Applying Sequence Analysis in Higher Education Research: A Life Course Perspective on Study Trajectories

Christina Haas[a]

[a]*Leibniz Institute for Educational Trajectories, Bamberg, Germany*

christina.haas@lifbi.de

Postprint Notice

This is the author accepted manuscript of:

Haas, C. (2022). Applying Sequence Analysis in Higher Education Research: A Life Course Perspective on Study Trajectories. In J. Huisman & M. Tight (Eds), *Theory and Method in Higher Education Research* (Vol. 8, pp. 127–147). Emerald Publishing Limited.

The final version of record is available at: https://doi.org/10.1108/S2056-375220220000008007

Applying Sequence Analysis in Higher Education Research: A Life Course Perspective on Study Trajectories

This contribution introduces sequence analysis to higher education research, an explorative technique aiming at detecting patterns, regularities and resemblance in time-ordered data. Thereby, it enables a holistic perspective on over-time developments and processes such as educational pathways or academic careers. In this contribution, the foundations and general logic of sequence analysis will be described. As an example, referring to the life course as a framing paradigm, sequence analysis is applied to reconstruct the study trajectories of a cohort of bachelor students in Germany. The potential of sequence analysis in three specific higher education research areas is outlined, that is, to study post-secondary education trajectories, academic careers and the development trajectories of higher education organizations. The conclusion discusses advantages and disadvantages, challenges and practicalities.

Keywords: Sequence analysis; study paths; educational trajectories; careers; longitudinal data; life course; study outcomes

Introduction

Higher education expansion is one of the fundamental societal developments of recent decades (Marginson, 2016; Schofer, Ramirez, & Meyer, 2021). Worldwide, the number of students has doubled in the last 15 years (UNESCO, 2022). At the same time, the enrolment of non-traditional students keeps rising, including part-time students or adult learners, accompanied by a diversifying study offer.

Consequently, gaining a better understanding of students' post-secondary pathways is also becoming a theme of crucial policy importance. So far, post-secondary study pathways have been studied mostly from a transition-oriented perspective. That is, educational attainment is often conceptualized as a sequence of educational episodes linked by transitions, often using statistical methods such as logistic regressions or event history analysis (Lacy, 2015; Mare, 1980). These research designs are particularly valuable to predict study outcomes (dropout, graduation) in relation to study conditions, sociodemographic characteristics or vis-a-vis a range of alternative outcomes, such as switching the study programme. In this vein, research has provided important insights on enrolment, field of study, dropout, study progress, transition into graduate or doctoral studies, study completion, achievement, and study success (Lassibille & Navarro Gomez, 2009; Lörz & Mühleck, 2019; Meggiolaro, Giraldo, & Clerici, 2017; Pfeffer & Goldrick-Rab, 2011).

Yet, as a conceptual disadvantage, the transitions-in-education perspective is bounded to a small number of alternative events that could be captured within a study trajectory. Consequently, study trajectories after an initial dropout or after switching majors are hardly considered in this perspective (Milesi, 2010). This undermines that the higher education attainment process is often more complicated given that students' actual trajectories frequently differ from some institutionally pre-defined ideal pathway through a study programme. For this reason, an alternative approach is presented in this contribution. This approach rests on the life course as a framing paradigm, emphasizing the actual course of attainment as a long-term process.

Methodologically, this contribution introduces sequence analysis to higher education research, an explorative technique aiming at detecting patterns, regularities and resemblance in time-ordered data. Although not yet widely applied in the area of (higher) education, sequence analysis enables a holistic understanding of educational pathways. This holds particularly true for trajectories in higher education, given that higher education is not compulsory and – in most systems – is less standardized compared to earlier educational stages. This makes sequence analysis a valuable method to reconstruct study trajectories to answer questions such as 'What are common trajectory patterns in a specific higher education context or at a specific university?'

Higher education and the life course

The Relationship Between Higher Education and the Life Course

Higher education and the life course are intertwined in several ways. First of all, higher education is a part of the educational trajectory and thus constitutes a specific life course passage. Second, on an individual level, higher education attainment affects greatly a range of other life course outcomes such as the future occupational career, unemployment risks, family formation or health (Hout, 2012; Vila, 2000). Third, societal developments and global changes also influence developments in higher education. For example, the increase in women's labour market participation goes hand in hand with the rise of female participation in higher education (Buchmann & DiPrete, 2006; Goldin, Katz, & Kuziemko, 2006). Trends towards lifelong learning and upskilling throughout the employment career increase the participation of non-traditional students in higher education, who enter with very different needs and objectives than more traditional student groups (Deil-Amen, 2011; Schuetze & Slowey, 2002; Tieben, 2020). Related to this is also the increasing demand for flexible study modalities such as remote or part-time programmes to allow for the combination of (full-time) work and higher education – a demand that is particularly accommodated by private organizations.

Higher Education as a Life Course Passage: Between Institutional Embeddedness and Individual Agency

Higher education is a specific passage of the educational life course, which is broadly defined as a 'sequence of age-graded roles and contexts with associated developmental tasks' (Crosnoe & Benner, 2016, p. 180). With enrolment, individuals become members of a specific higher education organization, i.e. they become university or college students. As such, they take courses in a pre-defined time frame and organizational setting – e.g. on campus or remotely; in a lecture or seminar format – while their study progress is assessed and certified with a degree.

Perceiving higher education studies from a sociological life course lens implies that attainment does not take place in isolation, but that it is structured by a pre-defined logic and shaped by various factors (Pallas, 2003; Sackmann & Wingens, 2003). These are institutional settings and configurations on the macro and meso level which 'structure social interactions and behavior via integrated systems of rules' (Gross & Hadjar, 2021, p. 382), leading to variation in 'college environmental differences and opportunities within the formal structures of institutions (their characteristics and programs) and social normative environments' (Hurtado, 2007, p. 94). The more obvious settings are legal regulations on the national, federal or organizational level that provide and regulate the study offer. Access standards may be defined on a national level, while higher education organizations define specific graduation requirements such as the number and type of course credits. Universities may also set their own rules regarding the circumstances of exclusion, for example when students fail multiple examinations or lack study progress.

