A Typology of CS Students' Preconditions for Learning

Maria Knobelsdorf
Freie Universität Berlin
Takustr. 9
D- 14195 Berlin
+48-30-838-75187
knobelsd@mi.fu-berlin.de

ABSTRACT

Problems that first year students encounter when majoring in Computer Science (CS) are complex and interrelated. We assume that CS majors drop the subject because, among other non-educational reasons, the teaching process and learning environment do not fit their preconditions for learning. Before meaningful educational interventions can be developed to address this issue, a profound understanding of students' learning backgrounds is needed. For this reason, we developed a biographical research approach, which allows us to analyze students' individual computing experiences retrospectively.

Students' computing experiences are individual and thus vary. However, students still share some common experiences, beliefs, and perceptions and a certain coherence or relationship should exist between them. Therefore, the objective of our research is to reconstruct typical patterns among the single characteristics of students' preconditions. For this purpose an empirically-based typology is planned. This paper presents our research design, providing a detailed description of how to develop an empirically-based typology.

Categories and Subject Descriptors

K.3.2 [**Computers and Education**]: Computer and Information Science Education – *Computer science education*, *Literacy, Self-assessment.*

General Terms

Experimentation, Human Factors.

Keywords

Typology, CS, Computers and Society, CS Education Research, Pedagogy, Computer Biographies.

1. INTRODUCTION

In Computer Science (CS), there is a consistently high drop-out rate, especially among female CS students. A myriad of issues contribute to this problem. For example, women may not study CS because they do not perceive themselves as computer scientists or identify with CS as a subject.

The reasons for high drop-out rates in CS are complex and interrelated and some aspects of this issue relate particularly to matters of CS Education. The latter is the focus of our research project at the Institute of Computer Science, Freie Universität Berlin. We investigate how students familiarize with CS, and we assume that this is influenced and formed by, among other factors, the interaction with CS artifacts. Our objective is to elaborate on this interaction in both formal and informal settings. For this purpose, we developed a biographical research design based on the principles of qualitative social research in which we analyze students' computing experiences retrospectively [31].

Our goal is to examine if these individual experiences comprise patterns that could be identified as types and embraced in a typology of CS students' preconditions for learning. The intent is to provide a practical and useful summary of the research field, as well as a theoretical background that could be used to develop educational interventions for teaching introductory CS classes. In this paper, we present our research approach and preliminary findings. The paper consists of four parts:

1. In section 2, we examine the drop-out problem in CS and provide a rationale for our research purpose.
2. In section 3, we present our research design. It includes the theoretical background, our research instrument, and an overview of the intended typology development.
3. In section 4, the methodology of an *empirically-based typology* is presented.
4. Finally, section 5 describes the current status of our intended typology development, summarizing studies conducted to date.

The paper concludes with section 6 where we discuss areas for future research.

2. RESEARCH MOTIVATION

Problems that first year students encounter when majoring in a subject are complex and interrelated. Before initiatives are developed to motivate more students to major in CS, it is necessary to understand why so many students who were accepted into CS programs, drop the subject so quickly. The drop-out rates in CS have always been extremely high at German universities for decades. Therefore, it is reasonable to explore the reasons for the high drop-out in CS in general.

The *Freie Universität Berlin* conducted in 2006 an empirical study throughout the university and examined the reasons why students drop a subject and leave the university [34]. Even though the results of that study are not the topic of this paper, it is interesting considering the several items of the study. The first main topic was students' *preconditions for learning*, which included the family's and the student's educational background, the reason for the subject choice, information on the subject and university to be chosen, and expectations of the subject and program. Next it surveyed students' *personal conditions* like family situation, financial situation, and health. The third main topic referred to *students' experiences* with their degree program and subject, including qualification requirements, conditions at the university, and learning experiences. Finally, the last main topic focused on the *reasons for dropping* the subject: personal and family reasons, financial situation, conditions at the university, motivation, disappointment, unmet expectations, performance and failure, and changing their profession. These items show how complex the reasons for dropping out are and how they are influenced by very different reasons like financial circumstances, how the subject is perceived, or learning environment.

As CS educators, we concentrate on matters of learning in order to contribute to the avoidance of drop-out. The preconditions for learning (including education and expectations of the subject) as well as students' experiences with the subject (including conditions at the university and learning experience) are the major points of interest. Considering the education process holistically is important, and it is crucial to how we understand learning.

From a constructivist perspective, learning is a process in which students construct knowledge and understanding individually. Students actively take part in this individual process. Learning becomes not only cognitive knowledge acquisition, but it also includes and affects all aspects of a student's personality [10]. This means that interest in and perceptions of CS do not arise suddenly; they develop gradually in a process of experience and understanding. Therefore, students enter CS class not as tabula rasa, but with some already acquired knowledge, ideas, and expectations. It is important to consider students' prior experiences and to incorporate students' everyday contexts into teaching [9], [33]. From now on, we subsume under the term *preconditions for learning* all aspects of students' educational background: every aspect of student's cognition and personality that will affect the further learning: e.g. preconceptions, pre-knowledge, beliefs, expectations, motivation, and interest.

Research aimed at understanding students' interest and involvement in CS was conducted mainly with a focus on gender. It revealed that students frequently have wrong, limited, or inadequate ideas about career opportunities in CS, as well as social environment and culture [8]. Beliefs about IT jobs and careers are highly biased and restricted to the cliché of a lonely male programmer in front of a computer-screen [27]. Students' preconceptions about CS should also be considered: Many students believe CS is primarily concerned with using and administrating computers [25]. Students who are comfortable using a computer believe to be successful in CS, as well [5], [36]. Previous research on students' knowledge when they begin a CS program confirmed their different levels of pre-knowledge [13], [16], [17], [24], [32]. Hence, it is reasonable to think that students' preconditions for learning CS have a major impact on their success in studying the subject.

