since they also represent a character disposition, but one that follows from an assumed or acquired attitude, an *ethos*. This influences one's own behavior in a way that can be partially contrary to one's own interests or motivation: One behaves in this way because one knows that this is what is expected of a member of the community. On the other hand, this does not necessarily have anything to do with recognition: Values/responsibility take effect even when there are no observers, because they stem from an internalized attitude.

Regarding the results of clustering the attributes, however, it is worth mentioning two serious methodological limitations: (1.) Carlone and Johnson and theoretical frameworks based on them were found in the corpus a few times, as shown in the results. Thus, it can be assumed that this has had some influence on the coding and grouping of the attributes. (2.) Moreover, the attributes are not the result of investigations of the articles, but are part of the conceptualization mentioned in the papers. In this respect, the attributes themselves and therefore the clustering lacks a direct empirical foundation. The direct link between a disciplinary identity as a computer scientist and values or accountability thus possesses a strong circumstantial evidence, but falls short of the final, empirical proof.

Nonetheless, the fact that values and responsibility form part of the identity as a computer scientist is hardly surprising on closer examination: social communities are characterized by shared values, this argument is already laid out by Lave and Wenger [67, p. 98] and others [1, p. 10]. Since computer scientists produce artifacts that have a direct

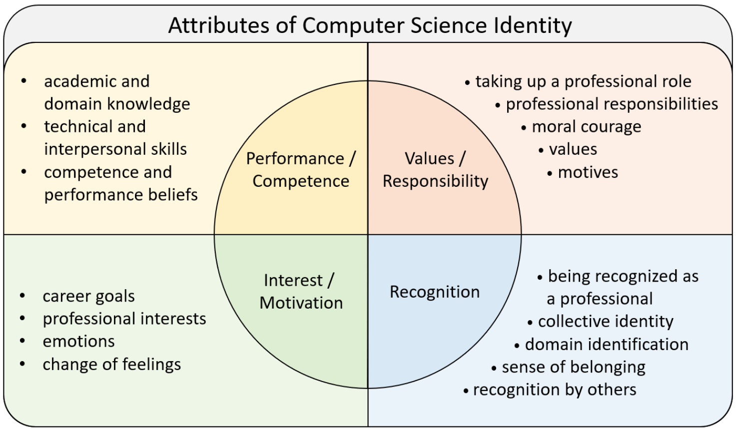


Fig. 7. Attributes of computer science identity.

impact on people's everyday lives, they are often ascribed a special responsibility that can be found, for example, in the ethical guidelines of the professional organizations ACM [48] and IEEE[7]. That responsibility and accountability is thus a component of computer scientist identity should not be surprising [114].

Nevertheless, this finding is of great importance from the point of view of diversity, retention, and attracting new talent: many of the values that prevail within the discipline are not currently the subject of systematic reflection; there is a frequently stated *ethics crisis* [100, 124] that affects not only the fields of machine learning and artificial intelligence. An intensified ethical debate within the discipline, which understands ethics more broadly – namely as a reflection of the values lived in computer science and their interrelationship with society – would not only ensure that computer science lives up to its social responsibility, but would also have the effect of positive self-assurance of its own identity. This could lead to the discipline becoming more attractive to a more diverse student body and to students who otherwise feel deterred by the values of the discipline (as described e.g. by [69]) not dropping out of their studies.

5.4 Recommendations for Future Research

From reading the articles for this and the K-12 review, as well as from comparing them with other disciplines, we can come up with some specific recommendations for actions on how to further develop the topic of identity in CER in order to promote a fruitful discourse within the discipline.

5.4.1 Selecting a theory.

As mentioned (see Section 2.2), strands of theory tend to be conceptualized, and sometimes even operationalized, in different ways. One's research interest and the identity theory used should match and promote each other accordingly. If it is about the individual identity of a person in connection with performance and/or motivational characteristics, then a psychological identity theory is recommended. If it is about identity in the social and historical context of the discipline, a socio-cultural or socio-political theory seems more desirable. Within the strands of theories there are popular representatives, such as Lave and Wenger for the socio-cultural ones. However, the theories of one strand (see figure 1) sometimes have considerable advantages and disadvantages for different research projects, accordingly it is worthwhile to have a closer look at and consequently select a theory.

---

[7] https://www.ieee.org/about/corporate/governance/p7-8.html

5.4.2 Contextualize the theory in your own research interest. The reasons for the selection of the theory and its relation to the research project should be clearly communicated and may form a good part of the background or related work. Likewise, it should be made clear why a chosen conceptualization and operationalization makes sense in the context of the theory and research project. Identity theory should not be used merely as a drop-in to a theory, but should be contextualized clearly and comprehensibly.

5.4.3 Develop the theory. Further development of the theory can already be done through critical reflection: Did identity make sense in the context of the research project? Has it paid off and where are the weaknesses? In this context, it would be especially desirable to take up and discuss long familiar criticisms of well-known theories, such as those of being unscientific and essentialism by Sfard and Prusak [110] of the theory of Lave and Wenger [67]. In addition to critical reflection, another way to advance the theory is to clearly show the connections between identity and other concepts, that is, to do what we called *Concept Location* (see Section 4.1). Articles with empirical research findings would often be able to do this, but rarely take the final step of explicitly making such connections known. In a broader sense, the further or even new development of theory would mean making adjustments and adoptions specific to computer science [84].

These suggestions are in part similar to those made by Danielsson et al. [26] in their review of science identity. However, Danielsson et al. have further valuable suggestions that should be taken into account by researchers.

5.5 Limitations

We have taken steps – such as forward and backward snowballing – to discover all publications that fall within our search grid but still cannot be sure that articles have not been overlooked. Furthermore, the problems concerning the definitional boundaries of identity, self-concept, self-perception and sense of belonging were pointed out (s. 3). In particular, it should be emphasized that the review focused on identity only in a educational context and did not consider all possible types of identity that occur in interaction with computer science.

Especially the database search led to articles that were not from computer science, but also from the science and engineering. Since this was foreseeable, it was determined in advance as a search criterion that such papers are included if they have a reference to computer science. In the case of an empirical paper, for example, this could be that a proportion of the group studied is made up of computer science students. This also satisfies the fact that computer science (especially internationally) has many other labels, such as Informatics, Computer Engineering, Information and Technology Studies, and so on. Although this may result in a loss of definitional precision, this is not supported by the results discussed above.

Regarding the evaluation of the attributes for RQ3, a certain circularity unfortunately cannot be avoided. The discussion of the research question above (see Section 5.3) draws attention to the exact problems.

6 CONCLUSION

Identity is still an emerging topic (within CS) and, for good reasons, is considered a powerful analytical tool. Nonetheless, the theoretical diversity and different understandings of identity can be overwhelming. The goal of this review was to provide an overview of the use and development of the concept of identity in computer science, while also creating a reference article that can serve as an introduction to the historical and conceptual development of the concept of identity. This should help future researchers in the field to orient themselves and lay a good foundation for further – and as shown in the discussion, much needed and fruitful – exploration of CS identity.

As outlined, the disciplinary refinement of the concept would be particularly desirable, just like its sharpening and localization vis-à-vis other concepts that have been intensively researched in recent years. This would help clarify perceived ambiguities and also have the potential to make clear the specific areas of application of the different concepts. Our analysis has shown that there is room for development in the area of concrete practices research especially considering Lave and Wenger's *Communities of Practice* [67] is the most commonly used theory. A closer examination of different types of practices offers the potential to open up the discipline to more types of participation and thereby make it attractive to more people. Similarly, values and responsibility as a subset of identity have not been a focus of research to date, but are of great importance especially in relation to challenges specific to computer science, such as promoting more diversity.

In comparison to identity reviews in related disciplines, it was striking here that many of the conceptualizations were shaped by more or less strong normative ideas about what a computer science identity should be (e.g. building professional identities in terms of values, attributes and motivation), as well as how a professional identity would help to provide student retention or higher motivation (e.g. specific practices or attributes were investigated and discussed of being conducive to achieving such an identity). With regard to diversity issues, it seems useful to detach the view from this normative level and to focus more (even though some papers already addressed this desideratum) on individual identity developments and ways of participation. How do successful individuals from underrepresented or historically marginalized groups develop identity in the discipline? How do they participate? In terms of conceptualization this would mean to shift from identity as pending variable back to identity as a theoretical lens of explaining.

APPENDIX

Below we present short summaries of all the papers contained in our corpus. All summaries follow the same structure. **Method**, **Participants**, **Identity Development or Usage** as well as the **Framework** are derived from the coding process described in Sec. 3. The **Adjacent concepts** include the coding for the motivation (s. Research Question 2) as well as other concepts that were mentioned in the papers in relation to identity.

Boyer et al. [9]: Increasing Technical Excellence, Leadership and Commitment of Computing Students through Identity-Based Mentoring (2010)
Intention: Evaluation of an intervention on Computing Identity Mentoring, that is, the impact of mentoring on performance and retention.
Method / Participants: Surveys / Beginners, Advanced
Identity Development or Usage: As Explanation for Other Concepts
Framework and Adjacent Concepts: No Framework / Retention, Performance, Self-efficacy, Leadership, Career planing
Results: As a result of their empirical study, the authors report that Computing Identity Mentoring has a positive impact on students' development as computer scientists.

Choe and Borrego [18]: Prediction of engineering identity in engineering graduate students (2019)
Intention: The authors observed that identity research exists for undergraduates (e.g. instruments) but not for master's and PhD students. Their study investigates whether the results for undergraduates are applicable to their target group and where the differences lie.
Method / Participants: Surveys, Interviews / Professionals
Identity Development or Usage: Transfer to other Target Group
Framework and Adjacent Concepts: Carlone & Johnson, Gee / Recruitment, Retention, Interpersonal skills, Curriculum Design
Results: The article reports on the development of a new instrument for an engineering identity scale consisting of four factors: Interest, Recognition, Competence, and Interpersonal Skills Competence. While the first three factors are already present in Carlone & Johnson, the Interpersonal Skill Competence factor is an extension to the more professional target group, reflecting the broader responsibilities of master's and PhD students.

Cohoon et al. [22]: Conflicted Identities and Sexism in Computing graduate programs (2010)
Intention: The paper discusses women's experiences with sexism and related coping strategies, such as minimizing their feminine identity.
Method / Participants: Focus groups / Advanced
Identity Development or Usage: As Theoretical Lens
Framework and Adjacent Concepts: Collective Identity / Diversity
Results: The study further examines the coping strategies of either denying sexism or one's own female identity as responses to dealing with sexism in CS.

Cummings et al. [25]: Computing Resilient Identity Development and Maintenance of Black Americans Who Earned a PhD in Computing (2019)
Intention: The authors examine factors that contribute to the success and identity development of African American PhDs.
Method / Participants: Interviews / Professionals
Identity Development or Usage: As Theoretical Lens
Framework and Adjacent Concepts: PVEST (Spencer et al.) / Diversity, Sense of Belonging, Resilience
Results: The study reports exploratory results that show how to develop and maintain a resilient computing identity.

Davis et al. [28]: Multiple Case Study of Nerd Identity in a CS1 Class (2014)
Intention: The paper discusses observations on student identities in a CS 1 course, specifically on the "nerd identity" and the consequences for multiple/diverse identities.
Method / Participants: Observation, Interviews / Beginners
Identity Development or Usage: As Theoretical Lens
Framework and Adjacent Concepts: Carlone & Johnson / Recruitment, Diversity
Results: The analysis shows how nerd identities enhance participation and engagement in CS, which is in contrast to much other research that portrays nerd identity as stereotypical or problematic.

Garcia et al. [41]: Examining the Computing Identity of High-Achieving Underserved Computing Students on the Basis of Gender, Field, and Year in School (2018)
Intention: The authors examine the success factors and computing identity of high-achieving underserved computing students.
Method / Participants: Surveys / Beginners, Advanced
Identity Development or Usage: As Theoretical Lens
Framework and Adjacent Concepts: Computing Identity Framework (Mahadeo et al.) / Diversity, Gender, Underserved Students
Results: The study identifies differences in computing identity between genders, first and post-second year students, and various study programs, as well as differences between sub-constructs.

Harrington et al. [52]: A Qualitative Analysis of Computing Students' Professional Identity and its Relationship to Strategies for Coping with Stressors in the Computing Disciplines (2007)
Intention: The paper examines students' professional identity development and related stressors, as well as coping strategies students describe to promote overall identity development.
Method / Participants: Focus groups / Beginners, Advanced
Identity Development or Usage: Concept Refinement
Framework and Adjacent Concepts: Professional identity / Retention
Results: As a result of this study, the overall Professional Development framework will be revised to include, for example, "nerd-being" as part of Professional Identity development. Furthermore, stressors of Professional Development are being discussed.

Hughes et al. [57]: Development of Leadership Self-Efficacy: Comparing Engineers, Other STEM, and Non-STEM Majors (2018)
Intention: The paper uses Lave and Wenger's community of practice theory to shed light on the co-development of engineering identity and leadership through an empirical study.
Method / Participants: Surveys / Advanced
Identity Development or Usage: Concept Refinement
Framework and Adjacent Concepts: Lave & Wenger / Technical mastery, Professionalism
Results: The preliminary results reported by the authors suggest a positive relationship between engineering identity and leadership. However, this link is very tenuous: both leadership and identity are fostered by internships, group projects, etc., so a dependence on a third factor seems likely.

Kapoor and Gardner-McCune [61]: Understanding CS Undergraduate Students' Professional Identity through the lens of their Professional Development (2019)
Intention: In this interview study, intrinsic and discipline-specific factors are listed and students' extracurricular community engagement is recorded to examine how students develop their professional identity.
Method / Participants: Interviews / Beginners, Advanced
Identity Development or Usage: As Basis for Interventions
Framework and Adjacent Concepts: Marcia (Identity Status Theory) / Professional identity, Curriculum design
Results: It was found that computer science students develop their professional identity between the second and third year of study. The authors also emphasize the need to build students' self-confidence through course interventions or projects in the first two years of computer science study.

Kapoor and Gardner-McCune [60]: Understanding Professional Identities and Goals of Computer Science Undergraduate Students (2018)
Intention: The study examines the relationship between career goals and student identity in order to draw conclusions about the design of curricula and retention programs.
Method / Participants: Surveys, Interviews / Beginners
Identity Development or Usage: Concept Refinement
Framework and Adjacent Concepts: Lave & Wenger / Retention, Diversity
Results: CS undergraduate students have various professional goals; their career goals depend on their professional identities. The authors suggest that current curricula be adapted to help students accomplish their career goals by offering specializations.

Kinnunen et al. [63]: Understanding initial undergraduate expectations and identity in computing studies (2018)
Intention: This study addresses the need for a better understanding of students' expectations, particularly their views of the field of computer science, through insights into their identities as computer science students and future professionals.
Method / Participants: Surveys, Essays / Beginners
Identity Development or Usage: As Explanation for Other Concepts
Framework and Adjacent Concepts: Lave & Wenger / Recruitment, Retention

Results: Students expect not only to be trained in aspects such as programming or systems development, but also to gain insights that go beyond the technical, such as the role that computer science can play in society. It is also apparent that students are not looking for specific job positions, but rather relate their future aspirations to specific types of work environments or job characteristics.

Kramer et al. [66]: A Narrative-Style Exploration of Undergraduate Engineering Students' Beliefs about Smartness and Identity (2019)
Intention: The paper presents the results of a narrative study of engineering students' educational trajectories in relation to their conceptions of intelligence and identity. The authors aim to better understand the nuanced interaction between these constructs.
Method / Participants: Interviews, Surveys / Beginners
Identity Development or Usage: As Explanation for Other Concepts
Framework and Adjacent Concepts: Gee / Beliefs about Ability and Smartness, Curriculum design
Results: Social experiences were found to have a significant impact on one's identity and confidence in one's intelligence. Of particular note are the female participants whose social isolation in engineering courses had a negative impact on their identity and beliefs.