Study trajectories are also shaped through less obvious channels, for example, through incentive schemes in the form of public allowances. In some systems, public study aid can be converted into a grant if students graduate within a study programme's standard study duration, thereby incentivizing students to finish on time. By contrast, public financial aid may be only granted to students of a certain age range or for a certain duration. Societal expectations regarding the timing and duration of educational phases may also impact study trajectories. For example, graduates with study interruptions and delayed degree completion might be regarded with suspicion by employers, whereas smooth study paths could be perceived positively as signals of perseverance and determination. Life course researchers also stress the role of age norms, defined as socially embedded expectations that guide actions and decisions with reference to one's age (Settersten & Mayer, 1997). Certain activity phases (education, family formation, employment or retirement) are related to an appropriate duration or timing within the life course. Observing one's surroundings and inferring from the most common age-transition links (Hillmert & Jacob, 2005, p. 418), older students could feel more pressured to complete higher education quickly compared to their younger fellow students who may not care too much about graduating as soon as possible.

However, such age norms may also differ from country to country. Some education and life course regimes – and consequently higher education systems – promote very uniform pathways, whereas other systems emphasize flexibility and are more open regarding the timing and duration of specific life course passages (Pallas, 2003). For example, life courses are generally less standardized in the United States, where higher education is less age-graded compared to countries such as Germany, where higher education is still mainly perceived as an institution that offers initial tertiary degrees to secondary school leavers rather than being a lifelong learning institution. Thus, most study programmes are limited to a full-time study offer, hampering the participation of students in other life phases, such as working students or students with care duties.

Despite these influences, students 'are not passive recipients of social influence' but have the capability to 'chart their own life course' (Crosnoe & Benner, 2016, p. 183). Prior to enrolment, prospective students make decisions on whether, where and what to study. After the initial enrolment decision, students make a series of smaller and supposedly less consequential decisions that ultimately shape their study trajectories, such as whether to continue their studies, or adapt and modify their earlier enrolment decision by switching the subject, changing the number of courses or extending the time frame of their studies beyond the originally intended duration.

This is a peculiarity in contrast to other – usually earlier – educational phases: higher education is not mandatory, less regulated and less standardized. Consequently, changes can be made all along the study trajectory and are not bound to specific time points. Students have more leeway regarding their study progress and timing of graduation. They may also change their major, withdraw from their studies and decide to re-enrol at a later point in time. In many systems, study interruptions (also known as stop-outs), and corrections of the initial study choice in terms of major or university changes, are possible and widespread. In general, this latitude of options is advantageous for many students as it allows them to flexibly adapt to unforeseen circumstances. On the other hand, this hampers some students from staying on track, given that more autonomy also requires more initiative and pro-active behaviour. For example, in setups where obligatory tracking points or pre-defined milestones are lacking, students cannot be compelled to action. Across OECD countries, 39% of all full-time students graduate from their studies in the indicated standard study duration – with great variation across countries. Adding three additional years, this figure rises to 67% (OECD, 2019, p. 208).

The variability of study trajectories in terms of duration, timing and unforeseen changes necessitates grasping study trajectories in a holistic sense.

An Introduction to Sequence Analysis in the Social Sciences

Originating from the life sciences, sequence analysis was introduced in the 1960s to analyze long strings of protein and DNA sequences (Abbott & Tsay, 2000). The first applications in a sociological context date back to the 1980s, with Andrew Abbott's pioneering works (Abbott, 1995). Since then, sequence analysis has become more established, particularly in life course, demographic or family research. In the early 2000s, criticism of the weaknesses of sequence analysis in social science applications turned out to be an impetus to improve the methodological toolbox and push its developments (Aisenbrey & Fasang, 2010). Currently, the field is still advancing, both in terms of application areas and methodological developments.

Sequence analysis is an explorative technique to detect regularities and resemblance in sequences. Its purpose is to provide a parsimonious model of social reality – here: time-order processes – that allows us to observe broader patterns that would otherwise remain unobserved. Theoretically, sequence analysis rests on the assumption that

> . . .specific social elements and experiences (depending on the order in which they occur) indirectly connect actors to each other. Actors' similar sequence experiences create affinities between them, and signal their shared location in social structure. (Cornwell, 2015, p. 32)

In contrast to other common quantitative social science methods, sequence analysis does not focus on single dependent variables (outcomes or events), but on sets of arrays that are ordered by time (Abbott, 1995; Abbott & Tsay, 2000). This can be repeated categorical information, such as an ordered range of events or activities over time, measured in years, months or hours.

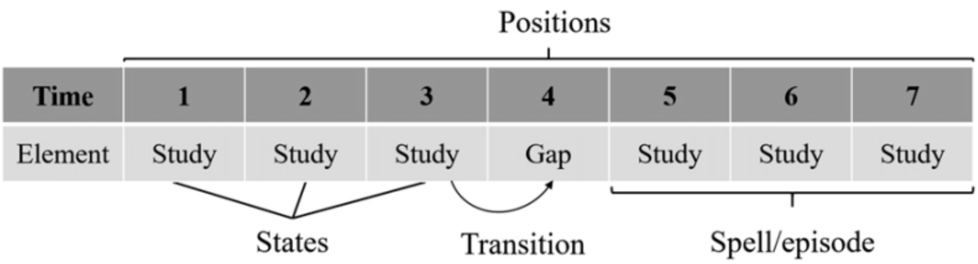

Fig. 1. Illustration of a Sequence and Sequence-Analytical Terminology.
Source: Author's depiction; following Martin and Wiggins (2011, p. 388).