We assume that CS majors drop the subject because, among other non-educational reasons, the teaching process and learning environment does not fit their preconditions for learning [3]. Before meaningful educational interventions are developed to address this issue, a profound understanding of students' learning backgrounds is needed. Hence, our research questions are:

1. What preconditions for learning do CS students have before starting university studies?
2. How do these preconditions develop and influence further learning?
3. What kind of a patterns, similarities or differences among the single characteristics of students' preconditions can we reconstruct?
4. How are these preconditions related to what is expected from students in the first year of studies?

In the next section, we present the research design that focuses on these research questions.

3. RESEARCH DESIGN

Based on our research questions, the research design considers CS students' preconditions for learning. Consistent with constructivism, we intend to examine these preconditions from the students' perspectives because we want to provide a background for teaching that allows the students' individual expectations to be met. Furthermore, we intend to examine these preconditions retrospectively because we are interested in students' perspectives on a specific moment: the beginning of their university studies. This purpose involves a biographical perspective on learning.

3.1 Biographical Research

Biographical research in education considers life as a process of learning and individuals' biographies as stories of learning. A biography (as opposed to curriculum vitae) is considered a subjective construction of reminiscent moments in life, where an individual describes particular situations and learning processes that were important for him or her. These processes refer not only to a formal setting. They also include experiences, changes, and decisions a subject went through and that established his or her self-conceptualization, world-view, and habits [11].

In his research on biography and education, Marotzki concludes that the process of creation of self-conceptualization and world-view is important for the construction of biographies: "The *perspective of individual sense- and meaning-making* leads directly to the approach of modern biographical research […] An understanding of learning and education […] becomes possible only when one comes to understand processes of learning and education as specific way of interpreting oneself and the world." ([26], p. 103). A research approach that focuses on biographical learning processes must therefore consider self-creation and world-making of individuals.

Our research design is based on this biographical approach. We are not interested in the entire biography, but in the parts that are relevant to CS Education. Therefore, we concentrate on all parts of a biography referring to learning, experiencing, and understanding CS. In particular, we are interested in every kind of interaction between one or more persons and CS artifacts. CS artifacts include both physical occurrences/values that can be referred to with the general term "information technology" as well as all non-physical occurrences/values that are referred to with the term "information science", e.g., algorithms, software, diagrams, etc. Since the students' interactions with CS artifacts comprise a broad field, our research approach focuses on the interaction with computers only. For more information, especially a detailed analysis about the role of computing experiences, see [31].

3.2 Methodology

We have developed a biographical research approach, which allows us to analyze students' individual computing[2] experiences retrospectively. Our data gathering method provides an autobiographical essay (usually hand-written) on computing experiences, which we call *a computer biography* [31]. We ask students to write down their computer biographies and encourage them to start with the first contact with a computer they can recall. We stimulate this writing process with "lure texts", which are quotes from other computer biographies. The question is intentionally open-ended to encourage the individuals to make their own decisions about which experiences were most significant. It is important to note that students are not asked about any specific aspects explicitly. The fact that certain references to different aspects occur indicates how important such experiences are to students' relationships with CS artifacts.

Computer biographies of CS majors explain why and how students chose to study CS. Such texts usually follow a typical narrative pattern and are constructed in a very coherent way. Additionally, we find important experiences that fostered or constrained the students' development. Since computing and CS are closely related (especially for novices), computer biographies reveal information about students' understanding and beliefs of CS [21].

According to the biographical perspective and constructivism, every student constructs knowledge individually and has different perceptions and beliefs about CS. Consequently, we should reconstruct the biographical learning process of each student and develop personalized interventions. However, our institution's structure and capacity make it impossible to achieve this degree of personalization. Therefore, effects of educational interventions are likely to be limited to these students, whose biographical learning processes "match" these interventions. However, interventions should reach all students.

Students' computer biographies are individual and thus vary. However, students still share some common experiences, beliefs, and perceptions. In addition, certain relationships should exist between several experiences, beliefs, and perceptions in a student's computer biography. This requires the reconstruction of some typical pathways in computer biographies and the development of a typology of students' biographical learning processes of CS.

What exactly is a typology? Typologies play a major role in conceptualizing complex social realities. A certain social reality is surveyed and empirical data is collected. A typology is the result of a data grouping process that provides a structured and reliable overview of this social reality. Data elements that correspond to one or several characteristics are merged together into one type. Types are constructed to structure and understand these characteristics with regards to their differences and similarities. This can be done with a theoretical or empirical purpose. "The construction of classes, categories, or types is a necessary aspect of the process of inquiry by means of which we reduce the complex to the simple, the unique to the general, and the occurrent to the recurrent." ([30], p. 3).

[2] The term computing refers to all kinds of computer usage and interaction.

Types and typologies can be determined by many different characteristics and for different purposes. "[Typologies] can be used for classificatory or descriptive purposes, as heuristic devices and as methodological conveniences." ([30], p. 8). Therefore, the objective of a typology is two-fold. The first purpose is descriptive and helps to structure the collected data in order to make it manageable and to provide an overview. It is convenient and useful when the social reality is extensive and of a complexity that can be reduced with a typology. The second purpose is heuristic and has a theory-building function: It is assumed that the correlation between the elements of a type is not incidental. It is reasonable that a certain relationship exists between the elements of a type . The output is of hypothetical quality and serves as a background for theory building (Kluge, pp. 43). "This capability is built into [types], since as composites they are given a structure with functional consequences, and hence types are systems." ([30], p.8)

The results of our research should reduce the multitude of elements in our computer biographies to a few groups and therefore provide an arrangement (primarily descriptive) and structuring of our research field (CS students' preconditions for learning). This will produce manageable results that can easily be used in CS class for diagnostic reasons. Our results should also serve heuristic purposes. Because our results will form a certain relationship between the elements of our research field, it will provide a theoretical background for proposing hypotheses and theory-construction in the field.