Lewis et al. [70]: "I Don't Code All Day": Fitting in Computer Science When the Stereotypes Don't Fit (2016)
Intention: The paper argues that, instead of changing stereotypes, it may be possible to challenge students' beliefs that stereotypes of CS are relevant to whether they can become a computer scientist.
Method / Participants: Interviews / Beginners
Identity Development or Usage: As Theoretical Lens
Framework and Adjacent Concepts: Sfard & Prusak / Retention, Diversity
Results: From the student interviews (and therefore from their perspective), the article identifies four characteristics to fit into CS: singularly focused on CS, asocial, competitive, male. It is discussed that some students have been able to reject these stereotypes and still feel they "fit in", especially when they know different role models.

Loui [72]: Ethics and the Development of Professional Identities of Engineering Students (2005)
Intention: The study presents a project in which students in an engineering ethics class discuss questions about the characteristics and responsibilities of professional engineers.
Method / Participants: Essays / Beginners, Advanced
Identity Development or Usage: Concept Expansion
Framework and Adjacent Concepts: No Framework / Professional identity, Four-stage model of role acquisition
Results: The results show that students learn about professionalism mainly from relatives and colleagues who are engineers, and rarely from technical engineering courses. Furthermore, by analyzing cases in groups and listening to different perspectives, some students understand professional responsibility at the end of the course not only as liability for mistakes, but in a broader sense as responsibility to society.

Mahadeo et al. [73]: Developing a Computing Identity Framework: Understanding Computer Science and Information Technology Career Choice (2020)

Intention: The authors adapt theory from science and engineering identity theory to present a construct for computing identity and empirically validate a related instrument.
Method / Participants: Surveys / Beginners
Identity Development or Usage: Construct Refinement
Framework and Adjacent Concepts: Hazari (Science/Engineering Identity) / Diversity, Retention
Results: The article reports on the instruments' development and validation, as well as on evidence that there are three subconstructs of Computing Identity.

Maher et al. [74]: The Connected Learner (2016)
Intention: Evaluation of a comprehensive intervention that focuses on increasing student connection and innovative teaching methods to help students form an identity.
Method / Participants: Surveys / Beginners, Advanced, Professional
Identity Development or Usage: Unclear
Framework and Adjacent Concepts: Lave & Wenger / Retention
Results: The author share preliminary results indicating that their intervention works.

McCartney and Sanders [78]: School/Work: Development of Computing Students' Professional Identity at University (2015)
Intention: This case study attempts to identify events that have had a significant impact on students' development during university computer science education.
Method / Participants: Interviews / Beginners
Identity Development or Usage: As Theoretical Lens
Framework and Adjacent Concepts: No Framework / Professional identity, Three dimensions of becoming an engineer
Results: Career-related factors appear to be more important than expected. Moreover, the authors report a complex interrelationship between these factors and the courses chosen. Courses influence students' perceptions of the world of work, and actual work experiences have an impact on students' expectations of their education.

Meharg et al. [80]: "So far back, I'm anonymous": Exploring Student Identity using Photovoice (2018)
Intention: The authors aim to investigate student identity in the transitional context of transfer students.
Method / Participants: Photovoice / Advanced
Identity Development or Usage: Unclear (Usage)
Framework and Adjacent Concepts: Gee, Identity formation through Crisis / Transition, College transfer
Results: College transfer students have different educational experiences and identity development. The authors suggest that institutions should pay particular attention to their needs.

Mishra [81]: Professional identity construction among software engineering students: A study in India (2016)
Intention: This study examines the process by which final year software engineering students construct their professional identities and explains the process of "identity morphing" as a mechanism by which students resolve the conflict/violation of their identities.

Method / Participants: Interviews / Advanced
Identity Development or Usage: Concept Expansion
Framework and Adjacent Concepts: No Framework / Professional identity
Results: The article suggests that derived self-esteem and perceived competence influence an individual's identity transition. The authors report on evidence in their data that individuals with higher levels of perceived competence had less conflict in their identity building process.

Nelson et al. [85]: A Qualitative Investigation on the Effectiveness of a Computing Identity Development Emailing List for African American Computer Scientists (2019)
Intention: This paper presents a qualitative interview study focused on African American doctoral students in computer science.
Method / Participants: Surveys, Interviews / Professionals
Identity Development or Usage: As Basis for Interventions
Framework and Adjacent Concepts: No Framework / Resilience, Diversity
Results: The authors' findings suggest that African American computer science identity can be effectively maintained within this intervention by providing a sense of belonging to the community, enabling optimistic outlooks, and promoting self-efficacy.

Nylén et al. [86]: Why are We Here? The Educational Value Model (EVM) as a Framework to Investigate the Role of Students' Professional Identity Development (2018)
Intention: The authors note that "the goal of professional identity development" is usually not reached during CS higher education and provide an analysis that points to factors that explain this situation.
Method / Participants: Literature Review / No Participants
Identity Development or Usage: Concept Expansion
Framework and Adjacent Concepts: Not coded / Belonging, Motivation
Results: Based on the literature review and identified challenges, the paper provides a theoretical framework to describe professional identity development in university education.

Parker [89]: Who I Am Becoming, Now: Toward a Computer Science Professional Identity Instrument (2018)
Intention: Description of the development and preliminary results of a professional identity survey instrument.
Method / Participants: Surveys, Interviews / Beginners, Advanced
Identity Development or Usage: Concept Refinement
Framework and Adjacent Concepts: Lave & Wenger / Career planning, Retention
Results: Future career goals and job titles, as well as how peers and supervisors perceive one's role, shape professional identity.

Parker [88]: How Do You Feel: Affective Expressions from Computer Science Senior Capstone Projects (2017)
Intention: In the context of final projects, the authors examine affective response as a factor leading to engagement and the formation of a professional identity during the transition from university to professional life.

Method / Participants: Interviews / Advanced
Identity Development or Usage: As Basis for Interventions
Framework and Adjacent Concepts: Lave & Wenger / Interest, Engagement
Results: Respondents experienced a range of affective responses during their project experience and their engagement appeared to be related to the impact of project outcomes. Thus, the capstone experience in such a real-world project is highlighted as a turning point in the trajectory of these students and represents "a bridge between the academic and professional communities of practice."

Peters [93]: Students' Experience of Participation in a Discipline—A Longitudinal Study of Computer Science and IT Engineering Students (2018)
Intention: The study follows students over time to capture their experiences of participation (as defined by Lave and Wenger) and how they change over the course of their studies.
Method / Participants: Self-reports, Interviews / Beginners, Advanced
Identity Development or Usage: Concept Refinement
Framework and Adjacent Concepts: Lave & Wenger / Retention, Diversity, Participation
Results: Besides an insightful theory section the article presents a comprehensive long-term study with several phenomenographic outcome spaces as results and an interesting discussion of the findings.

Peters and Pears [97]: Engagement in Computer Science and IT - What! A Matter of Identity? (2013)
Intention: The paper develops and illustrates the use of a new theoretical framework (Enwistle and Lave/Wenger) to systematically study the development of student identity in Computer Science and IT.
Method / Participants: Self-Report (Reflections) / Beginners
Identity Development or Usage: As Theoretical Lens
Framework and Adjacent Concepts: Lave & Wenger, Enwistle / Recruitment, Belonging
Results: The paper investigated that students collectively express conceptions of knowledge, as being absolute and provided by the university, which may imply a less sophisticated conception of learning. Furthermore, two groups of students with different negotiations of meaning are identified and discussed.

Peters et al. [96]: Preparing the Global Software Engineer (2015)
Intention: This study assesses student evaluations and reflections on a course in which students worked on a real-world problem in collaboration with a local client. The focus is on students' perceptions of software engineering, the perceived relevance of a global learning experience, and their role in reshaping their identities as global software engineers.
Method / Participants: Self-reports / Beginners, Advanced
Identity Development or Usage: As Basis for Interventions
Framework and Adjacent Concepts: Flanagan (Critical Incident Analysis) / Curriculum Design
Results: The analysis shows that educating the global software engineer is a complex endeavor. However, means of designing and assessing courses aimed at supporting students on their path to becoming global software engineers are presented.

Peters et al. [95]: Second year computer science and IT students' experience of participation in the discipline (2015)

Intention: The authors describe the participation experiences of second year computing students, and discuss participation as part of a broader goal of understanding the identity development of computing students.
Method / Participants: Interviews, Surveys / Beginners
Identity Development or Usage: As Theoretical Lens
Framework and Adjacent Concepts: Lave & Wenger / Motivation, Belonging
Results: The paper provides nuanced reflections on practices and participation in CS, and specifically the practice of problem solving as part of students' identity development.

Peters et al. [94]: First Year Computer Science and IT Students' of Participation in the Discipline (2014)
Intention: The goal of this paper is to analyze a specific aspect of the student experience, participation, in order to gain a better understanding of how computer science and IT students engage with CS prior to and during their studies.
Method / Participants: Surveys, Interviews / Beginners
Identity Development or Usage: As Theoretical Lens
Framework and Adjacent Concepts: Lave & Wenger / Motivation, Belonging
Results: The paper presents different experiences of student participation as a table.

Rodriguez and Lehman [103]: Developing the next generation of diverse computer scientists: the need for enhanced, intersectional computing identity theory (2017)
Intention: This theoretical paper explores the need for an enhanced, intersectional theory of computing identity in order to develop a diverse group of computer scientists for the future.
Method / Participants: Literature Review / No Participants
Identity Development or Usage: As Theoretical Lens
Framework and Adjacent Concepts: Not coded / Diversity, Retention
Results: The paper identifies a lack of intersectionality in current research, as well as of literature specific to computing identity, and recommends responding with further research from an intersectional perspective.

Rohde et al. [105]: Design Experiences, Engineering Identity, and Belongingness in Early Career Electrical and Computer Engineering Students (2019)
Intention: The authors examine the role of design experiences for identification and belongingness in engineering and the meaning of identification and belongingness for students.
Method / Participants: Surveys, Interviews / Beginners
Identity Development or Usage: Concept Expansion
Framework and Adjacent Concepts: Lave & Wenger / Retention, Belonging, Design Experience
Results: The authors conclude that design experiences lead to a stronger sense of belonging and move students from the periphery to the center of the CS community of practice.

Scott and Ghinea [108]: Measuring enrichment: the assembly and validation of an instrument to assess student self-beliefs in CS1 (2014)
Intention: The paper begins with the observation that self-beliefs may have an impact on academic success. Since there is no valid instrument to measure these beliefs , the authors attempt to fill this gap.
Method / Participants: Surveys / Beginners

Identity Development or Usage: As Basis for Interventions
Framework and Adjacent Concepts: Control-Value Theory of Achievement Emotion / Interest, Mindset, Improved learning
Results: The paper develops an instrument, but it is not documented in the appendix or by a link. The specific utility (e.g., reasons for use, etc.) of the instrument remains vague.

Sinclair and Kalvala [111]: Exploring societal factors affecting the experience and engagement of first year female computer science undergraduates (2015)
Intention: Based on a study with beginning students conducted in the UK, socio-cultural factors influencing study and CS identity are identified. Special attention is given to gender differences.
Method / Participants: Surveys / Beginners
Identity Development or Usage: As Theoretical Lens
Framework and Adjacent Concepts: Carlone & Johnson, Gee / Retention, Diversity
Results: The paper identifies a variety of issues related to the situation of women in computer science studies and identifies specific starting points for future interventions.

Smith-Orr and Garnett [112]: Motivation and identity in C++ the effects of music in an engineering classroom (2016)
Intention: The paper used a model of academic motivation (MUSIC) to guide course design within an introductory programming class in an attempt to increase engineering identity and sense of belonging among engineering students.
Method / Participants: Surveys / Beginners
Identity Development or Usage: Concept Refinement
Framework and Adjacent Concepts: MUSIC framework / Belonging, Motivation
Results: The authors note that the students' motivation (according to MUSIC) increased, but their sense of belonging and identity decreased; their intervention failed.

Taheri et al. [117]: A Structural Equation Model Analysis of Computing Identity Sub-Constructs and Student Academic Persistence (2018)
Intention: This research report examines the impact of CS identity on the persistence of computer science students by analyzing the effects of achievement/competence, recognition, interest, and sense of belonging on the academic persistence of computer science students.
Method / Participants: Surveys / Beginners, Advanced
Identity Development or Usage: Construct Definition, Refinement or Expansion
Framework and Adjacent Concepts: Carlone & Johnson / Retention, Persistence
Results: The results indicate that the authors' CS identity model is consistent with previous research on disciplinary identity and that students' academic persistence is directly influenced by their interest. The authors consider their model to be a useful analytical lens for further curricular and extracurricular activities.

Tate and Linn [118]: How Does Identity Shape the Experiences of Women of Color Engineering Students? (2005)
Intention: This study seeks to understand the experiences of women of color engineering students who persist and to

identify some of the dilemmas they face.
Method / Participants: Interviews / Advanced
Identity Development or Usage: As Theoretical Lens
Framework and Adjacent Concepts: Lave & Wenger / Diversity
Results: This study of women of color engineering students reveals how interactions between academic, intellectual and social identities jointly influence perceptions of educational experiences and career aspirations.

Taylor-Smith et al. [119]: Identity and Belonging for Graduate Apprenticeships in Computing (2019)
Intention: This study explores how Graduate Apprenticeship (GA) students experience their association with the university and their identities as students, but also as employees.
Method / Participants: Interviews / Advanced
Identity Development or Usage: Unclear
Framework and Adjacent Concepts: Stryker & Burke / Retention, Belonging
Results: Here the authors consider a cohort of CS students who are pursuing vocational training on the side. They hypothesize that these students acquire a strong IT professional identity in the workplace that also supports their development at the university. The analysis revealed that the students define themselves differently from traditional student identities and that there is a strong sense of belonging to their fellow students who are also in vocational education.

Tokmic et al. [120]: Salient Measures of an Engaged Computing Education Community (2019)
Intention: This paper presents the measurement and analysis of a comprehensive model of pedagogical change in the computing college and discusses how studying the construct of professional identity contributes to the knowledge base of the computing education community.
Method / Participants: Surveys / Beginners
Identity Development or Usage: Concept Expansion
Framework and Adjacent Concepts: No Framework / Professional identity, Engagement, Curriculum design
Results: The authors see the contribution of their work as defining peer learning and professional identity in computer science and providing computer science education with measures of psycho-social factors known to be important in the development of computer science students.

Tonso [121]: Student Engineers and Engineer Identity - Campus Engineer Identities as Figured World (2006)
Intention: This study attempts to explain the complex process of identity formation on a campus. The author examines how student engineers are recognized as engineers and how they already perceive themselves as engineers through their actions.
Method / Participants: Observation, Interviews / Beginners, Advanced
Identity Development or Usage: Concept Location
Framework and Adjacent Concepts: Holland et al. / Belonging, Motivation
Results: The preferred campus lifestyle has a tremendous impact not only on the individual as a student, but also on his or her career as an engineer. The author describes identity formation on campus as a process in which the individual's self-concept as an engineer leads to representations of the engineering self, and recognition as an engineer conveys a

sense of belonging.

Ulriksen et al. [123]: What do we know about explanations for drop out/opt out among young people from STM higher education programmes? (2010)
Intention: The paper provides a general overview on understandings of drop out/opt out from science, technology and mathematics (STM) higher education programs, by reviewing existing literature.
Method / Participants: Literature Review / No Participants
Identity Development or Usage: As Theoretical Lens
Framework and Adjacent Concepts: Not coded / Retention
Results: The authors conclude, that dropping out or opting out is influenced by a number of factors and interactions. However, identity construction is a relevant and underestimated factor.