As depicted in Fig. 1, sequences consist of an ordered range of states (also called elements) and transitions. Thus, a sequence could be understood as the more technical equivalent of a trajectory, course, path or career. Each position features one state which can be an attribute or status that may change within a finite range of states. A list of identical elements is called an episode (Sackmann & Wingens, 2003).

For study trajectories, possible states could be enrolment types such as full-time or part-time enrolment, the university or degree type enrolled for. These are ordered by time, using either a monthly or semester scale. The definition of sequence states is not trivial. It should be guided by parsimony whereas the constructed states should be mutually exclusive and exhaustive. Often, one is interested in more than one dimension, for example, when both the degree type and the higher education type are of interest. While multidimensional states are in themselves not a problem, the number of states may rise quickly. As an alternative, multi-channel sequence analysis might be an option (Gauthier, Widmer, Bucher, & Notredame, 2010).

Sequence analysis is an umbrella term for a range of related methods. At its core, the degree of (dis-)similarity between sequences is assessed based on some pre-defined criteria. Two strategies are widely used: first, assessing the similarity of all sequences in a sample by comparing each sequence against each other; and second, as an alternative, each sequence can be compared to an ideal-type or reference sequence.

Depending on the research question, different characteristics of sequences might be of interest, subsequently guiding the research process. Studer and Ritschard (2016, p. 483) stress the following as the most important ones:

- experienced states, that is, the states that appear in a specific sequence
- timing, that is, the position at which a state appears for the first time in a sequence
- duration, that is, sequence and spell lengths, and
- sequencing, that is, the order of states.

How is the similarity of sequences assessed? Multiple algorithms have been developed, each emphasizing different sequence attributes. However, their logic is often alike in that the similarity of two sequences is assessed by examining how many transformations in one sequence would be necessary for it to become identical to the other sequence. Consequently, the number of transformations to make them identical is summed up.

A number of transformation steps are linked to a priori defined parameters. Often, there is not only one way to make a sequence identical to another one. There are different strategies to set these parameters. The frequently used optimal matching approach identifies the lowest possible transformation cost when comparing two sequences based on two types of sequence alignment.

First, a sequence can be modified so that it is identical to another one by inserting and deleting (often abbreviated as indel) some elements. Conversely, sequence alignment can be also reached by substituting some elements with other elements at some positions (Fig. 2). Given the pre-defined cost parameters, sequence analysis will assess the similarity of two sequences based on the 'cheapest' alignment operations. Consequently, it generates for each pair of compared sequences a dissimilarity score which tells us how closely two sequences resembled each other.

There are many different alignment paradigms, allowing for both substitutions and indels or only one of them. The notion that sequence similarity should be theoretically grounded is important: what constitutes sequence similarity in the framework of a specific research question? Is it the order of elements as sequence attributes that matter, the positions where certain elements appear in a sequence or is it the length of episodes (Studer & Ritschard, 2014, 2016)? For example, if one is specifically interested in the timing and duration of episodes, then the indel operation costs should be set higher than the substitutions costs. Consequently, the preferred alignment operation yielding the lowest alignment costs will be

substituting the differing elements at the respective positions. This is known as the Hamming distance. By contrast, when the order of elements is a particularly important sequence attribute, alignment costs should be defined in such a way that substitutions costs are higher than indel operations, thereby avoiding the former. This measure to assess sequence similarity is called Levenshtein II distance.

There might be convincing theoretical arguments for choosing uneven substitution costs as some elements might be conceptually more distant from each other. For example, one may define a higher cost when aligning the elements 'full-time studies' and 'inactive' by contrast to 'full-time studies' and 'part-time studies'. Furthermore, there are many other alignment regimes that may set a specific focus in determining sequence similarity, such as spell-adjusted distances, transition-rate based transformation costs or modified optimal matching. Yet, in most situations, different transformation cost parameters do not lead to major differences in outcomes (Halpin, 2012; Robette & Bry, 2012; Studer & Ritschard, 2014, 2016).

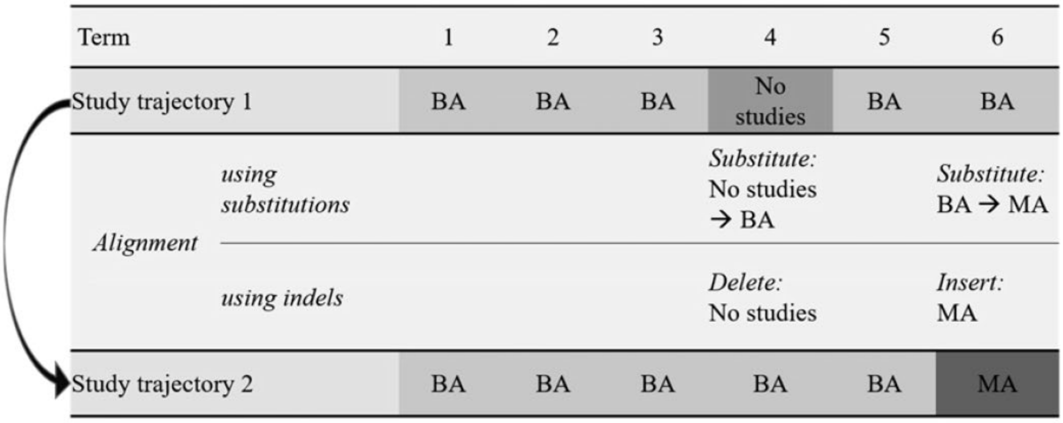

Fig. 2. Illustration of Sequence Alignment. *Note*: BA = Bachelor; MA = Master.