The next section presents a detailed methodology description of how to develop a typology. We rely on this methodological background in section 5, where we present the results of our previously conducted studies to serve as a background for the indented typology.

4. AN EMPIRICALLY-BASED TYPOLOGY

This section summarizes qualitative social research about the development of an empirically-based typology. Kluge[3] reviewed the main core of social literature and research about typology theories and methodology. She gives an account of this review in [18], referring, among others, to [2], [6], [7], [12], [14], [22], [23], [37]. The author was engaged in theoretical work as well as empirical research (*Sonderforschungsbereich "Statuspassagen und Risikolagen im Lebensverlauf"* of the University Bremen[4]), where she contributed as a qualitative social researcher in the domain of methodology. Drawing from her experiences, she proposes a normative model that summarizes the essential aspects of different typification[5] approaches by [12], [22], and [23]. These authors focus on different data gathering or analysis methods and generate different sorts of types. Their different proceedings can be summarized in a general model that can be adapted. Since each of these proceedings contains methods that are useful for our approach, we intent to implement Kluge's model and adapt methodology proposed by her in each stage.

[3] In this paper we refer to Kluge's work in [18]. Because the book is in German, content only is reflected. Kluge provides a brief English summary of her work in [19].

[4] Sfb 186, funded since 1988 by the DFG (German National Research Foundation).

[5] The word typification means the process of developing a typology.

4.1 Types and Typology

A type consists of a set of characteristics that are interrelated and logically connected in regards to content. Each characteristic has different parameter-values and can be understood as a dimension of comparison. Each case[6] is classified according to its parameter-values, and then the groups are compared. A type can be understood as some multidimensional space of parameter-values and is coined as a property-space. Barton & Lazarsfeld developed and described the theoretical background of property-spaces as well as multidimensional tables that represent property-spaces [4]. Table 1 is an example of a two-dimensional property-space containing characteristics A and B that are defined by parameter-values: A1, A2, B1, and B2. The numbers in the table show the arrangement of the cases in accordance to the characteristics' values (e.g., the number 10 indicates that there are 10 cases that express A1 and B1).

Table 1. An example of a two-dimensional table.

Characteristic A	**Characteristic B**	
	value B1	value B2
value A1	**10**	**3**
value A2	**7**	**1**

Types are developed from the grouping process. Therefore, each type should be homogenous inside (internal homogeneity) in order to form common characteristics. Among themselves, types should be highly heterogeneous (external heterogeneity) in order to broaden diversity of the research field. However, different types can form a typology only when they refer to the same property-space ([18], p. 42).

Typologies play a major role in conceptualizing complex social realities since classifications used in sciences are not appropriate for this purpose. There is a difference between a typology and a classification. A classification must be mutually exclusive and exhaustive. A type, on the other hand, combines characteristics that are not uniquely and exclusively allocated to it. There is no clear separation between types. Therefore, it is important to remember that a typology cannot reproduce the reality. Types are based on predefined characteristics and represent only a part of reality. Hence, generalizations must be handled cautiously ([18], p. 25).

"[T]he research practice is confronted with the problem how these types can be constructed systematically and transparently. In current sociological literature, there exist only few approaches in which the process of type construction is explicated and systematized in a detailed way. […] Also different concepts of type are used (e.g. ideal types, empirical types, structure types, prototypes etc.) or the concept of type is not defined explicitly at all." [19]. Therefore, Kluge proposes a four-stage model of an empirically-based typification ([18], pp. 260), which is presented in the next section.

[6] The term *case* means a data item, unit, or entity which can be a complete interview or a part of it, e.g. a certain decision every interviewed person is talking about. In our research approach a case is a complete computer biography.

4.2 A Four-stage Model of an Empirically-based Typification

The model generalized by Kluge consists of four main stages, where the first three stages can be repeated (see Figure 1). These stages are:

1. Developing the relevant dimensions of comparison
2. Case grouping and empirical regularities analysis
3. Analysis of coherence and typification
4. Types characterization

The four stages will be described in more detail in the next four subsections.

4.2.1 Developing the Relevant Dimensions of Comparison

The first stage forms characteristics and establishes dimensions of comparison. It is important to note that each case consists of all defined characteristics. Otherwise, cases cannot be compared with each other. A typology makes sense only when all types are related to each other. Thus, this stage is very important. Only the established dimensions of comparison form the basis of typology.

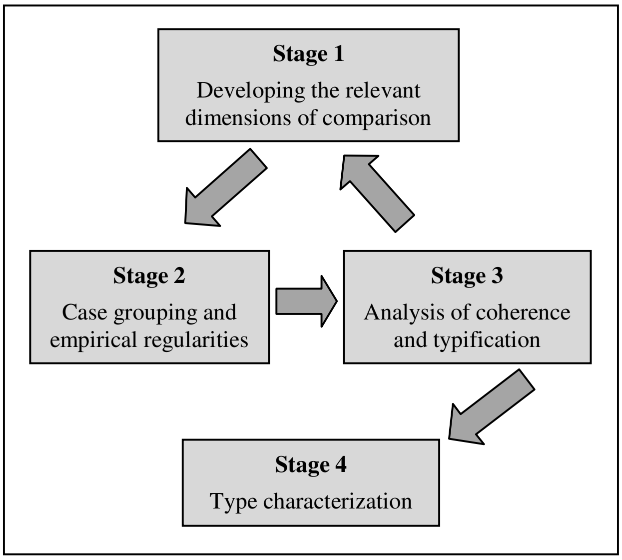

Figure 1. Model of the empirically-based typification ([18], p. 261).