Vesisenaho et al. [126]: Need for Study and Career Counselling in Computer Science (2009)
Intention: The article examines the relationship between career counseling methods and retention/motivation.
Method / Participants: Surveys / Beginners, Advanced
Identity Development or Usage: Unclear
Framework and Adjacent Concepts: No Framework / Career planing, Retention, Motivation
Results: There is no evidence on the use or development of the identity concept, but the authors recommend that CS training and curricula be viewed more from a career/professional perspective.

Washington et al. [127]: The Computer Science Attitude and Identity Survey (CSAIS): A Novel Tool for Measuring the Impact of Ethnic Identity in Underrepresented Computer Science Students (2016)
Intention: The purpose of this study is to measure the impact of ethnic identity on perceptions of CS as a subject and the decision to pursue it. Therefore, the Computer Science Cultural Attitude and Identity Survey (CSAIS) was developed to measure five important constructs that influence the attitudes and identity of CS undergraduate students of color.
Method / Participants: Surveys / Beginners
Identity Development or Usage: As Theoretical Lens
Framework and Adjacent Concepts: Collective identity / Diversity
Results: In addition to a comprehensive discussion of the identity of students of color, the authors present the instrument they developed, the “Computer Science Attitude and Identity Survey”.

REFERENCES

[1] Kwame Anthony Appiah. 2018. *The lies that bind: Rethinking identity.* Profile Books.

[2] Jane M. Armstrong and Richard A. Price. 1982. Correlates and Predictors of Women's Mathematics Participation. *Journal for Research in Mathematics Education* 13, 2 (1982), 99–109. https://doi.org/10.2307/748357

[3] Jeffrey Jensen Arnett. 2014. *Emerging adulthood: The winding road from the late teens through the twenties.* Oxford University Press.

[4] Lucy Avraamidou. 2014. Studying science teacher identity: current insights and future research directions. *Studies in Science Education* 50, 2 (2014), 145–179. https://doi.org/10.1080/03057267.2014.937171

[5] Joseph A. Bailey. 2003. Self-image, self-concept, and self-identity revisited. *Journal of the National Medical Association* 95 5 (2003), 383–6.

[6] Roy F. Baumeister. 1997. Chapter 26 - Identity, Self-Concept, and Self-Esteem: The Self Lost and Found. Academic Press, San Diego, 681–710. https://doi.org/10.1016/B978-012134645-4/50027-5

[7] Maureen Biggers, Anne Brauer, and Tuba Yilmaz. 2008. Student Perceptions of Computer Science: A Retention Study Comparing Graduating Seniors with Cs Leavers. In *Proceedings of the 39th SIGCSE Technical Symposium on Computer Science Education* (Portland, OR, USA) *(SIGCSE '08)*. Association for Computing Machinery, New York, NY, USA, 402–406. https://doi.org/10.1145/1352135.1352274

[8] Jo Boaler. 2000. *Multiple perspectives on mathematics teaching and learning.* Vol. 1. Greenwood Publishing Group.

[9] Kristy Elizabeth Boyer, E Nathan Thomas, Audrey S Rorrer, Deonte Cooper, and Mladen A Vouk. 2010. Increasing technical excellence, leadership and commitment of computing students through identity-based mentoring. In *Proceedings of the 41st ACM technical symposium on Computer science education.* 167–171.

[10] Eileen D Bunderson and Mary Elizabeth Christensen. 1995. An analysis of retention problems for female students in university computer science programs. *Journal of Research on Computing in Education* 28, 1 (1995), 1–18.

[11] Jennifer Burg, V. Paúl Pauca, William Turkett, Errin Fulp, Samuel S. Cho, Peter Santago, Daniel Cañas, and H. Donald Gage. 2015. Engaging Non-Traditional Students in Computer Science through Socially-Inspired Learning and Sustained Mentoring. In *Proceedings of the 46th ACM Technical Symposium on Computer Science Education (SIGCSE '15).* Association for Computing Machinery, New York, NY, USA, 639–644. https://doi.org/10.1145/2676723.2677266

[12] Brenda M Capobianco. 2006. Undergraduate women engineering their professional identities. *Journal of Women and minorities in Science and Engineering* 12, 2-3 (2006), 95–117.

[13] Brenda M. Capobianco, Heidi A. Diefes-Dux, and Meara M. Habashi. 2009. Generating measures of Engineering Identity Development among young learners. In *2009 39th IEEE Frontiers in Education Conference.* 1–6. https://doi.org/10.1109/FIE.2009.5350817 ISSN: 2377-634X.

[14] Brenda M. Capobianco, Brian F. French, and Heidi A. Diefes-Du. 2012. Engineering Identity Development Among Pre-Adolescent Learners. *Journal of Engineering Education* 101, 4 (2012), 698–716. https://doi.org/10.1002/j.2168-9830.2012.tb01125.x

[15] Heidi B. Carlone. 2012. Methodological Considerations for Studying Identities in School Science: An Anthropological Approach. In *Identity construction and science education research*, Maria Varelas (Ed.). SensePublishers, Rotterdam and Boston, 9–26.

[16] Heidi B. Carlone and Angela Johnson. 2007. Understanding the science experiences of successful women of color: Science identity as an analytic lens. *Journal of Research in Science Teaching* 44, 8 (2007), 1187–1218. https://doi.org/10.1002/tea.20237

[17] Heidi B. Carlone, Catherine M. Scott, and Cassi Lowder. 2014. Becoming (less) scientific: A longitudinal study of students' identity work from elementary to middle school science. *Journal of Research in Science Teaching* 51, 7 (2014), 836–869. https://doi.org/10.1002/tea.21150

[18] Nathan Hyungsok Choe and Maura Borrego. 2019. Prediction of engineering identity in engineering graduate students. *IEEE Transactions on Education* 62, 3 (2019), 181–187.

[19] Anna Chronaki and David Kollosche. 2019. Refusing mathematics: a discourse theory approach on the politics of identity work. *ZDM* 51, 3 (2019), 457–468. https://doi.org/10.1007/s11858-019-01028-w

[20] Abigail Clark and Rachel Louis Kajfez. 2019. Engineering Identity in Pre-College Students: A Literature Review. In *ASEE 2019 Annual Conference.*

[21] J McGrath Cohoon and William Aspray. 2006. *Women and information technology: Research on underrepresentation.* Vol. 1. The MIT Press.

[22] J McGrath Cohoon, Monica Nable, and Paul Boucher. 2011. Conflicted identities and sexism in computing graduate programs. In *2011 Frontiers in Education Conference (FIE).* IEEE, S1H–1.

[23] Lee J Cronbach and Paul E Meehl. 1955. Construct validity in psychological tests. *Psychological bulletin* 52, 4 (1955), 281.

[24] Richard L. Cruess, Sylvia R. Cruess, and Yvonne Steinert. 2016. Amending Miller's Pyramid to Include Professional Identity Formation. *Academic Medicine* 91, 2 (2016). https://journals.lww.com/academicmedicine/Fulltext/2016/02000/Amending_Miller_s_Pyramid_to_Include_Professional.17.aspx

[25] Robert Cummings, Byron M Lowens, Whitney Nelson, and Kinnis Gosha. 2019. Computing Resilient Identity Development and Maintenance of Black Americans Who Earned A Ph. D. in Computing. In *2019 IEEE Frontiers in Education Conference (FIE).* IEEE, 1–8.

[26] Anna T. Danielsson, Heather King, Spela Godec, and Anne-Sofie Nyström. 2023. The identity turn in science education research: a critical review of methodologies in a consolidating field. *Cultural Studies of Science Education* (jan 2023). https://doi.org/10.1007/s11422-022-10130-7

[27] Lisa Darragh. 2016. Identity research in mathematics education. *Educational Studies in Mathematics* 93, 1 (Sept. 2016), 19–33. https://doi.org/10.1007/s10649-016-9696-5

[28] Don Davis, Timothy Yuen, and Matthew Berland. 2014. Multiple case study of nerd identity in a CS1 class. In *Proceedings of the 45th ACM technical symposium on Computer science education.* 325–330.

[29] Jane E Dutton, Janet M Dukerich, and Celia V Harquail. 1994. Organizational images and member identification. *Administrative science quarterly* (1994), 239–263.

[30] J. Eccles. 1983. Expectancies, values and academic behaviors. *Achievement and achievement motives* (1983).

[31] Jacquelynne S. Eccles. 2009. Who Am I and What Am I Going to Do With My Life? Personal and Collective Identities as Motivators of Action. *Educational Psychologist* 44, 2 (2009), 78–89. https://doi.org/10.1080/00461520902832368

[32] Jacquelynne S. Eccles and Allan Wigfield. 2020. From expectancy-value theory to situated expectancy-value theory: A developmental, social cognitive, and sociocultural perspective on motivation. *Contemporary educational psychology* 61 (2020), 101859. https://doi.org/10.1016/j.cedpsych.2020.101859

[33] Yrjö Engeström. 2014. *Learning by Expanding: An Activity-Theoretical Approach to Developmental Research* (2 ed.). Cambridge University Press, Cambridge. https://doi.org/10.1017/CBO9781139814744

[34] Noel Entwistle. 2007. Conceptions of learning and the experience of understanding: Thresholds, contextual influences, and knowledge objects. (2007).

[35] E.H. Erikson. 1968. *Identity: youth and crisis.* Norton & Co., Oxford, England.

[36] Allan Fisher and Jane Margolis. 2002. Unlocking the clubhouse: the Carnegie Mellon experience. *ACM SIGCSE Bulletin* 34, 2 (June 2002), 79–83. https://doi.org/10.1145/543812.543836

[37] John C Flanagan. 1954. The critical incident technique. *Psychological bulletin* 51, 4 (1954), 327.

[38] Hanoch Flum and Avi Kaplan. 2012. Identity formation in educational settings: A contextualized view of theory and research in practice. *Contemporary Educational Psychology* 37, 3 (2012), 240 – 245. https://doi.org/10.1016/j.cedpsych.2012.01.003 Identity Formation in Educational Settings.

[39] Trine Foyn, Yvette Solomon, and Hans Jørgen Braathe. 2018. Clever girls' stories: the girl they call a nerd. *Educational Studies in Mathematics* 98, 1 (2018), 77–93. https://doi.org/10.1007/s10649-017-9801-4

[40] Francis Fukuyama. 2018. *Identity: The demand for dignity and the politics of resentment.* Farrar, Straus and Giroux.

[41] Atalie Garcia, M Ross, Z Hazari, M Weiss, Kenneth Christensen, and M Georgiopoulos. 2018. Examining the Computing Identity of High-Achieving Underserved Computing Students on the Basis of Gender, Field, and Year in School. *Collaborative Network for Engineering and Computing Diversity (CoNECD)* (2018).

[42] James Paul Gee. 2000. Chapter 3: Identity as an analytic lens for research in education. *Review of research in education* 25, 1 (2000), 99–125.

[43] Michail Giannakos, Ilias Pappas, Maria Letizia Jaccheri, and Demetrios Sampson. 2017. Understanding student retention in computer science education: The role of environment, gains, barriers and usefulness. *Education and Information Technologies* 22, 5 (9 2017), 2365–2382.

[44] Lara Gildehaus. 2021. Identität als Perspektive zur Genese individueller Wertehierarchien im Mathematikstudium. In *Beiträge zum Mathematikunterricht 2021*, Silke Ruwisch & Susanne Prediger Kerstin Hein, Cathleen Heil (Ed.). WTM-Verlag , Münster, Online unter https://eldorado.tu–dortmund.de/handle/2003/30630.

[45] Lara Gildehaus and Michael Liebendörfer. 2021. "I don't need this" – understanding preservice teachers disaffection with university mathematics. In *Proceedings of the 44th Conference of the International Group for the Psychology of Mathematics Education*, Changsri N. & Boonsena N Inprasitha, M. (Ed.). PME, Khon Kaen, 350–359.

[46] Allison Godwin, Geoff Potvin, Zahra Hazari, and Robynne Lock. 2016. Identity, Critical Agency, and Engineering: An Affective Model for Predicting Engineering as a Career Choice. *Journal of Engineering Education* 105, 2 (2016), 312–340. https://doi.org/10.1002/jee.20118

[47] Heather Honoré Goltz and Matthew Lee Smith. 2014. Forming and developing your professional identity: Easy as PI. *Health promotion practice* 15, 6 (2014), 785–789.

[48] Don Gotterbarn, Amy Bruckman, Catherine Flick, Keith Miller, and Marty J. Wolf. 2017. ACM Code of Ethics: A Guide for Positive Action. *Commun. ACM* 61, 1 (Dec. 2017), 121–128. https://doi.org/10.1145/3173016

[49] Gregor Große-Bölting, Yannick Schneider, and Andreas Mühling. 2019. It's like Computers Speak a Different Language: Beginning Students' Conceptions of Computer Science. In *Proceedings of the 19th Koli Calling International Conference on Computing Education Research* (Koli, Finland) *(Koli Calling '19)*. Association for Computing Machinery, New York, NY, USA, Article 2, 5 pages. https://doi.org/10.1145/3364510.3364527

[50] Gregor Große-Bölting, Yannick Schneider, and Andreas Mühling. 2020. Beginning Students' Conceptions of Computer Science: The Effect of the First Semester. In *2020 International Conference on Learning and Teaching in Computing and Engineering (LaTICE)*. Ho-Chi-Minh-City; Vietnam.

[51] G. Große-Bölting, D. Gerstenberger, L. Gildehaus, A. Mühling, and C. Schulte. 2021. Identity in K-12 Computer Education Research: A Systematic Literature Review. In *Proceedings of the 2021 ACM Conference on International Computing Education Research (ICER '21)*. Association for Computing Machinery, New York, NY, USA.

[52] K Vernard Harrington, Tracy Lewis, Wanda Smith, and France Belanger. 2007. A Qualitative Analysis of Computing Students' Professional Identity and its Relationship to Strategies for Coping with Stressors in the Computing Disciplines. *AMCIS 2007 Proceedings* (2007), 492.

[53] Zahra Hazari, Gerhard Sonnert, Philip M Sadler, and Marie-Claire Shanahan. 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 (2010), 978–1003.

[54] Michael Hewner. 2013. Undergraduate Conceptions of the Field of Computer Science. In *Proceedings of the Ninth Annual International ACM Conference on International Computing Education Research* (San Diego, San California, USA) *(ICER '13)*. Association for Computing Machinery, New York, NY, USA, 107–114. https://doi.org/10.1145/2493394.2493414

[55] Maureen P. Hogan. 2008. The Tale of Two Noras: How a Yup'ik Middle Schooler was Differently Constructed as a Math Learner. *Diaspora, Indigenous, and Minority Education* 2, 2 (2008), 90–114. https://doi.org/10.1080/15595690801894079

[56] Dorothy C Holland, William Lachicotte Jr, Debra Skinner, and Carole Cain. 2001. *Identity and agency in cultural worlds.* Harvard University Press.

[57] Bryce E Hughes, William J Schell, and Brett Tallman. 2018. Development of Leadership Self-Efficacy: Comparing Engineers, Other STEM, and Non-STEM Majors. In *2018 IEEE Frontiers in Education Conference (FIE).* IEEE, 1–5.

[58] Herminia Ibarra. 1999. Provisional selves: Experimenting with image and identity in professional adaptation. *Administrative science quarterly* 44, 4 (1999), 764–791.

[59] Jingoo Kang, Tuula Keinonen, and Anssi Salonen. 2019. Role of Interest and Self-Concept in Predicting Science Aspirations: Gender Study. *Research in Science Education* (Nov. 2019). https://doi.org/10.1007/s11165-019-09905-w

[60] Amanpreet Kapoor and Christina Gardner-McCune. 2018. Understanding professional identities and goals of computer science undergraduate students. In *Proceedings of the 49th ACM Technical Symposium on Computer Science Education.* 191–196.