When comparing all sequences with each other, the comparison of sequence pairs results in a matrix of dissimilarity indicators. Comparing all sequences to one particularly meaningful reference sequence, such as an ideal-type sequence, results in one dissimilarity indicator per sequence. In the former approach, sequence analysis is usually followed by cluster analysis, grouping similar sequences based on the dissimilarity matrix. Subsequently, the resulting variables of both approaches might be utilized as dependent variables using regression techniques.

Data analysis packages are available for R (TraMineR, Gabadinho, Ritschard, Mueller, & Studer, 2011) and Stata (SADI and SQ; Brzinsky-Fay, Kohler, & Luniak, 2006; Halpin, 2017). These packages come with a range of sequence description and visualization options. These can be helpful to gain an overview over trajectories or other longitudinal or processual data even when not applying sequence analysis.

As there are so many options in terms of sequence preparation and alignment cost setting, sequence analysis is a quite decision-intensive method. It has been compared to ‘fishing for patterns' where sequence similarity is based on arbitrarily defined transformation costs (Aisenbrey & Fasang, 2010, p. 429; Wu, 2000). Yet, this criticism referred to an early stage of sequence analysis when most applications were grounded on the default optimal matching algorithm. Many more alignment principles have been developed, allowing for an improved match between the theoretical importance of certain sequence attributes and the respective alignment algorithms. Although different cost parameters have been found to often yield fairly comparable results, selecting an appropriate alignment procedure instead of sticking to the default cost parameters has become a crucial step for sequence analysis.

It should be emphasized that sequence analysis is an exploratory method that aims at reducing complexity and providing a parsimonious model of social reality – here: of time-order processes – thereby allowing us to observe broader patterns that would otherwise remain unobserved. Its holistic focus makes it difficult to grasp the processes and mechanisms that impact or trigger the occurrence of certain events within such sequences (Aisenbrey & Fasang, 2010, pp. 434–5).

Example analysis: A reconstruction of bachelor students' study trajectories in Germany

Data Set and Data Preparation

This analysis uses the starting cohort five of the German National Education Panel Study (NEPS, Blossfeld & Roßbach, 2019).[1] The target group of this longitudinal survey are students who enrolled for the first time in German higher education in winter 2010/2011. The sample consists of students enrolling in public or state-approved private higher education. This refers mostly to students in bachelor's degree programmes at universities and universities of applied

[1] More information on the project and research infrastructure is provided on the website: https://www.neps-data.de/.

sciences. Students in nonconsecutive, longer degree courses were excluded in this analysis. The NEPS data seem particularly suitable for the following analysis given its panel design: students were surveyed twice a year, including a set of questions focusing on updates in students' higher education careers. This is vital, given that the reconstruction of trajectories requires detailed information about different enrolment phases, including the beginning and end of a period, information on major changes and interruptions between or within such phases, and whether or not these study phases have been successfully completed. Such detailed information is usually not covered in retrospective surveys. In contrast to administrative data that are often restricted to single organizations, data collection is not limited to a specific organization. On the other hand, as with other panel studies, panel attrition is a major validity threat, ranging between 50.6% and 73.5% compared to wave one participation (Zinn, Steinhauer, & Aßmann, 2017, pp. 16–17).

In terms of data preparation, overlapping enrolment spells (i.e. co-enrolment) were completely deleted or trimmed according to pre-defined set of rules. The Newspell ado for Stata (Kröger, 2015) is a helpful tool to clean such data.

Table 1 provides an overview over the analytic sample, comprising almost 7,000 students. More than two-thirds of these students enrolled at a university, about a quarter at a university of applied sciences and less than 3% in private higher education, which are in the majority of cases universities of applied sciences.

Post-bachelor degree completion trajectories were not considered. In principle, the data allow for a reconstruction of study trajectories of up to 65 months, i.e. almost 5.5 years. Many study trajectories are shorter – either because students dropped out or completed their studies. This is reflected in the actual average length of the trajectories displayed in Table 1: 45.55 months,

Table 1. Sample Description.

Student Sample by Higher Education Type	*N*	Percent
University	4,964	71.72
University of applied sciences	1,776	25.66
Private	181	2.62
Total	6,921	100.00
Sequence Length		**Months**
Longest possible length of sequence		65.00
Actual average length of sequence		45.55
Distribution of States (months)	**Person-months**	**Percent**
Bachelor studies	221,452	70.25
Bachelor studies, not finished	60,880	19.31
Other study phases	28,349	8.99
Study interruption	4,543	1.44
Total	315,224	100.00

i.e. nearly four years. Note that in many sequence-analytical applications, sequences are aligned to a common length, assigning an element for 'empty' positions at sequences that start later or end early. This is not the approach of the following analysis, thereby allowing a focus on overall length as well.

In the prepared data set, one study state is assigned to each month of each person. In total, four states are considered, as shown at the bottom of Table 1. The default state refers to studying in a bachelor degree programme. The second state comprises study phases in a bachelor degree course that were not finished. That is, students either dropped out from this programme or did not (yet) graduate at the end of the time frame. The third study state 'other' contains additional study phases. These are additional not completed study phases in other degree programmes, including bachelor degree programmes. The last study state comprises study interruptions.

Describing Study Trajectories

Based on Stata's SQ package, a couple of sequence indicators were calculated that are helpful to make the study trajectories more tangible (Table 2). On average, students experienced 1.31

episodes on their pathways towards a bachelor degree. This means that experiencing more than one episode (which must be by sample definition an episode of bachelor studies) is actually not so common, suggesting that the study trajectories look overall not too complex. Note that this number differs between university types: it is significantly lower among students enrolled at both universities of applied sciences and in private higher education compared to university students.