In order to substantiate the dimensions of comparison and to form further characteristics, collected data is analyzed intensively: each case is evaluated separately and then compared to all the others. Thematic coding by Glaser, Strauss, and Corbin is frequently used. First, the data is coded with thematic keywords and then, based on the keywords, the cases are compared to each other. This way, both a case study and comparison can be combined together very effectively. Similarities and differences between the cases can be elaborated on ([18], p. 266-269).

4.2.2 Case Grouping and Empirical Regularities Analysis

After establishing dimensions of comparison and their parameter-values, all cases can be grouped. Basically, there are two ways to proceed at this point. In a bottom-up process, the two most similar cases (i.e., two cases which have the same or similar parameter value for one characteristic) are merged iteratively together into a group or cluster (agglomerative process). In a top-down process, all cases are treated as one group that is divided into sub-groups with the same or similar parameter value according to one characteristic (divisive process) ([18], p. 270).

The agglomerative process is very time consuming because all cases must be compared to each other during each step. Hence, this process is conducted with computers, and agglomerative algorithms that perform cluster analysis are used. The disadvantage of this process is that it is difficult to trace which characteristics form the cluster, and one or two irrelevant characteristics can significantly distort the result. Combinations of characteristics that do not appear in the data are not incorporated. Only digressive cases, which could not be allocated to any cluster, can be found with adequate merging algorithms ([18], pp. 275).

Multidimensional tables that represent the dimensions of comparison are helpful to illustrate the grouping process [23]. Table 2 shows an example of a two-dimensional property-space. Multidimensional tables provide "a general view over all possible combinations which are *theoretically* conceivable. Since all possible combinations often do not exist in reality and/or the differences between individual combinations of attributes are not relevant for the research question, single fields of the attribute space can be summarized." [19].

4.2.3 Analysis of Coherence and Typification

Under the presumption that characteristics do not correlate randomly, an interrelation and logical connection in regards to content between the grouped characteristics must exist. The groups or clusters that were found in stage 2 become types when this coherence and connection can be identified. This process is based on the preliminary features of each group and on further characteristics concerning similarities and differences between the cases and the groups. There is no methodological advice on how to proceed at this point. As Kluge writes, the most difficult step is to systemize the analysis of sense coherence and logical connection of the grouped characteristics ([18], p. 279).

4.2.4 Types Characterization

The typification finishes with characterizing the types as comprehensively and as precisely as possible in regards to the relevant characteristics, their combinations, and their coherence. Because the cases of one type are not entirely equal in each characteristic, the problem lies in how to picture the similarities. Different forms of types exist for this purpose: prototypes are real cases that represent the type best; ideal types present the essential characteristics in their pure form; and if only opposite types exist, extreme types are useful.

If only extreme or ideal types are used, the risk of losing diversity and the appearance of inconsistency of the investigated reality arises, since the focus lies on the pure or extreme aspects. Abbreviations of types must also be used carefully because, again, this can cause a distortion of the results ([18], p. 280).

5. A TYPOLOGY OF CS STUDENTS' PRECONDITIONS FOR LEARNING

In this section we summarize the results of our previous studies that constitute the preliminary dimensions of comparison for the typification.

5.1 The Four Dimensions of Comparison

We analyzed the computer biographies from four different perspectives: *sense*, *structure*, *habits* and *pathway*. We will summarize this approach very briefly. For further reading, see [20], [21], [31].

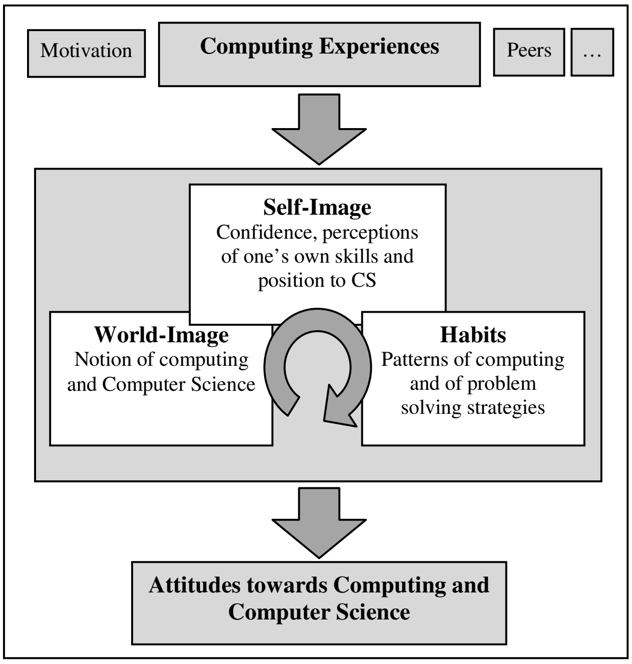

Figure 2. The analytical dimensions self-image, world-image, and habits specifying the biographical computing process ([31], p. 32).

Based on the biographical research approach discussed in section 3.1, we used a coding paradigm suggested by Tiefel [35] in her work on adapting Grounded Theory for the analysis of biographical learning processes: "a modified coding paradigm is proposed, with analytical perspectives geared to the reconstruction of subjective processes of making sense and constructing coherence." ([35], p. 66). However, according to constructivism, learning takes place through interaction. Therefore, Tiefel suggests considering interweaving the following perspectives into the analysis process and reconstructing the biographical learning process through them:

- The self creation of human beings and their subjective processes of sense-making are summed up under the notion of sense perspective and the self-image is then reconstructed from it. In our case, the self-image includes self-judgment and attitudes of one's own computer skills and orientation in the computer world.
- The relationship of human beings to the world and their coherence-creation are summed up under the notion of structure perspective and then the world-image is reconstructed from it. In our case, *world-image* includes personal theories and preconceptions about computing and CS.
- Finally, human beings' attitudes and contexts of interaction, reaction, and strategies of action are summed up under the notion of the habits perspective and then habits or behavior are reconstructed from it. In our case, *habits* include learning

strategies, typical performances with the computer, and reactions to problems.