[61] Amanpreet Kapoor and Christina Gardner-McCune. 2019. Understanding CS undergraduate students' professional identity through the lens of their professional development. In *Proceedings of the 2019 ACM Conference on Innovation and Technology in Computer Science Education.* 9–15.

[62] Amanpreet Kapoor and Christina Gardner-Mccune. 2022. Categorizing Research on Identity in Undergraduate Computing Education. In *Proceedings of the 22nd Koli Calling International Conference on Computing Education Research* (Koli, Finland) *(Koli Calling '22).* Association for Computing Machinery, New York, NY, USA, Article 16, 13 pages. https://doi.org/10.1145/3564721.3565948

[63] Päivi Kinnunen, Matthew Butler, Michael Morgan, Aletta Nylen, Anne-Kathrin Peters, Jane Sinclair, Sara Kalvala, and Erkki Pesonen. 2018. Understanding initial undergraduate expectations and identity in computing studies. *European Journal of Engineering Education* 43, 2 (2018), 201–218.

[64] Barbara Kitchenham. 2004. Procedures for Performing Systematic Reviews. *Keele, UK, Keele Univ.* 33 (Aug. 2004).

[65] Siu Cheung Kong and Yi Qing Wang. 2020. Formation of computational identity through computational thinking perspectives development in programming learning: A mediation analysis among primary school students. *Computers in Human Behavior* 106 (2020), 106230. https://doi.org/10.1016/j.chb.2019.106230

[66] Amy Kramer, Cassie Wallwey, Genevieve Thanh, Emily Dringenberg, and Rachel Kajfez. 2019. A Narrative-Style Exploration of Undergraduate Engineering Students' Beliefs about Smartness and Identity. In *2019 IEEE Frontiers in Education Conference (FIE).* IEEE, 1–9.

[67] Jean Lave and Etienne Wenger. 1991. *Situated Learning: Legitimate Peripheral Participation.* Cambridge University Press, Cambridge. https://doi.org/10.1017/CBO9780511815355

[68] Luzia Leifheit, Katerina Tsarava, Manuel Ninaus, Klaus Ostermann, Jessika Golle, Ulrich Trautwein, and Korbinian Moeller. 2020. SCAPA: Development of a Questionnaire Assessing Self-Concept and Attitudes Toward Programming. In *Proceedings of the 2020 ACM Conference on Innovation and Technology in Computer Science Education (ITiCSE '20).* Association for Computing Machinery, New York, NY, USA, 138–144. https://doi.org/10.1145/3341525.3387415

[69] Colleen Lewis, Paul Bruno, Jonathan Raygoza, and Julia Wang. 2019. Alignment of Goals and Perceptions of Computing Predicts Students' Sense of Belonging in Computing. In *Proceedings of the 2019 ACM Conference on International Computing Education Research* (Toronto ON, Canada) *(ICER '19).* Association for Computing Machinery, New York, NY, USA, 11–19. https://doi.org/10.1145/3291279.3339426

[70] Colleen M Lewis, Ruth E Anderson, and Ken Yasuhara. 2016. " I Don't Code All Day" Fitting in Computer Science When the Stereotypes Don't Fit. In *Proceedings of the 2016 ACM conference on international computing education research.* 23–32.

[71] Lilian L. Pozzer and Phoebe A. Jackson. 2015. Conceptualizing Identity in Science Education Research: Theoretical and Methodological Issues. In *Sociocultural Studies and Implications for Science Education*, Catherine Milne, Kenneth Tobin, and Donna DeGennaro (Eds.). Springer Netherlands, Dordrecht, 213–230.

[72] Michael C Loui. 2005. Ethics and the development of professional identities of engineering students. *Journal of Engineering Education* 94, 4 (2005), 383–390.

[73] Jonathan Mahadeo, Zahra Hazari, and Geoff Potvin. 2020. Developing a Computing Identity Framework: Understanding Computer Science and Information Technology Career Choice. *ACM Trans. Comput. Educ.* 20, 1, Article 7 (Jan. 2020), 14 pages. https://doi.org/10.1145/3365571

[74] Mary Lou Maher, Bojan Cukic, Larry Mays, Steven Rogelberg, Celine Latulipe, Jamie Payton, Audrey Rorrer, and Tonya Frevert. 2016. The Connected Learner: Engaging faculty to connect computing students to peers, profession and purpose. In *2016 IEEE Frontiers in Education Conference (FIE).* IEEE, 1–8.

[75] James E Marcia. 1966. Development and validation of ego-identity status. *Journal of personality and social psychology* 3, 5 (1966), 551.

[76] Eric Margolis and Stephen Laurence. 2021. Concepts. In *The Stanford Encyclopedia of Philosophy* (Spring 2021 ed.), Edward N. Zalta (Ed.). Metaphysics Research Lab, Stanford University.

[77] Philipp Mayring. 2014. *Qualitative content analysis: theoretical foundation, basic procedures and software solution.* AUT.

[78] Robert McCartney and Kate Sanders. 2015. School/Work: Development of Computing Students' Professional Identity at University. In *Proceedings of the eleventh annual International Conference on International Computing Education Research.* 151–159.

[79] George H Mead. 1913. The social self. *The Journal of Philosophy, Psychology and Scientific Methods* 10, 14 (1913), 374–380.

[80] Debbie Meharg, Sandra Cairncross, and Alison Varey. 2018. "So far back, I'm anonymous": Exploring Student Identity using Photovoice. In *2018 IEEE Frontiers in Education Conference (FIE).* IEEE, 1–9.

[81] Gunjan Tomer Sushanta Kumar Mishra. 2016. Professional identity construction among software engineering students: A study in India. *Information Technology & People* 29, 1 (2016), 146–172.

[82] John R. Morelock. 2017. A systematic literature review of engineering identity: definitions, factors, and interventions affecting development, and means of measurement. *European Journal of Engineering Education* 42, 6 (Nov. 2017), 1240–1262. https://doi.org/10.1080/03043797.2017.1287664

[83] Erica Mourão, João Felipe Pimentel, Leonardo Murta, Marcos Kalinowski, Emilia Mendes, and Claes Wohlin. 2020. On the performance of hybrid search strategies for systematic literature reviews in software engineering. *Information and Software Technology* 123 (2020), 106294.

[84] Greg L. Nelson and Andrew J. Ko. 2018. On Use of Theory in Computing Education Research. In *Proceedings of the 2018 ACM Conference on International Computing Education Research* (Espoo, Finland) *(ICER '18)*. Association for Computing Machinery, New York, NY, USA, 31–39. https://doi.org/10.1145/3230977.3230992

[85] Whitney Nelson, Robert T Cummings, and Kinnis Gosha. 2019. A Qualitative Investigation on the Effectiveness of a Computing Identity Development Emailing List for African American Computer Scientists. In *2019 IEEE Frontiers in Education Conference (FIE)*. IEEE, 1–7.

[86] Aletta Nylén, Mats Daniels, Arnold Pears, Åsa Cajander, Roger McDermott, and Ville Isomöttönen. 2018. Why are we here? The Educational Value Model (EVM) as a framework to investigate the role of students' professional identity development. In *2018 IEEE Frontiers in Education Conference (FIE)*. IEEE, 1–8.

[87] Daphna Oyserman, Kristen Elmore, and Smith. 2012. Self, self-concept, and identity. In *Handbook of self and identity*. 69–104.

[88] Rick Parker. 2017. How do you feel: Affective expressions from computer science senior capstone projects. In *2017 International Conference on Learning and Teaching in Computing and Engineering (LaTICE)*. IEEE, 35–42.

[89] Rick Parker. 2018. Who I am becoming, now: Toward a computer science professional identity instrument. In *2018 IEEE Frontiers in Education Conference (FIE)*. IEEE, 1–9.

[90] Anita Patrick and Maura Borrego. 2016. A Review of the Literature Relevant to Engineering Identity.. In *American Society for Engineering Education Annual Conference,*.

[91] Reinhard Pekrun and Raymond P Perry. 2014. Control-value theory of achievement emotions. In *International handbook of emotions in education*. Routledge, 130–151.

[92] Anne-Kathrin Peters. 2017. Learning Computing at University: Participation and Identity : A Longitudinal Study. (2017).

[93] Anne-Kathrin Peters. 2018. Students' experience of participation in a discipline—a longitudinal study of computer science and IT engineering students. *ACM Transactions on Computing Education (TOCE)* 19, 1 (2018), 1–28.

[94] Anne-Kathrin Peters, Anders Berglund, Anna Eckerdal, and Arnold Pears. 2014. First year computer science and IT students' experience of participation in the discipline. In *2014 International Conference on Teaching and Learning in Computing and Engineering*. IEEE, 1–8.

[95] Anne-Kathrin Peters, Anders Berglund, Anna Eckerdal, and Arnold Pears. 2015. Second year computer science and IT students' experience of participation in the discipline. In *Proceedings of the 15th Koli Calling Conference on Computing Education Research*. 68–76.

[96] Anne-Kathrin Peters, Waqar Hussain, Asa Cajander, Tony Clear, and Mats Daniels. 2015. Preparing the global software engineer. In *2015 IEEE 10th International Conference on Global Software Engineering*. IEEE, 61–70.

[97] Anne-Kathrin Peters and Arnold Pears. 2013. Engagement in Computer Science and IT–What! A Matter of Identity?. In *2013 Learning and Teaching in Computing and Engineering*. IEEE, 114–121.

[98] Patrice Potvin, Abdelkrim Hasni, Ousmane Sy, and Martin Riopel. 2018. Two Crucial Years of Science and Technology Schooling: A Longitudinal Study of the Major Influences on and Interactions Between Self-Concept, Interest, and the Intention to Pursue S&T. *Research in Science Education* (Aug. 2018). https://doi.org/10.1007/s11165-018-9751-6

[99] Darinka Radovic, Laura Black, Julian Williams, and Christian E. Salas. 2018. Towards conceptual coherence in the research on mathematics learner identity: a systematic review of the literature. *Educational Studies in Mathematics* 99, 1 (Sept. 2018), 21–42. https://doi.org/10.1007/s10649-018-9819-2

[100] Inioluwa Deborah Raji, Morgan Klaus Scheuerman, and Razvan Amironesei. 2021. You Can't Sit With Us: Exclusionary Pedagogy in AI Ethics Education. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (Virtual Event, Canada) *(FAccT '21)*. Association for Computing Machinery, New York, NY, USA, 515–525. https://doi.org/10.1145/3442188.3445914

[101] Kristy A. Robinson, You-kyung Lee, Emily A. Bovee, Tony Perez, S. Patrick Walton, Daina Briedis, and Lisa Linnenbrink-Garcia. 2019. Motivation in transition: Development and roles of expectancy, task values, and costs in early college engineering. *Journal of Educational Psychology* 111, 6 (2019), 1081–1102. https://doi.org/10.1037/edu0000331

[102] Kristy A. Robinson, Tony Perez, Amy K. Nuttall, Cary J. Roseth, and Lisa Linnenbrink-Garcia. 2018. From science student to scientist: Predictors and outcomes of heterogeneous science identity trajectories in college. *Developmental psychology* 54, 10 (2018), 1977–1992. https://doi.org/10.1037/dev0000567

[103] Sarah L Rodriguez and Kathleen Lehman. 2017. Developing the next generation of diverse computer scientists: the need for enhanced, intersectional computing identity theory. *Computer Science Education* 27, 3-4 (2017), 229–247.

[104] Sarah L Rodriguez, Charles Lu, and Morgan Bartlett. 2018. Engineering identity development: A review of the higher education literature. *International journal of education in mathematics, science and technology* 6, 3 (2018), 254–265.

[105] Jacqueline Rohde, Lisa Musselman, Brianna Benedict, Dina Verdín, Allison Godwin, Adam Kirn, Lisa Benson, and Geoff Potvin. 2019. Design experiences, engineering identity, and belongingness in early career electrical and computer engineering students. *IEEE Transactions on Education* 62, 3 (2019), 165–172.

[106] Edgar H Schein et al. 1978. *Career dynamics: Matching individual and organizational needs*. Vol. 6834. Addison-Wesley.

[107] Theresa Schnettler, Julia Bobe, Anne Scheunemann, Stefan Fries, and Carola Grunschel. 2020. Is it still worth it? Applying expectancy-value theory to investigate the intraindividual motivational process of forming intentions to drop out from university. *Motivation and Emotion* (2020).

https://doi.org/10.1007/s11031-020-09822-w

[108] Michael James Scott and Gheorghita Ghinea. 2014. Measuring enrichment: the assembly and validation of an instrument to assess student self-beliefs in CS1. In *Proceedings of the tenth annual conference on International computing education research*. 123–130.

[109] Catherine Sebastian, Stephanie Burnett, and Sarah-Jayne Blakemore. 2008. Development of the self-concept during adolescence. *Trends in Cognitive Sciences* 12, 11 (2008), 441–446. https://doi.org/10.1016/j.tics.2008.07.008

[110] Anna Sfard and Anna Prusak. 2005. Telling Identities: In Search of an Analytic Tool for Investigating Learning as a Culturally Shaped Activity. *Educational Researcher* 34 (May 2005), 14–22. https://doi.org/10.3102/0013189X034004014

[111] Jane Sinclair and Sara Kalvala. 2015. Exploring societal factors affecting the experience and engagement of first year female computer science undergraduates. In *Proceedings of the 15th Koli Calling Conference on Computing Education Research*. 107–116.

[112] Courtney S Smith-Orr and Andrew Garnett. 2016. Motivation and identity in C++ the effects of music in an engineering classroom. In *2016 IEEE Frontiers in Education Conference (FIE)*. IEEE, 1–5.

[113] Margaret Beale Spencer, Davido Dupree, and Tracey Hartmann. 1997. A phenomenological variant of ecological systems theory (PVEST): A self-organization perspective in context. *Development and psychopathology* 9, 4 (1997), 817–833.

[114] Bernd Carsten Stahl, Job Timmermans, and Brent Daniel Mittelstadt. 2016. The Ethics of Computing: A Survey of the Computing-Oriented Literature. *Comput. Surveys* 48, 4 (Feb. 2016), 55:1–55:38. https://doi.org/10.1145/2871196

[115] Reed Stevens, Kevin O'Connor, Lari Garrison, Andrew Jocuns, and Daniel M. Amos. 2008. Becoming an Engineer: Toward a Three Dimensional View of Engineering Learning. *Journal of Engineering Education* 97, 3 (2008), 355–368. https://doi.org/10.1002/j.2168-9830.2008.tb00984.x arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1002/j.2168-9830.2008.tb00984.x

[116] Sheldon Stryker and Peter J Burke. 2000. The past, present, and future of an identity theory. *Social psychology quarterly* (2000), 284–297.

[117] Mohsen Taheri, Monique Ross, Zahra Hazari, Mark Weiss, Michael Georgiopoulos, Ken Christensen, Tiana Solis, Atalie Garcia, and Deepa Chari. 2018. A Structural Equation Model Analysis of Computing Identity Sub-Constructs and Student Academic Persistence. In *2018 IEEE Frontiers in Education Conference (FIE)*. IEEE, 1–7.

[118] Erika D Tate and Marcia C Linn. 2005. How does identity shape the experiences of women of color engineering students? *Journal of Science Education and Technology* 14, 5-6 (2005), 483–493.

[119] Ella Taylor-Smith, Sally Smith, and Colin Smith. 2019. Identity and Belonging for Graduate Apprenticeships in Computing: The experience of first cohort degree apprentices in Scotland. In *Proceedings of the 2019 ACM Conference on Innovation and Technology in Computer Science Education*. 2–8.