The indicators in the following focus on the trajectory length. This refers to the entire time that a student was enrolled in higher education in the considered time frame from 2010 until bachelor degree completion or dropout. If a student is still enrolled at the end of the time frame in February 2016, it takes the maximum value of 65 months. On average, they have a length of 44.1 months – and thereby slightly differing from Table 1 as these estimates are weighted and survey-design adjusted. Students must not only be enrolled in the degree programme which they eventually complete but can have an enrolment history prior to that. This is reflected in the difference between the length of the entire trajectory and the degree-completed episode, which is on average 41.3 months long. In terms of duration, the trajectories of students in universities of applied sciences are significantly longer. This makes sense given that the indicated standard study duration is often seven or even eight semesters, whereas it is usually six semesters at the universities. Comparing these two length indicators also shows that university students spend on average more time in other study phases that do not lead to a bachelor degree.

Turning to specific states in the lower half of Table 2, 77% of all students in this sample experienced an episode that was completed with a bachelor degree. In other words, more than three quarters of all students obtained a bachelor's degree within the observation window. The share is slightly higher among students at universities of applied sciences and significantly higher among students in private institutions. Reflecting this, the exact opposite holds for study

Table 2. Descriptive Sequence Indicators.

	Total	University Type		
		University	UAS	Private
Mean number of episodes	1.31	1.38	1.23***	1.09***
Length in months (mean): Duration to bachelor degree completion				
Entire trajectory	44.1	43.8	45.1**	40.7**
Degree-completed bachelor studies episode only	41.3	40.4	43.0***	40.5
Proportion of students who experienced enrolment state, in %				
Bachelor studies	76.6%	74.2%	79.2%*	90.2%***
Bachelor studies, not completed	34.3%	39.0%	28.7%***	11.1%***
Other study phase	12.4%	14.7%	9.2%***	3.9%***
Study interruption	5.5%	6.2%	4.4%*	3.4%*

Note: *p <0.05, **p <0.01, ***p <0.001.

phases that ended without obtaining a degree. The share is 39% among university students, indicating that almost 40% of all university students were at least at one point in their study career enrolled in a bachelor degree course that they left without completion. These students switched degree course, dropped out from university or were still enrolled but did not yet complete their degree course at the end of the time frame. University students also reported significantly more other study phases. Overall, the proportion of students who reported a study interruption is low and again is slightly lower among students at the universities of applied sciences.

Table 3 provides another descriptive overview by displaying the most common sequences of all students by university type. Note that this refers to similar sequences in terms of the order of episodes, ignoring differences in duration. Overall, there is no great diversity in the study trajectories towards a bachelor degree: almost two-thirds (65.4%) of all students remain enrolled in the same bachelor degree programme until graduation. Other significant trajectory types are enrolling and not completing (12.2%), enrolling and switching to another bachelor degree course, either immediately (8.4%) or interrupting one's study career in between (2.1%). Five percent of all students switched from one programme to another and did not yet graduate

Table 3. Five Most Common Sequences by Higher Education Sector.

		University Type		
	Total	University	UAS	Private
BA, completed	65.4%	60.3%	71.4%	88.4%
BA, not completed	12.2%	12.5%	12.3%	5.9%
BA, not completed > BA, completed	8.4%	10.1%	6.0%	
BA, not completed > BA, not completed	5.0%	5.7%	4.2%	
BA, not completed > interruption > BA, not completed	2.1%		2.3%	3.1%
BA, not completed > Other degree programme		2.1%		
Total (most common sequences)	93.0%	90.7%	96.3%	97.5%

Note: Weighted and survey-design adjusted. Sequences with less than 2% of all students not shown. UAS = university of applied sciences.

after more than five years. These five trajectory types account for 93% of all study trajectories. Conversely, 7% of all students follow other, less common trajectory types. Some differences by university type appear. First, more students at universities of applied sciences follow the more common study trajectories. A major difference is the percentage of students following a smooth, linear trajectory (i.e. enrolling for only one bachelor degree course over the course of their trajectory and completing it): 88.4% and 71.4% in private higher education and at the universities of applied sciences, respectively, compared to 60.3% at the universities.

Cluster Analysis

These descriptive indicators show that the study trajectories of bachelor students in Germany seem overall not too complex, but that those of university students are more diversified and discontinuous compared to those in private higher education and at the universities of applied sciences.

This description of sequences was followed by sequence analysis. Intending to focus specifically on sequence order instead of the duration of episodes, the costs for indel operations were set inferior to the costs for substitutions. Thus, the alignment procedure favours indels over substitutions. Subsequently, a hierarchical cluster analysis was conducted based on the resulting dissimilarity matrix. The selection of the optimal number of clusters is based on theoretical considerations, visual inspection (dendrogram) and goodness-of-fit indicators (Calinski-Harabasz index and the Duda-Hart cluster stopping indices, cf. Cornwell, 2015; Halpin, 2016).

The results are depicted in Fig. 3. Four clusters are shown. On the left side, these clusters are displayed as sequence index plots, that is, individual trajectories are depicted. The x-axis refers to time whereas the individual sequences are stacked upon each other. The percentage plots shown on the right-hand side refer to the same clusters but provide a different view: they display the distribution of states at each month. That is, they do not show individual trajectories, but provide distributional bar graphs for each month and thus provide a cumulative perspective on these clusters.

Cluster 1 comprises students who remained enrolled in the same bachelor programme over the course of their studies and also completed this degree course successfully. Fig. 3 points to great variation in terms of study duration, ranging from less than three years to more than five years. This trajectory type comprises two-thirds of all students – with variation across higher education types, ranging from 61% at the universities to 88% at the private institutions (Table 4).