These three perspectives form an analytical point of view on the holistic biographical learning process ([31], p. 31). Based on Tiefel's coding paradigm, Figure 2 illustrates an analytical approach: computing events are experienced individually and influenced by internal and external factors. These experiences are part of the biographical learning process of CS and therefore affect students' world-image, self-image, and habits related to CS. Different experiences in students' lives are interrelated. With each new experience, these three dimensions, separated only on an analytical level, are affected ([31], p. 32).

While analyzing the biographies, we realized that the biographical process is a further analytical perspective on the three dimensions because through the biographical process the world-image, self-image, and habits develop, change, and interact. The biographies of CS majors who had just entered the university revealed three periods. We call the first period the *introductory period*. It starts with the first contact with a computer. It contains experiences and situations that are initiated either by coincidence or by others. After the introductory period, a *period of development* begins. It is characterized by meaningful experiences in which students develop their interests. Then a *decision period* might take place. It contains important experiences that are crucial for the future. These experiences are described in more detail than other events in a biography [21]. Additionally, we analyzed one biography of a PhD student who graduated in CS several years ago where we examined a period that follows the decision period. Therefore, we assume that the process likely continues after the decision period.

As a result, the four dimensions (world-image, self-image, habits, and process) establish the dimensions of comparisons of our typology. These four dimensions provide many different grouping combinations, which also depend on how many dimensions each attribute has. The *process* dimension, for instance, can have three attributes: Introductory Period, Development Period, and Decision Period. Table 2 shows an example of a possible four-dimensional table of the corresponding property-space. W1, W2, S1, S2, H1, and H2 are not further specified parameter-values of the dimensions world-image, self-image, and habits and are shown just for illustrative reasons.

Table 2. Example of a possible four-dimensional table for the property-space of computer biographies.

		Process		
		Introductory Period	Development Period	Decision Period
World-image	W1			
	W2			
Self-image	S1			
	S2			
Habits	H1			
	H2			

The results of our studies (described in the next subsection) can serve as preliminary parameter-values of these four characteristics and form possible dimensions of comparison.

5.2 Parameter-Values

In our previous studies, we surveyed biographies of students majoring in different subjects: CS, Bioinformatics, Mathematics, Psychology, CS Education, and German Philology. The comparison between CS-affiliated and non-affiliated students helped in contrasting and understanding the biographies of CS majors. At the beginning, we analyzed the biographies using the Grounded Theory approach and open and axial coding ([1], pp. 271). In the last two studies, we used qualitative content analysis by Mayring [29]. We collected a large number of parameter-values of world-image, self-image, habits, and process. These characteristics are summarized below.

5.2.1 Psychology and German Philology Students

We have found the following attributes among students majoring in Psychology and German Philology: CS is perceived as a closed world that a person can only enter with special skills (a "clubhouse"). CS is an incomprehensible and complicated subject. The computer is perceived as a CS artifact and also as a tool used for working. Students' only interrelation with CS happens using the CS artifact computer.

Relating to self-image, the students believe that computer scientists are using the computer in a different way (more professional) and are able to understand "the mystery" behind it. The students believe that they are not capable of learning computer-based skills because they are missing a certain "pre-understanding" and "skills" (a special gene) that computer scientists have naturally. Therefore, these students see themselves as outsiders of the CS world.

These students are mainly autonomous learners, and they often feel helpless and left alone with problems they cannot solve and understand (learnt helplessness, attribution theory). They prefer to be taught how to use the computer and this is what they expect from a CS class at school. When using a computer, they want to understand how something works before they try to perform it themselves.

5.2.2 CS Students

We have found the following attributes among the CS majors: CS is perceived as a closed world a person can only enter with special skills, and these students think they have these skills. Based on these beliefs, the students see themselves as insiders, and the computer is omnipresent for them. As for their self-conception, they see themselves as "born to be computer scientists". They are interested in computers because they are fascinating, and computer activities are fun. They view computer problems as a challenge. They are mainly autonomous learners (learning by doing) and enjoy it. Consequently, these students often overestimate their skills and do not respond to formal learning environments.

Among the CS majors, we also found the following attributes: CS is perceived as a closed and interesting world a person can enter by changing his or her status from a user to a designer. They think that they are capable of learning things connected with a computer, and they are interested in computers because they can produce something on their own. They are mainly autonomous learners (learning by doing) and enjoy this situation, too. In contrast to the characteristics in the paragraph above, these students do not perceive themselves as being born with these skills; they accept that such skills are developed. Therefore, they are more willing to accept learning in formal settings.

5.2.3 *Bioinformatics Students*

Among the Bioinformatics majors, we found the following characteristics: CS is perceived as a fun and creative world, where a person can always discover and learn new things. A computer is a tool for creating. Concerning their self-conception, they think that they are capable of learning things connected with a computer, and computer activities are fun. These students are mainly autonomous learners (learning by doing) who enjoy trying things out in a playful way. In comparison to the CS students, these subjects did not identify with the computer, just as psychology students did not. But in contrast to psychology students, they were not afraid or did not feel intimidated by the computer.

5.2.4 *Summary*

Table 3, Table 4, Table 5, and Table 6 summarize the aforementioned characteristics according to the four dimensions: world-image, self-image, habits, and periods.