[120] Farah Tokmic, Syeda Fatema Mazumder, Audrey Rorrer, Tonya Frevert, Mary Lou Maher, and Celine Latulipe. 2019. Salient Measures of an Engaged Computing Education Community. In *2019 IEEE Frontiers in Education Conference (FIE)*. IEEE, 1–5.

[121] Karen L Tonso. 2006. Student engineers and engineer identity: Campus engineer identities as figured world. *Cultural studies of science education* 1, 2 (2006), 273–307.

[122] Stuart A. Tross, Jeffrey P. Harper, Lewis W. Osher, and Linda M. Kneidinger. 2000. Not Just the Usual Cast of Characteristics: Using Personality To Predict College Performance and Retention. *Journal of College Student Development* 41, 3 (2000), 323–34.

[123] Lars Ulriksen, Lene Møller Madsen, and Henriette T Holmegaard. 2010. What do we know about explanations for drop out/opt out among young people from STM higher education programmes? *Studies in science education* 46, 2 (2010), 209–244.

[124] Moshe Y. Vardi. 2018. Are We Having an Ethical Crisis in Computing? *Commun. ACM* 62, 1 (Dec. 2018), 7. https://doi.org/10.1145/3292822

[125] Monique Verhoeven, Astrid M. G. Poorthuis, and Monique Volman. 2019. The Role of School in Adolescents' Identity Development. A Literature Review. *Educational Psychology Review* 31, 1 (2019), 35–63. https://doi.org/10.1007/s10648-018-9457-3

[126] Mikko Vesisenaho, Helena Puhakka, Jussi Silvonen, Erkki Sutinen, Marjatta Vanhalakka-Ruoho, Pirkko Voutilainen, and Leena Penttinen. 2009. Need for study and career counselling in computer science. In *2009 39th IEEE Frontiers in Education Conference*. IEEE, 1–6.

[127] Alicia N Washington, Shaefny Grays, and Sudhipta Dasmohapatra. 2016. The Computer Science Attitude and Identity Survey (CSAIS): A Novel Tool for Measuring the Impact of Ethnic Identity in Underrepresented Computer Science Students. In *Proceedings of the ASEE's 123rd Annual Conference & Exposition 2016*.

[128] Etienne Wenger. 1999. *Communities of practice: Learning, meaning, and identity*. Cambridge university press.

[129] Robin Whittemore, Susan K Chase, and Carol Lynn Mandle. 2001. Validity in qualitative research. *Qualitative health research* 11, 4 (2001), 522–537.

[130] Jesse L. M. Wilkins. 2004. Mathematics and Science Self-Concept: An International Investigation. *The Journal of Experimental Education* 72, 4 (July 2004), 331–346. https://doi.org/10.3200/JEXE.72.4.331-346

[131] Claes Wohlin. 2014. Guidelines for Snowballing in Systematic Literature Studies and a Replication in Software Engineering. In *Proceedings of the 18th International Conference on Evaluation and Assessment in Software Engineering* (London, England, United Kingdom) *(EASE '14)*. Association for Computing Machinery, New York, NY, USA, Article 38, 10 pages. https://doi.org/10.1145/2601248.2601268

Appendix B

Questionnaires

B.1 First survey in winter semester 21/22

B1. **Bitte geben Sie für die anonyme Zuordnung von weiteren Befragungen einen Code an, den Sie wie folgt selbst konstruieren:**

Sie nehmen die ersten beiden Buchstaben des Vornamens Ihrer Mutter, gefolgt vom (zweistelligen) Geburtsmonat Ihrer Mutter. Anschließen nehmen Sie analog dazu die ersten beiden Buchstaben und den (zweistelligen) Geburtsmonats Ihres Vaters.

Beispiel: Heißt Ihre Mutter Anna und ist im Oktober geboren und Ihr Vater heißt Bernd und ist im Januar geboren, so ist Ihr Code: an10be01

C1. **An welcher Hochschule studieren Sie?**

Universität Kiel ☐

Universität Paderborn ☐

C2. **Ist dies Ihr erstes Studium?**

Ja ☐

Nein, ich habe gewechselt von (Universität + Fach): ☐

C3. Wie ist Ihre Studiensituation?

Erstes Semester - Informatik im 1-Fach Studiengang ☐
Höheres Semester - Informatik im 1-Fach Studiengang ☐
Erstes Semester - Informatik im 2-Fach Studiengang ☐
Höheres Semester - Informatik im 2-Fach Studiengang ☐
Wirtschaftsinformatik ☐
Informatik im Nebenfach ☐
Sonstiges ☐

Sonstiges

C4. Welches Geschlecht haben Sie?

weiblich ☐
männlich ☐
Sonstiges ☐

Sonstiges

C5. Wie alt sind Sie?

Unter 17 ☐
17 ☐
18 ☐
19 ☐
20 ☐
21 - 25 ☐
Über 25 ☐

D1. **Was *tun* Informatiker*innen?**

Zählen Sie bitte im Folgenden so viele unterschiedliche, für Informatiker*innen typische Handlungen auf, wie Ihnen einfallen. Trennen Sie ihre Aufzählung mit Kommata oder schreiben Sie pro Zeile eine Handlung auf.

Programmieren, ...

E1. **Haben Sie bereits in irgendeiner Form programmiert?**

Ja ☐

Nein ☐

E2. **Wir schätzen Sie Ihre Programmierkenntnisse zu diesem Zeitpunkt ein?**

1 ☐

2 ☐

3 ☐

4 ☐

5 ☐

E3. **Wie viele Zeilen umfasst das längste von Ihnen selbst bzw. im Team entwickelte Programm?**

Weniger als 100 Zeilen ☐

Weniger als 1000 Zeilen ☐

Weniger als 10000 Zeilen ☐

Mehr als 10000 Zeilen ☐

E4. **In wie vielen Programmiersprachen haben Sie bereits selbst programmiert?**

1 ☐

2 ☐

3 ☐

4 ☐

5 oder mehr ☐

E5. **In welchem Alter haben Sie das erste Mal programmiert?**

☐☐☐☐☐☐☐☐☐☐

E6. **Wie sind Sie dazu gekommen, selbst zu programmieren?**

Durch die Schule (z.B. Schulfach, Projektwoche, …) ☐

Durch außerschulische Aktivitäten (z. B. Freizeitangebote, Computerkurse, …) ☐

Durch meine Eltern ☐

Durch Freunde ☐

Durch Medien (Buch, Film, youTube, …) ☐

Sonstiges ☐

Sonstiges

E7. **Mit welcher Programmiersprache / welchem System hatten Sie den ersten Kontakt zur Programmierung?**

Java ☐

Processing ☐

Scratch ☐

Lego Mindstorms ☐

Arduino / Rasberry Pi ☐

Javascript ☐

Python ☐

Ruby ☐

PHP ☐

C / C++ / C# ☐

Swift / Cocoa ☐

Sonstiges ☐

Sonstiges

F1. Inwieweit treffen die folgenden Aussagen auf Sie zu?

	Stimme überhaupt nicht zu				Stimme voll zu
Meine Familie sieht mich als computeraffine Person	☐	☐	☐	☐	☐
Meine Freunde/Kommilitonen sehen mich als computeraffine Person	☐	☐	☐	☐	☐
Meine Ausbilder/Dozenten sehen mich als computeraffine Person	☐	☐	☐	☐	☐
Computerthemen wecken meine Neugierde	☐	☐	☐	☐	☐
Ich schaue mir gerne Foren, soziale Medien oder Online-Videos zu computerbezogenen Themen an	☐	☐	☐	☐	☐
Computerprogrammierung ist für mich interessant	☐	☐	☐	☐	☐
Ich kann Computeraufgaben gut bewältigen (z. B. Programmieren und Einrichten von Servern)	☐	☐	☐	☐	☐
Ich verstehe die Konzepte, die den Computerprozessen zugrunde liegen	☐	☐	☐	☐	☐
Andere bitten mich um Hilfe bei Software (Anwendungen/Programme)	☐	☐	☐	☐	☐
Ich reflektiere meinen Umgang mit Computern und Programmierung	☐	☐	☐	☐	☐
Meine Auseinandersetzung mit Computern und computerbezogenen Themen folgt Werten, Idealen oder festen Vorstellungen	☐	☐	☐	☐	☐
Ich sehe mich als Fachperson mit einer professionellen Verantwortung	☐	☐	☐	☐	☐
Ich fühle mich verantwortlich für das, was ich programmiere/mit dem Computer mache	☐	☐	☐	☐	☐

G1. Inwieweit treffen die folgenden Aussagen auf Sie zu?

Falls Sie noch nicht programmieren bzw. meinen, dass Sie auf eine Frage keine Antwort geben können, dann lassen Sie diese bitte aus.

	Stimme überhaupt nicht zu				Stimme voll zu
Ich finde es spannend, neue Ideen durch Programmieren auszudrücken.	☐	☐	☐	☐	☐
Ich bin glücklich, dass ich mich durch das Programmieren ausdrücken kann.	☐	☐	☐	☐	☐
Ich habe eine Leidenschaft für das Programmieren, weil ich damit neue Dinge erschaffen kann.	☐	☐	☐	☐	☐
Ich fühle mich inspiriert, wenn ich mehr Programmierkenntnisse erwerbe.	☐	☐	☐	☐	☐
Programmieren hilft mir Informationen und Wissen mit anderen zu teilen.	☐	☐	☐	☐	☐
Programmieren hilft mir mit anderen zusammenzuarbeiten, um nach Anwendungen im wirklichen Leben zu suchen.	☐	☐	☐	☐	☐
Programmieren verbindet mich mit Menschen, die so denken, wie ich.	☐	☐	☐	☐	☐

	Stimme überhaupt nicht zu				Stimme voll zu
Programmieren ist eng mit meinem/unserem täglichen Leben verbunden.	☐	☐	☐	☐	☐
Nachdem ich Programmieren gelernt habe, habe ich ein tieferes Verständnis dafür, wie digitale Technologie funktioniert.	☐	☐	☐	☐	☐
Programmieren kann mein Verständnis für die technologische Welt stärken.	☐	☐	☐	☐	☐
Ich denke über das Wissen nach, das ich durch das Programmieren erworben habe.	☐	☐	☐	☐	☐
Ich denke sorgfältig über die möglichen Probleme nach, die beim Programmieren auftreten können.	☐	☐	☐	☐	☐
Ich verstehe, dass jedes Programmierwerkzeug (z. B. Python o. Java) seine Grenzen hat.	☐	☐	☐	☐	☐

G2. Inwieweit treffen die folgenden Aussagen auf Sie zu?

Falls Sie noch nicht programmieren bzw. meinen, dass Sie auf eine Frage keine Antwort geben können, dann lassen Sie diese bitte aus.

	Stimme überhaupt nicht zu				Stimme voll zu
Der Inhalt von Programmieraufgaben weckt mein Interesse am Programmierenlernen.	☐	☐	☐	☐	☐
Ich denke, Programmieren macht Spaß.	☐	☐	☐	☐	☐
Ich fühle mich von Programmieraufgaben wirklich angezogen.	☐	☐	☐	☐	☐
Ich fühle mich von Programmieraufgaben angeregt.	☐	☐	☐	☐	☐
Ich fühle mich mit meinen Mitstudierenden verbunden, wenn ich mit ihnen an Programmieraktivitäten teilnehme.	☐	☐	☐	☐	☐
Das Programmierenlernen mit meinen Mitstudierenden gibt mir ein starkes Gefühl der Zugehörigkeit.	☐	☐	☐	☐	☐
Ich diskutiere gerne mit meinen Mitstudierenden über programmierbezogene Themen.	☐	☐	☐	☐	☐
Ich erkenne die Tatsache an, dass ich ein Mitglied einer Programmiervorlesung bin.	☐	☐	☐	☐	☐
Ich möchte mehr über das Programmieren lernen.	☐	☐	☐	☐	☐
Ich möchte Programmierkenntnisse nutzen, um Probleme in der realen Welt zu lösen.	☐	☐	☐	☐	☐
Ich möchte mein Programmierwissen nutzen, um neue Objekte zu entwerfen.	☐	☐	☐	☐	☐
Programmieren wird ein Teil meines Lebens sein.	☐	☐	☐	☐	☐
Ich möchte in einem Beruf arbeiten, in dem ich meine Programmierkenntnisse und -fähigkeiten einsetzen kann.	☐	☐	☐	☐	☐
Ich möchte mit Menschen zusammenarbeiten, die auch gerne programmieren.	☐	☐	☐	☐	☐
Ich möchte an Programmieraktivitäten mit Menschen teilnehmen, die ähnliche Interessen wie ich haben.	☐	☐	☐	☐	☐

	Stimme überhaupt nicht zu				Stimme voll zu
Ich möchte das, was ich beim Programmieren gelernt habe, in meiner zukünftigen Arbeit anwenden.	☐	☐	☐	☐	☐

B.2 Second survey in summer semester 23

B1. **Bitte geben Sie für die anonyme Zuordnung von weiteren Befragungen einen Code an, den Sie wie folgt selbst konstruieren:**

Sie nehmen die ersten beiden Buchstaben des Vornamens Ihrer Mutter, gefolgt vom (zweistelligen) Geburtsmonat Ihrer Mutter. Anschließen nehmen Sie analog dazu die ersten beiden Buchstaben und den (zweistelligen) Geburtsmonats Ihres Vaters.

Beispiel: Heißt Ihre Mutter Anna und ist im Oktober geboren und Ihr Vater heißt Bernd und ist im Januar geboren, so ist Ihr Code: an10be01

C1. **An welcher Hochschule studieren Sie?**

Christian-Albrechts-Universität zu Kiel ☐
Universität Paderborn ☐
RWTH Aachen ☐
KIT - Karlsruher Institut für Technologie ☐
Universität Hildesheim ☐
Sonstiges ☐

Sonstiges

C2. **Ist dies Ihr erstes Studium?**

Ja ☐
Nein, ich habe gewechselt von (Universität + Fach): ☐

C3. **Wie ist Ihre Studiensituation?**

Informatik im 1-Fach Studiengang ☐
Informatik im 2-Fach Studiengang ☐
Wirtschaftsinformatik ☐
Informatik im Nebenfach ☐
Sonstiges ☐

Sonstiges

C4. **In welchem Fachsemester Ihres Studiums befinden Sie sich?**

C5. **Welches Geschlecht haben Sie?**

weiblich ☐
männlich ☐
Sonstiges ☐

Sonstiges

C6. **Wie alt sind Sie?**

Unter 17 ☐
17 ☐
18 ☐
19 ☐
20 ☐
21 - 25 ☐
Über 25 ☐

D1. **Was *tun* Informatiker*innen?**

Zählen Sie bitte im Folgenden so viele unterschiedliche, für Informatiker*innen typische Handlungen auf, wie Ihnen einfallen. Trennen Sie ihre Aufzählung mit Kommata oder schreiben Sie pro Zeile eine Handlung auf.

Programmieren, ...