Cluster 2 also comprises students who obtained a bachelor degree, but their trajectories were not as smooth compared to those in cluster 1. All of these students switched at least once from one degree programme into another along the way, sometimes also interrupting their studies for some time. This applies to about 10% of all students, with the proportion is higher at the universities compared to other higher education types.

Cluster 3 comprises 20% of students who switched from one programme to another and who, except for a tiny minority of cases, did not (yet) complete their bachelor studies. About half of these students were still enrolled, i.e. they may still complete their studies after the five-year

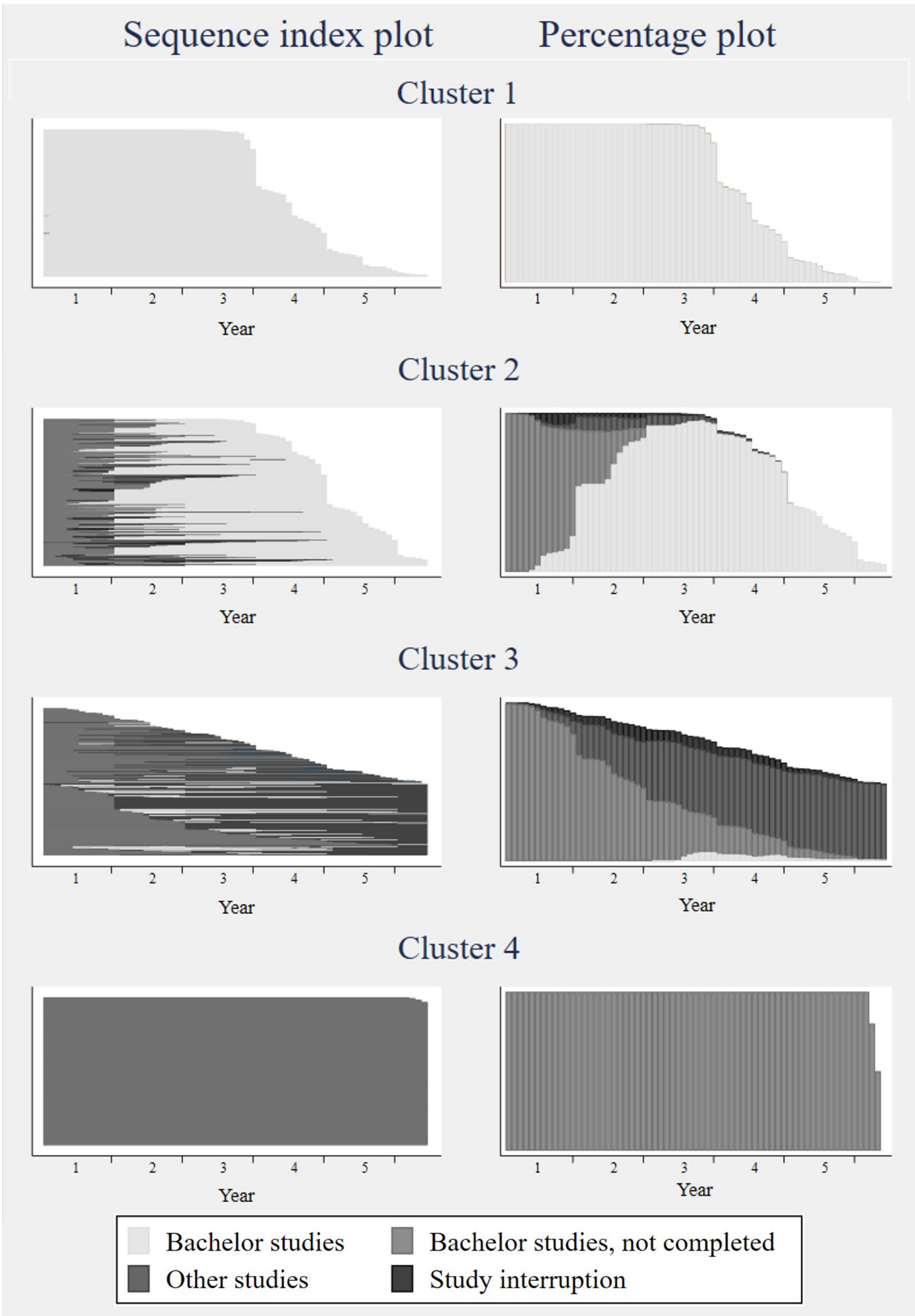

Fig. 3. Sequence Index Plots and Sequence Percentage Plots of Four Trajectory Cluster.

time frame, while the other half has left without a degree. There seems to be no consistent time point at which students drop out. There is great variability, ranging from less than one year up to more than five years after enrolment. This trajectory type is less common among students enrolled at private institutions, compared to students at the universities and universities of applied sciences.

The last cluster, cluster 4, comprises about 5% of all students. These are students who remain enrolled for the bachelor degree course for which they initially enrolled for in winter 2010 until the end of the window of observation after almost 5.5 years. Increasing the time window would be necessary to know whether these students will eventually complete their studies or not. Again, this applies more often to students at the universities compared to private higher education.

The potential of sequence analysis in higher education research

This section discusses applications of sequence analysis in three different higher education research areas: (1) post-secondary education pathways, (2) academic career trajectories and (3) development trajectories of higher education organizations. Recent applications of sequence analysis will be discussed for the first two themes, whereas, to the best of my knowledge, the latter has not yet been covered extensively although sequence analysis applications seem promising in this regard.

Post-Secondary Education and Education-Work Trajectories

The first set of studies considered here – perhaps the field with the most sequence analysis applications so far in higher education-related research – covers adolescents' trajectories through higher education or beyond into the labour market. Regarding the former, higher education researchers are mostly interested in gaining a better understanding of students' progression pathways, where they go astray and identifying predictors of certain study trajectory patterns. In this vein, being particularly interested in the dropout risk of traditional versus non-traditional students, Sánchez-Gelabert (2021) reconstructed and contrasted the study trajectories of undergraduate students in one Spanish region in distance and presence mode studies. Given the abundance of study opportunities and the rather de-standardized school-to-work pathways in the United States, sequence analysis seems particularly vital to gain an overview over the major patterns.