Table 3. Attributes of the world-image dimension

Attributes of the world-image dimension		
Clubhouse	**W1** CS is a closed world	**W1.1** only a person with special skills can enter
		W1.2 a person can enter by changing their status from a user to a designer
Nature CS	**W2** CS is a world	**W2.1** where a person can always discover and learn new things
		W2.2 that is fun and creative
		W2.3 that is interesting
		W2.4 incomprehensible
Nature Artifact	**W3** The computer is	**W3.1** a toy
		W3.3 a tool (to work with: a pragmatic view)
		W3.4 a tool (for creating: a creative view)

Table 4. Attributes of the self-image dimension.

Attributes of the self-image dimension		
Self-conception	**S1** Concerning the “CS world”	**S1.1** I am an insider.
		S1.2 I am an outsider.
	S2 Concerning myself	**S2.1** I was born to become a computer scientist.
		S2.2 I became a computer scientist.
		S2.3 I know that one can become a computer scientist, but this process is not completed for me yet.
		S3.2 I know that one can become a computer scientist, but I will never be one.
Learning	**S3**	**S3.1** I am able to learn things on the computer.
		S3.2 I am not capable of learning things at the computer.
Sensation	**S4** Computer activities are	**S4.1** fun
		S4.2 dull
Interest	**S5** I am interested in computers because	**S5.1** they are fascinating.
		S5.2 I can produce something on my own.
		S5.3 they are useful and helpful.
Motivation	**S6** At the computer, I’m motivated most when	**S6.1** I can do some context-based things.
		S6.2 I can perform, try different roles.
		S6.3 the activities include creativity.
		S6.4l I can work independently and be self-determined.

Table 5. Attributes of the habits dimension

Attributes of the habits dimension		
Reactions	**H1** To computer problems	**H1.1** I feel helpless.
		H1.2 I appreciate the challenge.
Learning behavior	**H2** Things I can do on the computer	**H2.1** I am a self-learner (learning by doing).
		H2.2 I was taught.
Behavior	**H3** When I do something on the computer	**H3.1** I simply try things out.
		H3.2 I try to understand things before I do them.

Table 6. Attributes of the Process dimension.

Attributes of the Process dimension		
Transition	**B1** A transition	**B1.1** has been experienced from use to design
		B1.2 has not been experienced
	B2 A development	**B2.1** has been experienced from a regular use to a professional use
		B2.2 has not been experienced
Period	**P1** Introductory Period	
	P2 Development Period	
	P3 Decision Period	

Currently, we are working on further characteristics. We examine stereotypes in CS: how students reproduce them and what kind of influence they have for successful learning [15]. We also plan a study about mindsets based on the self-theories by [10].

5.3 Further Proceedings

In this subsection, we outline how we plan to continue our research project and the intended typology. We describe the data collection, analysis, and typology stages, and we provide a timeline for these activities.

5.3.1 Data Collection

The dimensions seem to be constant. Each new aspect is a further attribute to one of the dimensions. Since all the examined attributes have been elaborated on in different studies, we were not able to compare all cases to all characteristics. These attributes form a certain dimensions of comparison but are preliminary for the development of a typology. In order to construct types, the data must be based on all attributes (see section 4.2.1). Therefore, it is necessary to survey new data that will refer to a certain dimension of comparison. It is also necessary to survey new data that will provide new attributes or further information on the existing one. For this purpose, we collected at the beginning of the winter-semester 2008 new computer biographies of first year CS students at our institute. In a second data collection step, we are planning to conduct semi-structured interviews with a subset of the same students in order to gain additional information. The intended typology will be based on this data.

5.3.2 Data Analysis

In the process of data analysis that corresponds to stage one and two of the empirically-based typology (see sections 4.2.1 and 4.2.2), we will use the qualitative content analysis by Mayring [29].

Qualitative content analysis by Mayring "[…] is defined as […] an approach of empirical, methodological controlled analysis of texts within their context of communication, following content analytical rules and step model, without rash quantification." [28]. Within this model, a category system is developed and several approaches are possible. As Mayring suggests, "[t]he main idea of the procedure is to formulate a criterion of definition, derived from theoretical background and research question, which determines the aspects of the textual material taken into account. Following this criterion the material is worked through and categories are tentative and step by step deduced. Within a feedback loop those categories are revised, eventually reduced to main categories and checked in respect to their reliability." [28].

Coding methods in Grounded Theory are not restricted, which is an advantage when a research question is open and very little is known about the research field. The disadvantage is that many steps are not standardized, nor well-defined. Therefore, a lot of expertise and capacity is necessary for decision-making and analysis. Since we have already conducted our study, we gained some knowledge and understanding of our research field. Using the typology, we aim to specify and structure our results. Therefore, we need a standardized and well-defined method to analyze our data effectively, and qualitative content analysis fits these criteria.

5.3.3 Research Schedule

Data collection is conducted in the winter-semester 2008, followed by data analysis and selection of students for interviews. The objective of the semi-structured interviews with the CS majors is to get more information on the single attributes. Next, we will collect computer biographies of the non-CS majors, analyzing and comparing them to the data of the CS majors in order to obtain a high contrast level. This data will be used for the grouping process and construction of the subsequent stages of typology. Table 7 provides an overview of the future activities.

Table 7. Overview of future activities.