E1. **Wir schätzen Sie Ihre Programmierkenntnisse zu diesem Zeitpunkt ein?**

1 ☐
2 ☐
3 ☐
4 ☐
5 ☐

E2. **Wie viele Zeilen umfasst das längste von Ihnen selbst bzw. im Team entwickelte Programm?**

Weniger als 100 Zeilen ☐
Weniger als 1.000 Zeilen ☐
Weniger als 10.000 Zeilen ☐
Weniger als 100.000 Zeilen ☐
Mehr als 100.000 Zeilen ☐

E3. **In wie vielen Programmiersprachen haben Sie bereits selbst programmiert?**

1 ☐
2 ☐
3 ☐
4 ☐
5 oder mehr ☐

E4. **In welchem Alter haben Sie das erste Mal programmiert?**

E5. **Wie sind Sie dazu gekommen, selbst zu programmieren?**

- Durch die Universität/Hochschule ☐
- Durch die Schule (z.B. Schulfach, Projektwoche, ...) ☐
- Durch außerschulische Aktivitäten (z. B. Freizeitangebote, Computerkurse, ...) ☐
- Durch meine Eltern ☐
- Durch Freunde ☐
- Durch Medien (Buch, Film, youTube, ...) ☐
- Sonstiges ☐

Sonstiges

☐

E6. **Mit welcher Programmiersprache / welchem System hatten Sie den ersten Kontakt zur Programmierung?**

- Java ☐
- Processing ☐
- Scratch ☐
- Lego Mindstorms ☐
- Arduino / Rasberry Pi ☐
- Javascript ☐
- Python ☐
- Ruby ☐
- PHP ☐
- C / C++ / C# ☐
- Swift / Cocoa ☐
- Sonstiges ☐

Sonstiges

☐

F1. Inwieweit treffen die folgenden Aussagen auf Sie zu?

	Stimme überhaupt nicht zu				Stimme voll zu
Meine Familie sieht mich als computeraffine Person	☐	☐	☐	☐	☐
Meine Freunde/Kommilitonen sehen mich als computeraffine Person	☐	☐	☐	☐	☐
Meine Ausbilder/Dozenten sehen mich als computeraffine Person	☐	☐	☐	☐	☐
Computerthemen wecken meine Neugierde	☐	☐	☐	☐	☐
Ich schaue mir gerne Foren, soziale Medien oder Online-Videos zu computerbezogenen Themen an	☐	☐	☐	☐	☐
Computerprogrammierung ist für mich interessant	☐	☐	☐	☐	☐
Ich kann Computeraufgaben gut bewältigen (z. B. Programmieren und Einrichten von Servern)	☐	☐	☐	☐	☐
Ich verstehe die Konzepte, die den Computerprozessen zugrunde liegen	☐	☐	☐	☐	☐
Andere bitten mich um Hilfe bei Software (Anwendungen/Programme)	☐	☐	☐	☐	☐
Ich reflektiere meinen Umgang mit Computern und Programmierung	☐	☐	☐	☐	☐
Meine Auseinandersetzung mit Computern und computerbezogenen Themen folgt Werten, Idealen oder festen Vorstellungen	☐	☐	☐	☐	☐
Ich sehe mich als Fachperson mit einer professionellen Verantwortung	☐	☐	☐	☐	☐
Ich fühle mich verantwortlich für das, was ich programmiere/mit dem Computer mache	☐	☐	☐	☐	☐

Appendix C

Factor Analysis: Model Results

Table C.1. Model: Baseline. * Indicates parameters fixed for model identification.

Parameter	Estimate	SE	z	p
	Factor Loadings			
base				
IDM_REC1	1.00*	0.00		
IDM_REC2	1.16	0.13	9.22	.000
IDM_REC3	0.85	0.13	6.61	.000
IDM_INT1	0.64	0.10	6.33	.000
IDM_INT2	0.87	0.13	6.45	.000
IDM_INT3	0.34	0.09	3.64	.000
IDM_COMP1	1.08	0.14	7.91	.000
IDM_COMP2	0.93	0.13	7.11	.000
IDM_COMP3	1.24	0.17	7.47	.000
IDM_VAL1	0.64	0.14	4.46	.000
IDM_VAL2	0.47	0.14	3.26	.001
IDM_VAL3	0.77	0.16	4.79	.000
IDM_VAL4	0.21	0.12	1.73	.084
	Intercepts			
IDM_REC1	4.27	0.09	46.03	.000
IDM_REC2	3.71	0.10	37.86	.000
IDM_REC3	3.15	0.10	31.90	.000
IDM_INT1	4.22	0.07	57.00	.000
IDM_INT2	3.58	0.10	35.86	.000
IDM_INT3	4.43	0.07	64.47	.000
IDM_COMP1	3.48	0.10	35.55	.000
IDM_COMP2	3.56	0.09	38.71	.000
IDM_COMP3	3.45	0.12	28.57	.000
IDM_VAL1	3.52	0.10	34.19	.000
IDM_VAL2	3.10	0.10	30.36	.000
IDM_VAL3	2.45	0.11	21.41	.000
IDM_VAL4	4.02	0.09	46.53	.000

Table C.2. Model: Mahadeo Items. * Indicates parameters fixed for model identification.

Parameter	Estimate	SE	z	p
		Factor Loadings		
recognition				
IDM_REC1	1.00*	0.00		
IDM_REC2	1.25	0.11	11.12	.000
IDM_REC3	0.78	0.11	7.12	.000
interest				
IDM_INT1	1.00*	0.00		
IDM_INT2	0.91	0.15	6.16	.000
IDM_INT3	0.60	0.09	7.05	.000
competence				
IDM_COMP1	1.00*	0.00		
IDM_COMP2	0.80	0.10	8.09	.000
IDM_COMP3	1.06	0.13	8.12	.000
		Intercepts		
IDM_REC1	4.27	0.09	45.96	.000
IDM_REC2	3.71	0.10	38.06	.000
IDM_REC3	3.20	0.10	32.50	.000
IDM_INT1	4.22	0.07	57.11	.000
IDM_INT2	3.58	0.10	35.78	.000
IDM_INT3	4.42	0.07	64.50	.000
IDM_COMP1	3.48	0.10	35.45	.000
IDM_COMP2	3.55	0.09	38.59	.000
IDM_COMP3	3.46	0.12	28.82	.000

Table C.3. Model: Mahadeo vs. Values. * Indicates parameters fixed for model identification.

Parameter	Estimate	SE	z	p
		Factor Loadings		
mahadeo				
IDM_REC1	1.00*	0.00		
IDM_REC2	1.16	0.12	9.95	.000
IDM_REC3	0.81	0.12	6.71	.000
IDM_INT1	0.63	0.10	6.49	.000
IDM_INT2	0.85	0.13	6.61	.000
IDM_INT3	0.34	0.09	3.70	.000
IDM_COMP1	1.03	0.13	8.00	.000
IDM_COMP2	0.86	0.12	7.00	.000
IDM_COMP3	1.18	0.16	7.48	.000
values				
IDM_VAL1	1.00*	0.00		
IDM_VAL2	1.05	0.18	5.87	.000
IDM_VAL3	1.03	0.22	4.77	.000
IDM_VAL4	0.38	0.14	2.82	.005
		Intercepts		
IDM_REC1	4.27	0.09	46.00	.000
IDM_REC2	3.71	0.10	37.86	.000
IDM_REC3	3.16	0.10	32.04	.000
IDM_INT1	4.22	0.07	57.01	.000
IDM_INT2	3.58	0.10	35.84	.000
IDM_INT3	4.43	0.07	64.46	.000
IDM_COMP1	3.48	0.10	35.55	.000
IDM_COMP2	3.56	0.09	38.75	.000
IDM_COMP3	3.45	0.12	28.52	.000
IDM_VAL1	3.50	0.10	33.97	.000
IDM_VAL2	3.08	0.10	30.30	.000
IDM_VAL3	2.45	0.11	21.47	.000
IDM_VAL4	4.02	0.09	46.47	.000

Table C.4. Model: VAL complete. * Indicates parameters fixed for model identification.

Parameter	Estimate	SE	z	p
		Factor Loadings		
recognition				
IDM_REC1	1.00*	0.00		
IDM_REC2	1.26	0.11	11.08	.000
IDM_REC3	0.78	0.11	7.10	.000
interest				
IDM_INT1	1.00*	0.00		
IDM_INT2	0.91	0.15	6.12	.000
IDM_INT3	0.61	0.09	7.10	.000
competence				
IDM_COMP1	1.00*	0.00		
IDM_COMP2	0.84	0.10	8.34	.000
IDM_COMP3	1.09	0.13	8.40	.000
values				
IDM_VAL1	1.00*	0.00		
IDM_VAL2	0.99	0.17	5.74	.000
IDM_VAL3	1.03	0.22	4.75	.000
IDM_VAL4	0.37	0.13	2.80	.005
		Intercepts		
IDM_REC1	4.27	0.09	45.98	.000
IDM_REC2	3.71	0.10	38.08	.000
IDM_REC3	3.21	0.10	32.50	.000
IDM_INT1	4.22	0.07	57.11	.000
IDM_INT2	3.58	0.10	35.79	.000
IDM_INT3	4.42	0.07	64.51	.000
IDM_COMP1	3.48	0.10	35.42	.000
IDM_COMP2	3.55	0.09	38.58	.000
IDM_COMP3	3.46	0.12	28.82	.000
IDM_VAL1	3.50	0.10	33.93	.000
IDM_VAL2	3.07	0.10	30.08	.000
IDM_VAL3	2.44	0.11	21.46	.000
IDM_VAL4	4.02	0.09	46.45	.000

Table C.5. Model: without VAL1. * Indicates parameters fixed for model identification.

Parameter	Estimate	SE	z	p
		Factor Loadings		
recognition				
IDM_REC1	1.00*	0.00		
IDM_REC2	1.26	0.11	11.02	.000
IDM_REC3	0.78	0.11	7.11	.000
interest				
IDM_INT1	1.00*	0.00		
IDM_INT2	0.91	0.15	6.15	.000
IDM_INT3	0.60	0.09	7.07	.000
competence				
IDM_COMP1	1.00*	0.00		
IDM_COMP2	0.84	0.10	8.36	.000
IDM_COMP3	1.07	0.13	8.34	.000
values				
IDM_VAL2	1.00*	0.00		
IDM_VAL3	1.92	0.57	3.35	.001
IDM_VAL4	0.48	0.17	2.80	.005
		Intercepts		
IDM_REC1	4.27	0.09	46.00	.000
IDM_REC2	3.71	0.10	38.09	.000
IDM_REC3	3.20	0.10	32.51	.000
IDM_INT1	4.22	0.07	57.11	.000
IDM_INT2	3.58	0.10	35.79	.000
IDM_INT3	4.42	0.07	64.51	.000
IDM_COMP1	3.48	0.10	35.52	.000
IDM_COMP2	3.55	0.09	38.60	.000
IDM_COMP3	3.46	0.12	28.84	.000
IDM_VAL2	3.09	0.10	30.38	.000
IDM_VAL3	2.44	0.11	21.35	.000
IDM_VAL4	4.02	0.09	46.53	.000

Table C.6. Model: without VAL2. * Indicates parameters fixed for model identification.

Parameter	Estimate	SE	z	p
	Factor Loadings			
recognition				
IDM_REC1	1.00*	0.00		
IDM_REC2	1.26	0.11	11.03	.000
IDM_REC3	0.78	0.11	7.10	.000
interest				
IDM_INT1	1.00*	0.00		
IDM_INT2	0.90	0.15	6.16	.000
IDM_INT3	0.60	0.09	7.04	.000
competence				
IDM_COMP1	1.00*	0.00		
IDM_COMP2	0.86	0.10	8.46	.000
IDM_COMP3	1.11	0.13	8.48	.000
values				
IDM_VAL1	1.00*	0.00		
IDM_VAL3	1.16	0.26	4.40	.000
IDM_VAL4	0.34	0.16	2.08	.038
	Intercepts			
IDM_REC1	4.27	0.09	46.00	.000
IDM_REC2	3.71	0.10	38.08	.000
IDM_REC3	3.21	0.10	32.51	.000
IDM_INT1	4.22	0.07	57.11	.000
IDM_INT2	3.58	0.10	35.79	.000
IDM_INT3	4.42	0.07	64.51	.000
IDM_COMP1	3.48	0.10	35.43	.000
IDM_COMP2	3.55	0.09	38.56	.000
IDM_COMP3	3.46	0.12	28.85	.000
IDM_VAL1	3.51	0.10	34.11	.000
IDM_VAL3	2.45	0.11	21.45	.000
IDM_VAL4	4.02	0.09	46.46	.000

Table C.7. Model: without VAL3. * Indicates parameters fixed for model identification.

Parameter	Estimate	SE	z	p
		Factor Loadings		
recognition				
IDM_REC1	1.00*	0.00		
IDM_REC2	1.25	0.11	11.09	.000
IDM_REC3	0.78	0.11	7.10	.000
interest				
IDM_INT1	1.00*	0.00		
IDM_INT2	0.90	0.15	6.14	.000
IDM_INT3	0.60	0.09	7.00	.000
competence				
IDM_COMP1	1.00*	0.00		
IDM_COMP2	0.84	0.10	8.30	.000
IDM_COMP3	1.11	0.13	8.45	.000
values				
IDM_VAL1	1.00*	0.00		
IDM_VAL2	0.58	0.19	3.11	.002
IDM_VAL4	0.18	0.11	1.63	.103
		Intercepts		
IDM_REC1	4.27	0.09	45.96	.000
IDM_REC2	3.71	0.10	38.06	.000
IDM_REC3	3.20	0.10	32.49	.000
IDM_INT1	4.22	0.07	57.11	.000
IDM_INT2	3.58	0.10	35.78	.000
IDM_INT3	4.42	0.07	64.50	.000
IDM_COMP1	3.48	0.10	35.33	.000
IDM_COMP2	3.55	0.09	38.54	.000
IDM_COMP3	3.46	0.12	28.80	.000
IDM_VAL1	3.50	0.10	33.74	.000
IDM_VAL2	3.08	0.10	30.07	.000
IDM_VAL4	4.02	0.09	46.44	.000

Table C.8. Model: without VAL4. * Indicates parameters fixed for model identification.

Parameter	Estimate	SE	z	p
		Factor Loadings		
recognition				
IDM_REC1	1.00*	0.00		
IDM_REC2	1.26	0.11	11.07	.000
IDM_REC3	0.78	0.11	7.10	.000
interest				
IDM_INT1	1.00*	0.00		
IDM_INT2	0.91	0.15	6.12	.000
IDM_INT3	0.60	0.09	7.07	.000
competence				
IDM_COMP1	1.00*	0.00		
IDM_COMP2	0.84	0.10	8.37	.000
IDM_COMP3	1.09	0.13	8.42	.000
values				
IDM_VAL1	1.00*	0.00		
IDM_VAL2	0.95	0.17	5.59	.000
IDM_VAL3	0.99	0.21	4.62	.000
		Intercepts		
IDM_REC1	4.27	0.09	45.98	.000
IDM_REC2	3.71	0.10	38.08	.000
IDM_REC3	3.21	0.10	32.50	.000
IDM_INT1	4.22	0.07	57.11	.000
IDM_INT2	3.58	0.10	35.79	.000
IDM_INT3	4.42	0.07	64.51	.000
IDM_COMP1	3.48	0.10	35.42	.000
IDM_COMP2	3.55	0.09	38.59	.000
IDM_COMP3	3.46	0.12	28.83	.000
IDM_VAL1	3.50	0.10	33.97	.000
IDM_VAL2	3.08	0.10	30.14	.000
IDM_VAL3	2.45	0.11	21.49	.000

Bibliography

[AO05] Hilary Arksey and Lisa O'Malley. "Scoping studies: towards a methodological framework". In: *International Journal of Social Research Methodology* 8.1 (Feb. 2005), pp. 19–32. DOI: `10.1080/1364557032000119616`.

[App18] Kwame Anthony Appiah. *The lies that bind: Rethinking identity*. Profile Books, 2018.

[BB16] Ilias Bergstrom and Alan F. Blackwell. "The practices of programming". In: *2016 IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC)*. IEEE, Sept. 2016. DOI: `10.1109/vlhcc.2016.7739684`.

[BBY08] Maureen Biggers, Anne Brauer, and Tuba Yilmaz. "Student perceptions of computer science: a retention study comparing graduating seniors with cs leavers". In: *Proceedings of the 39th SIGCSE Technical Symposium on Computer Science Education*. SIGCSE '08. Portland, OR, USA: Association for Computing Machinery, 2008, pp. 402–406. DOI: `10.1145/1352135.1352274`.