Against this background, Monaghan (2020) reconstructed the college attendance patterns of a cohort of US individuals between the ages of 18 and 39, identifying four distinct trajectory types: rapid completers, marginal college-goers, lifelong learning students and delayed completers. In a subsequent step, the author linked these to employment and family formation patterns. Noteworthy about this study is its long time span, which also allows grasping more non-traditional, extended enrolment trajectories. In a similar vein, Boylan (2020) reconstructs the trajectories of a sample of students in the United States over 107 months, identifying six distinct post-secondary pathways and their relationship to students' social origin. With regard to transitions from higher education into the labour market, and being specifically interested in social reproduction mechanisms, Duta, Wielgoszewska, and Iannelli (2021) reconstructed the occupational trajectories of graduates based on the 1970 British Cohort Study. They distinguished broad activity categories (in education, employed, inactive, unemployed), but further differentiated employment states by actual occupational status.

These sequence analysis-related contributions range from more fine-grained applications on study and enrolment patterns to broader applications covering higher education studies vis-a-vis other life course stages such as employment or family formation.

Academic Career Trajectories

In a similar vein, a few recently published studies have reconstructed academic career trajectories. Interestingly, some focus on very successful academics, such as full professors, exceptionally successful or internationally mobile academics.

Resting upon a theoretical approach of resource accumulation across the professional life course, Benz, Bühlmann, and Mach (2021) developed a typology of professors' careers at Swiss universities, from pre-tenure positions to full professorships. Rossier (2020) conceptualized academic careers as acquiring different sorts of capital, reconstructing the careers of economics and business studies professors based on the type of positions held, differentiating positions outside academia, academic and managerial positions in academia, and thereby identifying four distinct career patterns leading to a professorship.

Korom (2020) reconstructs the career paths of highly recognized professors in economics and sociology – Nobel laureates in economics and the most widely cited sociologists – by their respective institutional affiliations. Combining sequence analysis with network analysis, he finds that institutional affiliation is a better predictor of scientific excellence in the field of economics compared to sociology. Vinkenburg, Connolly, Fuchs, Herschberg, and Schels (2020) retrace the career histories of applicants to a specific European Research Council (ERC) research grant scheme across the academic positions held and by type of institution, also including non-research states within the trajectories. Among other findings, they do not identify any great gender differences in the career trajectories of ERC applicants. By means of sequence analysis, Czaika and Toma (2017) analyze the international mobility patterns of Indian academics based on their social origin and performance. They were interested in path dependencies in international mobility, identifying how far earlier international mobility forecasts later mobility.

There are still many open questions with regard to academic career trajectories that could be potentially answered in a sequence analysis framework. However, particularly with reference to the question of what predicts 'successful' academic careers, adequate sampling is vital. What does it tell us if we reconstruct the trajectories of those who reached a certain prestigious position while the trajectories of those who pursued a similar path but took a different line along the way remain unobserved? In this vein, it seems particularly relevant to retrace the trajectories of an entire cohort of early-stage academics or doctoral students and follow their pathways through academia, identifying at which stage academic careers start diverging or identifying early markers of successful versus problematic academic careers.

Development Trajectories of Higher Education Organizations

Although sequence analysis has been predominantly used to retrace individual-level trajectories, it also holds potential to reconstruct meso- or even macro-level developments. A potential application area in this regard is the development of higher education organizations, for example, vertical or horizontal differentiation. In this vein, sequence analysis could be a tool to observe the evolution of different types of universities over time. For example, by reconstructing university ranking positions over time, we may learn about the evolution pathways of excellent universities, such as the degree of stability or volatility as well as path dependencies in their development. Universities' student numbers or their study offer (two-year only, BA granting, MA granting, PhD granting) could be used to reconstruct the development of newer universities in order to gain some insights on more or less successful development pathways. By doing so, we may learn more about common development patterns, stability and volatility of universities' standing or the role of path dependencies in higher education organizations' development.

Discussion and outlook

Sequence analysis is a useful method in several areas of higher education research, specifically where the development (pathways, careers) of entities – be it sequence analysis in this respect is its holistic perspective which enables the detection of common trajectory patterns, irregularities and deviations.

A couple of reflections and practical recommendations complete this contribution. As a rather general caveat, the data preparation process might be extensive and not as straightforward as with other quantitative data analysis techniques. First of all, sequence analysis requires information on spells. That is, it is not only enough to have a longitudinal dataset but start and end dates of the episodes need to be known (unless one is only interested in the order of states and not in duration). With regard to students' trajectories, administrative data sources might have an edge over survey data given that they are not dependent on an individual's disposition to participate (again) in a survey. Depending on the research questions, panel attrition could be a major problem for sequence analysis approaches given that the trajectories of those dropping out from the survey remain unobserved, potentially inducing bias. On the other hand, administrative data sources usually fall short in that they provide not much beyond the usual information on demographic characteristics, and rarely provide in-depth information on students' pre-higher education pathways, motivation or attitude.

The utilized data also need to cover a sufficiently long time span. In the reconstruction of study trajectories, the time span covered should be beyond the standard study duration or the expected standard pathway. Otherwise – as illustrated in the example analysis in this contribution – one could end up with trajectories with unknown outcomes. Consequently, it remains unclear for a group of students whether they will still eventually complete a degree after the window of observation or whether these students are dropouts.