Activity	Purpose
Collect new computer biographies of first year CS students (on their first day at the university) Collect computer biographies of non-CS majors Type biographies	To survey the current data To survey the contrast data To prepare for data analysis
Analyze collected data Choose students for interviews Develop semi-structured interviews	To gain some new characteristics To choose interviewees
Conduct interviews with the same first year CS students two months after they started their studies	To gain more information on characteristics
Analyze interviews (stage 1) Grouping process (stage 2) Analyze coherence and typification (stage 3) Types characterization (stage 4)	To construct a typology
Conduct interviews with the same interviewees, one year later,	To gain information about students' further learning process at the university
Analyze interviews	To examine CS program influences on further learning

In section 2, we stated four research questions we intend to answer with this research project:

1. What preconditions for learning do CS students have before starting university studies?
2. How do these preconditions develop and influence further learning?
3. What kind of a patterns, similarities or differences among the single characteristics of students' preconditions can we reconstruct?
4. How are these preconditions related to what is expected from students in the first year of studies?

Using the typology, we will answer questions 1-3 and make certain predictions about students' development in their university studies. In order to answer questions 4 we will analyze what is expected from CS students in the first year of studies and to compare this with the typology result. At this early stage, we have not developed a methodological approach for this objective, yet. Finally, we intend to conduct semi-structured interviews with the same CS students one year later and question them about their CS studies. The interview structure will be developed based on the typology.

It would be appropriate to test the typology using quantitative methods like a standardized questionnaire, but this would be an additional research project.

6. CONCLUSION

In this paper, a research design that combines different theoretical and methodological approaches from sociology, psychology, education, and CS was presented. We outlined how we adapt theory and methods to answer a research question from CS Ed (Education). How is this research design supposed to be evaluated? Certainly, it would be possible to evaluate each aspect independently. However, empirical methods from social sciences, whether qualitative or quantitative, are not "recipes" for data survey and analysis. The main challenge is that using a given methodology properly involves adaption of its epistemological and ontological context as well.

CS Ed research often approaches using methodology in an algorithmic way. However, we need empirical evidence to provide valid and sustainable results. For instance, when Grounded Theory is used to develop a theory about *student understanding of programming concepts*, the researchers have to think and behave in the tradition of qualitative empirical research, similar to sociologists. However, by doing this, they will depart from CS Ed research. This raises the question: how do we judge research design like the one presented in this paper? Overall, is such a research design still CS Ed research? Where do social sciences end and where does CS Ed starts?

The research design described in this paper intends to examine CS students' learning backgrounds retrospectively. The intention is to analyze the preconditions for learning that CS students have and how these preconditions develop and influence further learning. Finally, the objective is to reconstruct patterns, similarities or differences among the single characteristics of students' preconditions. For this purpose an empirically-based typology is planned. It must be discussed if this research approach is appropriate for the research purpose.

As for methodology, different stages are planned. It must be discussed if the data collection and analysis is appropriate. A clear question is whether the property-space is complete. The qualitative social research talks about data saturation, but remarks that the researchers must decide themselves when the property-space is complete. Finally, the research project presented in this paper intends to examine how CS students' preconditions are related to what is expected from students in the first year of studies. Based on the knowledge of these preconditions, the overall goal is to better understand why CS students drop the subject due to learning reasons. Therefore, this paper concludes by asking if the intended research design is suitable to this purpose.

7. ACKNOWLEDGMENT

I would like to thank Carsten Schulte, Essi Lahtinen, Josh Tenenberg, and the reviewers for their helpful comments that improved this paper so much.

8. REFERENCES

[1] Andreas Böhm 2004 Theoretical Coding: Text Analysis in Grounded Theory. In A Companion to Qualitative Research (Flick, Uwe, Kardorff, Ernst von and Steinke, Ines, eds.). Sage Publications Ltd, 270–275

[2] Bailey, K. D. 1973 Monothetic and Polythetic Typologies and their Relation to Conceptualization, Measurement and Scaling. American Sociological Review 38, 18–33

[3] Barker, L. J., Garvin-Doxas, K. and Jackson, M. 2002 Defensive climate in the computer science classroom. In Proceedings of the 33rd SIGCSE technical symposium on Computer science education, 43-47

[4] Barton, A. H. 1955 The Concept of Property-Space in Social Sciences. In The Language of Social Sciences (Lazarsfeld, Paul F. and Rosenberg, Morris, eds.). Free Press, 40–53

[5] Beaubouef, T. 2003 Why Computer Science Students Need Language. ACM SIGCSE Bulletin 35, 4, 51–54

[6] Becker, H. S. 1968 Through Values to Social Interpretation. Essays on Social Contexts, Actions, Types, and Prospects, Greenwood Press

[7] Capecchi, V. 1968 On the Definition of Typology and Classification in Sociology. Quality and Quantity 2, 1-2, 9–30

[8] Cohoon, J. M. and Aspray, W. 2006 Women and Information Technology. Research on Underrepresentation, MIT Press

[9] Donovan, M. S. and Bransford, J. D. 2005 How students learn. History, mathematics, and science in the classroom, National Academies Press

[10] Dweck, C. 2000 Self-theories: their role in motivation, personality, and development, Psychology Press

[11] Ecarius, J. 2006 Biographieforschung und Lemen/ Biographical Research and Education[7]. In Handbuch erziehungswissenschaftliche Biographieforschung (Krüger, Heinz-Hermann and Marotzki Winfried, ed.). VS Verlag, 91–108

[12] Gerhardt, U. 1984 Typenkonstruktion bei Patientenkarrieren/Typification and Patient-careers[7]. In Biographie und soziale Wirklichkeit (Kohli, Martin and Robert, Günter, eds.). Metzlersche Verlagsbuchhandlung, 53–77

[13] Greening, T. 1998 Computer Science: Through the Eyes of Potential Students. In Proceedings of the 3rd Australasian conference on Computer Science Education, ACE 1998, 145-154

[14] Hempel, C. G. and Oppenheim, P. 1936 Der Typusbegriff im Lichte der neuen Logik, A.W. Sijt Hoff's Uitgeversmaatschappij N.V.