[BC00] Rogers Brubaker and Frederick Cooper. "Beyond 'identity'". In: *Theory and society* 29.1 (2000), pp. 1–47.

[BC22] Emma O Brien and Aoife Chawke. "Tell me a story: Using personas to enhance student engagement". In: *All Ireland Journal of Higher Education* 14.1 (2022).

[BC95] Eileen D. Bunderson and Mary Elizabeth Christensen. "An analysis of retention problems for female students in university computer science programs". In: *Journal of Research on Computing in Education* 28.1 (1995), pp. 1–18.

[BDK+23] Bailey Braaten, Emily Dringenberg, Amy Kramer, and Rachel Kajfez. "You're an Engineer? You Must Be Really Smart! A Theoretical Discussion of the Need to Integrate "Smart" into Engineering Identity Research". In: *Studies in Engineering Education* 4.2 (2023), pp. 22–37. DOI: `10.21061/see.86`.

[Ber13] Peter Berger. *Computer und Weltbild: habitualisierte Konzeptionen von der Welt der Computer*. Springer-Verlag, 2013.

[Bro15] Timothy A Brown. *Confirmatory factor analysis for applied research*. Guilford publications, 2015.

[BWJ+06] Astrid MW Bulte, Hanna B Westbroek, Onno de Jong, and Albert Pilot. "A research approach to designing chemistry education using authentic practices as contexts". In: *International journal of science education* 28.9 (2006), pp. 1063–1086.

[CCS16] Richard L. Cruess, Sylvia R. Cruess, and Yvonne Steinert. “Amending miller’s pyramid to include professional identity formation”. In: *Academic Medicine* 91.2 (2016). ISSN: 1040-2446. URL: `https://journals.lww.com/academicmedicine/Fulltext/2016/02000/Amending_Miller_s_Pyramid_to_Include_Professional.17.aspx`.

[CJ07] Heidi B Carlone and Angela Johnson. “Understanding the science experiences of successful women of color: Science identity as an analytic lens”. en. In: *Journal of Research in Science Teaching: The Official Journal of the National Association for Research in Science Teaching* 44.8 (2007), pp. 1187–1218. DOI: `10.1002/tea.20237`.

[CL96] Seth Chaiklin and Jean Lave. *Understanding practice: Perspectives on activity and context*. Cambridge University Press, 1996.

[Con90] Jere Confrey. “Chapter 1: A Review of the Research on Student Conceptions in Mathematics, Science, and Programming”. In: *Review of Research in Education* 16.1 (1990), pp. 3–56. DOI: `10.3102/0091732X016001003`.

[CS18] Melanie M. Cooper and Ryan L. Stowe. “Chemistry education research – From personal empiricism to evidence, theory, and informed practice”. In: *Chemical reviews* 118.12 (2018), pp. 6053–6087.

[Dar16] Lisa Darragh. “Identity research in mathematics education”. en. In: *Educational Studies in Mathematics* 93.1 (Sept. 2016), pp. 19–33. ISSN: 1573-0816. DOI: `10.1007/s10649-016-9696-5`. URL: `https://doi.org/10.1007/s10649-016-9696-5`.

[Den05] Peter J. Denning. “Is Computer Science Science?” In: *Commun. ACM* 48.4 (Apr. 2005), pp. 27–31. ISSN: 0001-0782. DOI: `10.1145/1053291.1053309`.

[DK20] Catherine D’Ignazio and Lauren F. Klein. *Data Feminism*. en. The MIT Press, 2020.

[DKG+23] Anna T. Danielsson, Heather King, Spela Godec, and Anne-Sofie Nyström. “The identity turn in science education research: a critical review of methodologies in a consolidating field”. In: *Cultural Studies of Science Education* (Jan. 2023). DOI: `10.1007/s11422-022-10130-7`.

[DM17] Don Davis and Lee L. Mason. “A behavioral phenomenological inquiry of maker identity.” In: *Behavior Analysis: Research and Practice* 17.2 (May 2017), pp. 174–196. DOI: `10.1037/bar0000060`.

[DZ12] Ira Diethelm and Stefan Zumbrägel. “An investigation of secondary school students’ conceptions on how the internet works”. In: *Proceedings of the 12th Koli Calling International Conference on Computing Education Research*. Ed. by Mikko-Jussi Laakso. New York, NY: ACM, 2012, pp. 67–73. ISBN: 9781450317955. DOI: `10.1145/2401796.2401804`.

[Ecc83] J. Eccles. “Expectancies, values and academic behaviors”. In: *Achievement and achievement motives* (1983). (Visited on 07/29/2020).

[Ede07] Amnon H Eden. “Three paradigms of computer science”. In: *Minds and machines* 17.2 (2007), pp. 135–167.

[EMP99] Yrjö Engeström, Reijo Miettinen, and Raija-Leena Punamäki. *Perspectives on activity theory*. Cambridge University Press, 1999.

[Eng14] Yrjö Engeström. *Learning by Expanding: An Activity-Theoretical Approach to Developmental Research*. 2nd ed. Cambridge: Cambridge University Press, 2014. DOI: `10.1017/CBO9781139814744`.

[Eri68] E.H. Erikson. *Identity: youth and crisis*. Identity: youth and crisis. Oxford, England: Norton & Co., 1968.

[EW20] Jacquelynne S. Eccles and Allan Wigfield. “From expectancy-value theory to situated expectancy-value theory: A developmental, social cognitive, and sociocultural perspective on motivation”. In: *Contemporary educational psychology* 61 (2020), p. 101859. DOI: `10.1016/j.cedpsych.2020.101859`.

[FK12] Hanoch Flum and Avi Kaplan. “Identity formation in educational settings: A contextualized view of theory and research in practice”. In: *Contemporary Educational Psychology* 37.3 (2012). Identity Formation in Educational Settings, pp. 240–245. DOI: `https://doi.org/10.1016/j.cedpsych.2012.01.003`.

[Fla54] John C. Flanagan. “The critical incident technique.” In: *Psychological bulletin* 51.4 (1954), p. 327.

[Fle19a] Ludwik Fleck. “Denkstile und Tatsachen. Gesammelte Schriften und Zeugnisse”. In: ed. by Sylwia Werner and Claus Zittel. suhrkamp taschenbuch wissenschaft, 2019. Chap. Über einige spezifische Merkmale des ärztlichen Denkens (1927), pp. 41–51.

[Fle19b] Ludwik Fleck. “Denkstile und Tatsachen. Gesammelte Schriften und Zeugnisse”. In: ed. by Sylwia Werner and Claus Zittel. suhrkamp taschenbuch wissenschaft, 2019. Chap. Schauen, Sehen, Wissen (1947), pp. 390–418.

[Fle19c] Ludwik Fleck. “Denkstile und Tatsachen. Gesammelte Schriften und Zeugnisse”. In: ed. by Sylwia Werner and Claus Zittel. suhrkamp taschenbuch wissenschaft, 2019. Chap. Über die wissenschaftliche Beobachtung und die Wahrnehmung im allgmeinen (1935), pp. 211–238.

[Fle19d] Ludwik Fleck. “Denkstile und Tatsachen. Gesammelte Schriften und Zeugnisse”. In: ed. by Sylwia Werner and Claus Zittel. suhrkamp taschenbuch wissenschaft, 2019. Chap. Wissenschaftstheoretische Probleme (1946), pp. 369–389.

[Fle19e] Ludwik Fleck. “Denkstile und Tatsachen. Gesammelte Schriften und Zeugnisse”. In: ed. by Sylwia Werner and Claus Zittel. suhrkamp taschenbuch wissenschaft, 2019. Chap. Zur Frage der Grundlagen der medizinischen Erkenntnis (1935), pp. 239–259.

[FM02] Allan Fisher and Jane Margolis. “Unlocking the clubhouse: the Carnegie Mellon experience”. In: *ACM SIGCSE Bulletin* 34.2 (June 2002), pp. 79–83. ISSN: 0097-8418. DOI: `10.1145/543812.543836`. (Visited on 08/21/2020).

[Fuk18] Francis Fukuyama. *Identity: The demand for dignity and the politics of resentment*. Farrar, Straus and Giroux, 2018.

[GBD17] Virginia Grande, Anders Berglund, and Mats Daniels. “Experiences of teachers in computing as role models: a phenomenographic study”. In: *Proceedings of the 17th Koli Calling International Conference on Computing Education Research*. Koli Calling ’17. Koli, Finland: Association for Computing Machinery, 2017, pp. 133–137. DOI: `10.1145/3141880.3141901`.

[GD17] Virginia Grande and Mats Daniels. “A diversity lens on the last decade of the fie conference: role models for the engineering community”. In: *2017 IEEE Frontiers in Education Conference (FIE)*. IEEE. 2017, pp. 1–8.

[Gee00] James Paul Gee. “Chapter 3: Identity as an analytic lens for research in education”. In: *Review of research in education* 25.1 (2000), pp. 99–125.

[GGG+21] G. Große-Bölting, D. Gerstenberger, L. Gildehaus, A. Mühling, and C. Schulte. “Identity in K-12 Computer Education Research: A Systematic Literature Review”. In: *Proceedings of the 2021 ACM Conference on International Computing Education Research*. ICER ’21. New York, NY, USA: Association for Computing Machinery, 2021.

[GGG+23] Gregor Große-Bölting, Dietrich Gerstenberger, Lara Gildehaus, Andreas Mühling, and Carsten Schulte. “Identity in Higher Computer Education Research: A Systematic Literature Review”. In: *ACM Trans. Comput. Educ.* (June 2023). Just Accepted. DOI: `10.1145/3606707`.

[Gib03] Donald E. Gibson. “Developing the Professional Self-Concept: Role Model Construals in Early, Middle, and Late Career Stages”. In: *Organization Science* 14.5 (2003), pp. 591–610.

[GPH+16] Allison Godwin, Geoff Potvin, Zahra Hazari, and Robynne Lock. “Identity, Critical Agency, and Engineering: An Affective Model for Predicting Engineering as a Career Choice”. en. In: *Journal of Engineering Education* 105.2 (2016), pp. 312–340. DOI: `10.1002/jee.20118`.

[GPJ+17] Michail Giannakos, Ilias Pappas, Maria Letizia Jaccheri, and Demetrios Sampson. "Understanding student retention in computer science education: the role of environment, gains, barriers and usefulness". In: *Education and Information Technologies* 22.5 (Sept. 2017), pp. 2365–2382.

[GWS21] Dietrich Karl-Heinz Gerstenberger, Felix Winkelnkemper, and Carsten Schulte. "Nutzung der Personas-Methode zum Umgang mit der Heterogenität von Informatik-Studierenden". In: *9. Fachtagung Hochschuldidaktik Informatik (HDI)* (2021), p. 49.

[Hac22] Cort-Denis Hachmeister. *Informatik: Mehr Studienabschlüsse und mehr weibliche Studierende nötig – CHE Hochschuldaten*. de-DE. Online. Nov. 2022. URL: `https://hochschuldaten.che.de/informatik-mehr-studienabschlusse-und-mehr-weibliche-studierende-noetig/` (visited on 08/29/2023).

[Hat08] John Hattie. *Visible learning a synthesis of over 800 meta-analyses relating to achievement. A synthesis of over 800 meta-analyses relating to achievement*. Taylor & Francis Group, 2008. ISBN: 9780203887332.

[Hew13] Michael Hewner. "Undergraduate conceptions of the field of computer science". In: *Proceedings of the Ninth Annual International ACM Conference on International Computing Education Research*. ICER '13. San Diego, San California, USA: Association for Computing Machinery, 2013, pp. 107–114. DOI: `10.1145/2493394.2493414`.

[HHS22] Ulrich Heublein, Christopher Hutzsch, and Robert Schmelzer. *Anhang zum DZHW Brief 05 2022 - Die Entwicklung der Studienabbruchquoten in Deutschland*. Online. May 2022. URL: `https://www.dzhw.eu/pdf/pub_brief/dzhw_brief_05_2022_anhang.pdf`.

[HLS+01] Dorothy C. Holland, William Lachicotte Jr., Debra Skinner, and Carole Cain. *Identity and agency in cultural worlds*. Harvard University Press, 2001.

[HM21] Helen Hickmann and Lydia Malin. *Fachkräftereport September 2021–Der Fachkräftemangel nimmt wieder zu*. Sept. 2021.

[HMM+21] Tra Huynh, Adrian Madsen, Sarah McKagan, and Eleanor Sayre. "Building personas from phenomenography: a method for user-centered design in education". In: *Information and Learning Sciences* 122.11/12 (2021), pp. 689–708.

[HRO21] Cynthia Hampton, David Reeping, and Desen Sevi Ozkan. "Positionality statements in engineering education research: a look at the hand that guides the methodological tools". In: *Studies in Engineering* (Mar. 2021). DOI: `10.21061/see.13`.

[HSS+10] Zahra Hazari, Gerhard Sonnert, Philip M Sadler, and Marie-Claire Shanahan. "Connecting high school physics experiences, outcome expectations, physics identity, and physics career choice: a gender study". In: *Journal of research in science teaching* 47.8 (2010), pp. 978–1003.

[Hub21] Aleata Hubbard Cheuoua. "Confronting inequities in computer science education: a case for critical theory". In: *Proceedings of the 52nd ACM Technical Symposium on Computer Science Education*. SIGCSE '21. Virtual Event, USA: Association for Computing Machinery, 2021, pp. 425–430. DOI: `10.1145/3408877.3432453`.

[Hüs22] Marc Hüsch. *Erstmals mehr weibliche als männliche Studierende an deutschen Hochschulen – CHE Hochschuldaten*. de-DE. Online. Sept. 2022. URL: `https://hochschuldaten.che.de/erstmals-mehr-weibliche-als-maennliche-studierende-an-deutschen-hochschulen/` (visited on 08/29/2023).

[Joa93] Hans Joas. *Pragmatism and social theory*. University of Chicago Press, 1993.

[JW12] Samireh Jalali and Claes Wohlin. "Systematic literature studies: database searches vs. backward snowballing". In: *Proceedings of the ACM-IEEE International Symposium on Empirical Software Engineering and Measurement*. ESEM '12. Lund, Sweden: Association for Computing Machinery, 2012, pp. 29–38. DOI: `10.1145/2372251.2372257`.

[KB14] Kari Kuutti and Liam J. Bannon. "The Turn to Practice in HCI: Towards a Research Agenda". In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '14. Toronto, Ontario, Canada: Association for Computing Machinery, 2014, pp. 3543–3552. DOI: `10.1145/2556288.2557111`.

[KG22] Amanpreet Kapoor and Christina Gardner-Mccune. "Categorizing research on identity in undergraduate computing education". In: *Proceedings of the 22nd Koli Calling International Conference on Computing Education Research*. Koli Calling '22. Koli, Finland: Association for Computing Machinery, 2022. DOI: `10.1145/3564721.3565948`.

[KGH+13] Karin Kleinn, Monika Götsch, Yvonne Heine, and Britta Schinzel. "Das DFG-Projekt "Weltbilder der Informatik"". In: *Informatik-Spektrum* 36.3 (2013), pp. 251–256. URL: `https://doi.org/10.1007/s00287-013-0700-5`.

[KH08] Eckhard Klieme and Johannes Hartig. "Kompetenzkonzepte in den Sozialwissenschaften und im erziehungswissenschaftlichen Diskurs". In: *Kompetenzdiagnostik: Zeitschrift für Erziehungswissenschaft* (2008), pp. 11–29.

[Kit04] Barbara Kitchenham. "Procedures for Performing Systematic Reviews". In: *Keele, UK, Keele Univ.* 33 (Aug. 2004).