[15] Hewner, M. and Knobelsdorf, M. 2008 Understanding Computing Stereotypes with Self-Categorization Theory. In Proceedings of the 8th Baltic Sea Conference on Computing Education Research, Koli 2008, Finland

[7] This book/text is only available in German. The German title is translated by the author of this article.

[16] Hoffman, M. E. and Vance, D. R. 2005 Computer literacy: what students know and from whom they learned it. ACM SIGCSE Bulletin 37, 1, 356-360

[17] Kinnunen, P. and Malmi, L. 2008 CS minors in a CS1 course. In Proceeding of the fourth International Computing Education Research Workshop, ICER 2008, 79-90

[18] Kluge, S. 1999 Empirisch begründete Typenbildung/ Empirically Grounded Typification[7]. Zur Konstruktion von Typen und Typologien in der qualitativen Sozialforschung, Leske + Budrich, Opladen

[19] Kluge, S. 2000 Empirically grounded construction of types and typologies in qualitative social research. Forum: Qualitative Social Research [Online Journal] 1, 1 (2000), http://www.qualitative-research.net/fqs-texte/1-00/ 1-00kluge-e.htm

[20] Knobelsdorf, M. and Schulte, C. 2007 Das informatische Weltbild von Studierenden/Students' CS world-view[7]. In Didaktik der Informatik in Theorie und Praxis. 12. GI-Fachtagung Informatik und Schule - INFOS 2007, 69–79

[21] Knobelsdorf, M. and Schulte, C. 2008 Computer Science in Context - Pathways to Computer Science. 7th Baltic Sea Conference on Computing Education Research, Koli 2007. In Conferences in Research and Practice in Information Technology (Australian Computer Society, Inc., ed.), Sydney, Australia

[22] Kuckartz, U. 1990 Computerunterstützte Suche nach Typologien in qualitativen Interviews. In Fortschritte der Statistik-Software 2. SOFTSTAT '89. 5. Konferenz über die wissenschaftliche Anwendung von Statistik-Software. (Faulbaum, Frank, Haux, Reinhold and Jöckel, Karl-Heinz, eds.). Gustav Fischer, New York, 495–502

[23] Lazarsfeld, P. F. and Barton, A. H. 1951 Qualitative Measurement in the Social Sciences. In The Policy Sciences (Lerner, Daniel and Lasswell, Harold, eds.). Stanford University Press, 155–192

[24] Lewandowski, G., Bouvier, D. J., McCartney, R., Sanders, K. and Simon, B. 2007 Commonsense computing (episode 3): concurrency and concert tickets. In Proceedings of the third International Computing Education Research Workshop, ICER 2007, 133-144

[25] Maaß, S. a. W. H. 2006 Programmieren, Mathe und ein bisschen Hardware…Wen lockt dies Bild der Informatik?/ Programming, maths, and a bit of hardware…who is attracted by this picture of CS?[7] Informatik Spektrum 29, 1, 125–132

[26] Marotzki, W. 2004 Qualitative Biographical Research. In A companion to qualitative research (Flick, Uwe, Kardorff, Ernst von and Steinke, Ines, eds.). SAGE, 101–107

[27] Martin, C. D. 2004 Draw a computer scientist. In Proceedings of the 9th Annual Conference on Innovation and Technology in Computer Science Education. ITiCSE 2004, 11–12

[28] Mayring, P. 2000 Qualitative Content Analysis. Forum: Qualitative Social Research [Online Journal] 1, 2 (2000), Art. 20. http://nbn-resolving.de/urn:nbn:de:0114-fqs0002204

[29] Mayring, P. 2004 Qualitative Content Analysis. In A Companion to Qualitative Research (Flick, U. and Kardoff E. and Steinke I. von, eds.). Sage Publications Ltd, 266–269

[30] McKinney, J. C. 1969 Typification, Typologies, and Social Theory. Social Forces 48, 1, 1–12

[31] Schulte, C. and Knobelsdorf, M. 2007 Attitudes towards Computer Science - Computing Experiences as a Starting Point and Barrier to Computer Science. In Proceedings of the third International Computing Education Research Workshop. ICER 200, 27–38

[32] Simon, S., Fincher, S., Robins, A., Baker, B., Box, I., Cutts, Q., Raadt, M. de, Haden, P., Hamer, J., Hamilton, M., Lister, R., Petre, M., Sutton, K., Tolhurst, D. and Tutty, J. 2006 Predictors of success in a first programming course. In Proceedings of the 8th Austalian conference on Computing education, ACE '06, 189-196

[33] Sinatra, G. M. 2005 The "Warming Trend" in Conceptual Change Research: The Legacy of Paul R. Pintrich. Educational Psychologist 40, 2, 107–115

[34] Thiel, F., Blüthmann, I., Lepa, S. and Ficzko, M. 2007 Ergebnisse der Befragung der Studierenden in den Bachelorstudiengängen an der Freien Universität Berlin Sommersemester 2006. http://www.ewi-psy.fu-berlin.de/einrichtungen/arbeitsbereiche/schulentwicklungsforschung/forschung/bachelorbefragung.html

[35] Tiefel, S. 2005 Coding in terms of Grounded Theory. Modifying coding guidelines for the analysis of biographical learning within a theoretical framework of learning and education. ZBBS 6, 1, 65–84

[36] Turner, E. H. and Turner, R. M. 2005 Teaching entering students to think like computer scientists. In Proceedings of the 36th SIGCSE technical symposium on Computer science education, 307–311.

[37] Weber, M. 1988/1904 Die "Objektivität" sozialwissenschaftlicher und sozialpolitischer Erkenntnis/ The Objectivity of the Sociological and Social-Political Knowledge. In Gesammelte Aufsätze zur Wissenschaftslehre (Winckelmann, J., ed.). Mohr, Tübingen, 146–214