[KMP13] Päivi Kinnunen, Maija Marttila-Kontio, and Erkki Pesonen. "Getting to know computer science freshmen". In: *Proceedings of the 13th Koli Calling International Conference on Computing Education Research – Koli Calling '13*. ACM Press, 2013. DOI: `10.1145/2526968.2526975`.

[KS05] Maria Knobelsdorf and Carsten Schulte. "Computer biographies-a biographical research perspective on computer usage and attitudes towards informatics". In: *Proceedings of the 5th Annual Finnish/Baltic Sea Conference on Computer Science Education, Koli*. Vol. 2006. 2005.

[KS07] Amy K. Kiefer and Denise Sekaquaptewa. "Implicit stereotypes and women's math performance: how implicit gender-math stereotypes influence women's susceptibility to stereotype threat". In: *Journal of Experimental Social Psychology* 43.5 (2007), pp. 825–832. DOI: `https://doi.org/10.1016/j.jesp.2006.08.004`.

[KW20] Siu Cheung Kong and Yi Qing Wang. "Formation of computational identity through computational thinking perspectives development in programming learning: a mediation analysis among primary school students". In: *Computers in Human Behavior* 106 (2020), p. 106230. DOI: `https://doi.org/10.1016/j.chb.2019.106230`.

[Lav09] Jean Lave. "The practice of learning". In: *Contemporary theories of learning*. Routledge, 2009, pp. 208–216.

[LBR+19] Colleen Lewis, Paul Bruno, Jonathan Raygoza, and Julia Wang. "Alignment of goals and perceptions of computing predicts students' sense of belonging in computing". In: *Proceedings of the 2019 ACM Conference on International Computing Education Research*. ICER '19. Toronto ON, Canada: Association for Computing Machinery, 2019, pp. 11–19. DOI: `10.1145/3291279.3339426`.

[LGB+08] Anna Lichtwarck-Aschoff, Paul van Geert, Harke Bosma, and Saskia Kunnen. "Time and identity: a framework for research and theory formation". In: *Developmental Review* 28.3 (2008), pp. 370–400.

[LW91] Jean Lave and Etienne Wenger. *Situated Learning: Legitimate Peripheral Participation*. Learning in Doing: Social, Cognitive and Computational Perspectives. Cambridge: Cambridge University Press, 1991. DOI: `10.1017/CBO9780511815355`.

[May14] Philipp Mayring. *Qualitative content analysis: theoretical foundation, basic procedures and software solution*. AUT, 2014.

[MBO06] E. M. Mercier, B. Barron, and K. M. O'Connor. "Images of self and others as computer users: the role of gender and experience". In: *Journal of Computer Assisted Learning* 22.5 (2006), pp. 335–348. DOI: `https://doi.org/10.1111/j.1365-2729.2006.00182.x`.

[McC73] David C McClelland. "Testing for competence rather than for intelligence." In: *American psychologist* 28.1 (1973), p. 1.

[MCG+21] Melissa Høegh Marcher, Ingrid Maria Christensen, Paweł Grabarczyk, Therese Graversen, and Claus Brabrand. "Computing educational activities involving people rather than things appeal more to women (CS1 appeal perspective)". In: *Proceedings of the 17th ACM Conference on International Computing Education Research*. ACM, Aug. 2021. DOI: `10.1145/3446871.3469761`.

[MCM16] Allison Master, Sapna Cheryan, and Andrew N Meltzoff. "Computing whether she belongs: stereotypes undermine girls' interest and sense of belonging in computer science." In: *Journal of educational psychology* 108.3 (2016), p. 424.

[Mea13] George H Mead. "The social self". In: *The Journal of Philosophy, Psychology and Scientific Methods* 10.14 (1913), pp. 374–380.

[MHP20] Jonathan Mahadeo, Zahra Hazari, and Geoff Potvin. "Developing a computing identity framework: understanding computer science and information technology career choice". In: *ACM Trans. Comput. Educ.* 20.1 (Jan. 2020). DOI: `10.1145/3365571`.

[Mid18] Mary Midgley. *What Is Philosophy for?* en. Bloomsbury Publishing, Sept. 2018.

[Mid88] Mary Midgley. "Myths of intellectual isolation". In: *Proceedings of the Aristotelian Society* 89 (1988), pp. 19–32.

[Mid90] Mary Midgley. "The use and uselessness of learning". In: *European Journal of Education* 25.3 (1990), pp. 283–294. DOI: `10.2307/1503318`.

[MPM+20] Erica Mourão, João Felipe Pimentel, Leonardo Murta, Marcos Kalinowski, Emilia Mendes, and Claes Wohlin. "On the performance of hybrid search strategies for systematic literature reviews in software engineering". In: *Information and Software Technology* 123 (2020), p. 106294.

[MRP01] N. Mays, E. Roberts, and J. Popay. "Synthesising research evidence". In: *Studying the organisation and delivery of health services: research methods*. Ed. by N Fulop, P Allen, A Clarke, and N Black. London: Routledge, 2001, pp. 188–220. URL: `https://researchonline.lshtm.ac.uk/id/eprint/15408/`.

[MSY09] Reijo Miettinen, Dalvir Samra-Fredericks, and Dvora Yanow. "Re-turn to practice: An introductory essay". In: *Organization studies* 30.12 (2009), pp. 1309–1327.

[NBG02] Brian A Nosek, Mahzarin R Banaji, and Anthony G Greenwald. "Math = male, me = female, therefore math != me." In: *Journal of personality and social psychology* 83.1 (2002), p. 44.

[Nic12] Davide Nicolini. *The social theories of practice underpinning the strategy-as-practice*. Henry Stewart Talks, 2012.

[NK18] Greg L. Nelson and Andrew J. Ko. "On use of theory in computing education research". In: *Proceedings of the 2018 ACM Conference on International Computing Education Research*. ICER '18. Espoo, Finland: Association for Computing Machinery, 2018, pp. 31–39. DOI: `10.1145/3230977.3230992`.

[OHT94] Penelope J Oakes, S Alexander Haslam, and John C Turner. *Stereotyping and social reality*. Blackwell Publishing, 1994.

[Pap05] Marina Papastergiou. "Students' mental models of the internet and their didactical exploitation in informatics education". In: *Education and Information Technologies* 10.4 (2005), pp. 341–360. DOI: `10.1007/s10639-005-3431-7`.

[Pea86] Roy D Pea. "Language-independent conceptual "bugs" in novice programming". In: *Journal of educational computing research* 2.1 (1986), pp. 25–36.

[Pet17] Anne-Kathrin Peters. "Learning Computing at University: Participation and Identity : A Longitudinal Study". eng. In: (2017).

[Pet18] Anne-Kathrin Peters. "Student' experience of participation in a discipline: a longitudinal study of computer science and it engineering students". In: *ACM Trans. Comput. Educ.* 19.1 (Sept. 2018), 5:1–5:28. ISSN: 1946-6226.

[Pfe09] V. Pfeifer. *Ethisch argumentieren: eine Anleitung anhand von aktuellen Fallanalysen*. Ethisch argumentieren / Eine Anleitung anhand von aktuellen Fallanalysen. Schöningh, 2009. ISBN: 9783140250504.

[PHG+13] Lawrence A. Palinkas, Sarah M. Horwitz, Carla A. Green, Jennifer P. Wisdom, Naihua Duan, and Kimberly Hoagwood. "Purposeful sampling for qualitative data collection and analysis in mixed method implementation research". In: *Administration and Policy in Mental Health and Mental Health Services Research* 42.5 (Nov. 2013), pp. 533–544. DOI: `10.1007/s10488-013-0528-y`.

[PP13] Anne-Kathrin Peters and Arnold Pears. "Engagement in Computer Science and IT–What! A Matter of Identity?" In: *2013 Learning and Teaching in Computing and Engineering*. IEEE. 2013, pp. 114–121.

[PR14] Anne-Kathrin Peters and Detlef Rick. "Identity development in computing education: theoretical perspectives and an implementation in the classroom". In: *Proceedings of the 9th Workshop in Primary and Secondary Computing Education*. WiPSCE '14. New York, NY, USA: Association for Computing Machinery, Nov. 2014, pp. 70–79. DOI: `10.1145/2670757.2670787`.

[PR66] James C Parker and Louis-j Rubin. “Process as content: curriculum design and the application of knowledge”. In: *Curriculum series* (1966).

[PW95] William R. Penuel and James V. Wertsch. “Vygotsky and identity formation: a sociocultural approach”. In: *Educational Psychologist* 30.2 (Mar. 1995), pp. 83–92. doi: `10.1207/s15326985ep3002_5`.

[QL17] Yizhou Qian and James Lehman. “Students’ misconceptions and other difficulties in introductory programming: a literature review”. In: *ACM Trans. Comput. Educ.* 18.1 (Oct. 2017). doi: `10.1145/3077618`.

[RBD+20] Steven O. Roberts, Carmelle Bareket-Shavit, Forrest A. Dollins, Peter D. Goldie, and Elizabeth Mortenson. “Racial inequality in psychological research: trends of the past and recommendations for the future”. In: *Perspectives on Psychological Science* 15.6 (2020), pp. 1295–1309. doi: `10.1177/1745691620927709`.

[RLB18] Sarah L Rodriguez, Charles Lu, and Morgan Bartlett. “Engineering identity development: a review of the higher education literature”. In: *International journal of education in mathematics, science and technology* 6.3 (2018), pp. 254–265.

[RP16] Michael T. Rücker and Niels Pinkwart. “Review and discussion of children’s conceptions of computers”. In: *Journal of Science Education and Technology* 25.2 (2016), pp. 274–283. doi: `10.1007/s10956-015-9592-2`.

[SBL+18] Linda J. Sax, Jennifer M. Blaney, Kathleen J. Lehman, Sarah L. Rodriguez, Kari L. George, and Christina Zavala. “Sense of belonging in computing: the role of introductory courses for women and underrepresented minority students”. In: *Social Sciences* 7.8 (July 2018), p. 122. doi: `10.3390/socsci7080122`.

[Sch02] Toni Schmader. “Gender identification moderates stereotype threat effects on women’s math performance”. In: *Journal of Experimental Social Psychology* 38.2 (2002), pp. 194–201. doi: `https://doi.org/10.1006/jesp.2001.1500`.

[Sch93] Andreas Schwill. “Fundamentale Ideen der Informatik”. In: *Zentralblatt für Didaktik der Mathematik* 25.1 (1993), pp. 20–31.

[SdR94] John P. Smith III, Andrea A. diSessa, and Jeremy Roschelle. “Misconceptions reconceived: a constructivist analysis of knowledge in transition”. In: *Journal of the Learning Sciences* 3.2 (1994), pp. 115–163. doi: `10.1207/s15327809jls0302\_1`.

[Ser87] Michel Serres. *Der Parasit*. Ed. by M. Bischoff. Suhrkamp-Taschenbuch Wissenschaft. Suhrkamp, 1987. isbn: 9783518282779.

[SK07] Carsten Schulte and Maria Knobelsdorf. "Attitudes towards computer science-computing experiences as a starting point and barrier to computer science". In: *Proceedings of the third international workshop on Computing education research*. ACM. 2007, pp. 27–38.

[SLM+13] Alexander von Stetten, Sven Laumer, Christian Maier, Tim Weitzel, Andreas Eckhardt, and Elke Guhl. *Recruiting Trends 2013 : Eine empirische Untersuchung mit den Top-1.000-Unternehmen aus Deutschland sowie den Top-300-Unternehmen aus den Branchen Automotive, Finanzdienstleistung und IT*. deu. Tech. rep. Otto-Friedrich-Universität Bamberg und Goethe-Universität Frankfurt am Main, Centre of Human Resources Information Systems (CHRIS), July 2013. URL: `https://fis.uni-bamberg.de/handle/uniba/2361` (visited on 08/29/2023).

[SM17] Yannick Schneider and Andreas Mühling. "Das Konzept Nature of Computer Science". In: *Informatische Bildung zum Verstehen und Gestalten der digitalen Welt, Proceedings der 17. GI-Fachtagung Informatik und Schule, INFOS 2017, 13.-15. September 2017, Oldenburg*. 2017, pp. 123–126. URL: `https://dl.gi.de/20.500.12116/4306`.

[SP05] Anna Sfard and Anna Prusak. "Telling Identities: In Search of an Analytic Tool for Investigating Learning as a Culturally Shaped Activity". In: *Educational Researcher* 34.4 (May 2005), pp. 14–22. DOI: `10.3102/0013189X034004014`.

[Sta20] Statista Research Department. *US: Female computes sciences graduates, by degree*. en. Online. Jan. 2020. URL: `https://www.statista.com/statistics/1251335/female-graduates-computer-sciences-us/` (visited on 08/29/2023).

[TGA+18] Pablo Vicente Torres-Carrión, Carina Soledad González-González, Silvana Aciar, and Germania Rodríguez-Morales. "Methodology for systematic literature review applied to engineering and education". In: *2018 IEEE Global Engineering Education Conference (EDUCON)*. 2018, pp. 1364–1373. DOI: `10.1109/EDUCON.2018.8363388`.

[TS08] Matti Tedre and Erkki Sutinen. "Three traditions of computing: what educators should know". In: *Computer Science Education* 18.3 (2008), pp. 153–170. DOI: `10.1080/08993400802332332`.

[Var07] Roli Varma. "Women in computing: The role of geek culture". In: *Science as culture* 16.4 (2007), pp. 359–376.

[VPV19] Monique Verhoeven, Astrid M. G. Poorthuis, and Monique Volman. "The role of school in adolescents' identity development. a literature review". In: *Educational Psychology Review* 31.1 (2019), pp. 35–63. ISSN: 1573-336X. URL: `https://doi.org/10.1007/s10648-018-9457-3`.

[WC10] Allan Wigfield and Jenna Cambria. "Students' achievement values, goal orientations, and interest: definitions, development, and relations to achievement outcomes". In: *Developmental review* 30.1 (2010), pp. 1–35.

[Wei71] Carl Friedrich von Weizsäcker. *Die Einheit der Natur: Studien von Carl Friedrich von Weizsäcker*. Carl Hanser, 1971. ISBN: 9783446113862.

[Wen99] Etienne Wenger. *Communities of practice: learning, meaning, and identity*. Cambridge university press, 1999.

[Woh14] Claes Wohlin. "Guidelines for snowballing in systematic literature studies and a replication in software engineering". In: *Proceedings of the 18th International Conference on Evaluation and Assessment in Software Engineering*. EASE '14. London, England, United Kingdom: Association for Computing Machinery, 2014. DOI: `10.1145/2601248.2601268`.

[Won16] Billy Wong. "I'm good, but not that good: digitally-skilled young people's identity in computing". In: *Computer Science Education* 26.4 (2016), pp. 299–317. DOI: `10.1080/08993408.2017.1292604`.

[XW17] Yu Xiao and Maria Watson. "Guidance on conducting a systematic literature review". In: *Journal of Planning Education and Research* 39.1 (Aug. 2017), pp. 93–112. DOI: `10.1177/0739456x17723971`.

[ZSK08] A. Zendler, C. Spannagel, and D. Klaudt. "Process as content in computer science education: empirical determination of central processes". In: *Computer Science Education* 18.4 (Dec. 2008), pp. 231–245. DOI: `10.1080/08993400802390553`.

Abbreviations

ACM Association for Computing Machinery

CCS Computing Classification System, s. Ch. 5

CER Computing Education Research

CFA Confirmatory factor analysis, s. Sec. 4.1

CoP Community of Practice, s. Ch. 3

CS Computer Science

CSE Computer Science Education

CSER Computer Science Education Research

EFA Exploratory factor analysis, s. Sec. 4.1

IEEE Institute of Electrical and Electronics Engineers

STEM Science, Technology, Engineering, Mathematics: the science, technology and mathematics oriented school subjects

List of Figures

List of Tables