Most data sets are not readily prepared for sequence analysis approaches. Extensive data cleaning could be necessary as survey information is incomplete, inconsistent or contradictory. Referring once more to study trajectories, it might be necessary to define a set of rules regarding which enrolment episodes to keep, which ones to eliminate and which ones to trim in the case of overlapping enrolment information provided or parallel enrolment episodes.

Another recommendation is to carefully consider one's expectations towards the sequence analysis outcomes beforehand. Sequence analysis is an exploratory, data-reducing and pattern-detecting method. This should be kept in mind, considering that it may complicate testing a priori defined hypotheses and mechanisms.

In general, sequence analysis is of great potential in application areas that are characterized by a high degree of volatility, complexity and variability, and thus do not allow gaining an understanding of the data patterns using other descriptive methods. For example, sequence analysis seems particularly useful in less standardized educational contexts, such as in retracing the course enrolment patterns of US community college students. Given the increasing availability of student register data or longitudinal educational panel studies in many countries, there is a great potential for sequence analysis in student research.

Sequence analysis does not free us from a general claim brought forward towards higher education research, namely the necessity for explicit theorizing (Kosmützky, 2016). To avoid 'fishing for patterns' (Wu, 2000), sequence analysis needs to be theory-grounded, both with regard to the definition of states as well as when setting the cost parameters for sequence alignment operations. Although popular, the default optimal matching approach might not always be the best solution. Instead, researchers should carefully examine the sequence characteristics of interest: is it about the duration of episodes, their occurrence or the overall complexity of sequences (Cornwell, 2015)? Although different cost parameters have been found to often yield fairly comparable results, a range of alternative alignment principles have been developed, allowing for an improved match between the theoretical importance of certain sequence attributes and the respective alignment principles.

Related to this, another suggestion is to use sequence analysis more explicitly in comparative approaches. Leaving major challenges such as data availability and equivalence aside for a moment, sequence analysis is a great method to compare developments across groups, institutions or countries. The majority of studies, including the example provided here as well as the reviewed studies in the previous section make use of a default strategy – sequence and cluster analysis – followed by a prediction of cluster types using some type of regression. However, a major potential of sequence analysis lies in description – reducing complex data on over-time developments to meaningful indicators and visualizations – across groups of entities. An alternative approach here might be to compare the sequences of pre-defined groups from the outset. One could stress commonalities and differences without the detour of clustering, but instead develop a set of meaningful indicators that could be compared across groups. In this vein, comparing trajectories of individuals with common starting points and observing their divergent development over time is of interest. For instance, we are not only concerned in the formation of academic careers of those who were successful, but in a comparison of the 'successful' group with those who took a different pathway. In this vein, panel data sets such as the NEPS student panel but also the German National Academics Panel Study (NACAPS, Briedis et al., 2018), gathering survey data from the early stage of doctoral studies throughout the academic career, seem promising.

Ultimately, this contribution aimed to bring together higher education research and sequence analysis as a methodological supplement in cases where timing and duration is of interest. The methodological toolbox is continuously advanced – there are many applications that go beyond the possibilities presented here and the often considered 'sequence analysis →cluster analysis →regression' approach. In this regard, advances towards combining sequence and transition or event-centred perspectives seem promising as well as the multi-channel sequence analysis or advancements that combine sequence analysis with network analysis (Blanchard, Bühlmann, & Gauthier, 2014; Cornwell, 2015; Milesi, 2010; Ritschard & Studer, 2018; Rossignon, Studer, Gauthier, & Le Goff, 2018).

Note: This contribution includes unpublished parts from the dissertation: ‘Higher education trajectories and social origin in Germany and the United States: A comparative sequence-analytical approach’ by Christina Haas, University of Luxembourg, 2021.

Acknowledgements

This paper uses data from the National Educational Panel Study (NEPS; NEPS-Netzwerk, 2021; see Blossfeld & Roßbach, 2019). The NEPS is carried out by the Leibniz Institute for Educational Trajectories (LIfBi, Germany) in cooperation with a nationwide network.

References

Abbott, A. (1995). Sequence analysis: New methods for old ideas. Annual Review of Sociology, 21(1), 93–113.

Abbott, A., & Tsay, A. (2000). Sequence analysis and optimal matching methods in sociology: Review and prospect. Sociological Methods & Research, 29(1), 3–33.

Aisenbrey, S., & Fasang, A. E. (2010). New life for old ideas: The "second wave" of sequence analysis bringing the "course" back into the life course. Sociological Methods & Research, 38(3), 420–462.

Benz, P., Bühlmann, F., & Mach, A. (2021). The transformation of professors' careers: Standardization, hybridization, and acceleration? Higher Education, 81(5), 967–985.

Blanchard, P., Bühlmann, F., & Gauthier, J.-A. (2014). Advances in sequence analysis: Theory, method, applications (Vol. 2). Cham: Springer.

Blossfeld, H.-P., & Roßbach, H.-G. (2019). Education as a lifelong process: The German national educational panel study (NEPS) (Edition ZfE). Wiesbaden: Springer VS.

Boylan, R. L. (2020). Predicting postsecondary pathways: The effect of social background and academic factors on routes through school. Socius, 6, 1–25.

Briedis, K., Lietz, A., Ruß, U., Schwabe, U., Weber, A., Birkelbach, R., & Hoffstätter, U. (2018). Nacaps 2018. Daten- und Methodenbericht zur National Academics Panel Study. Hannover: DZHW.

Brzinsky-Fay, C., Kohler, U., & Luniak, M. (2006). Sequence analysis with Stata. STATA Journal, 6(4), 435.

Buchmann, C., & DiPrete, T. A. (2006). The growing female advantage in college completion: The role of family background and academic achievement. American Sociological Review, 71(4), 515–541.

Cornwell, B. (2015). Social sequence analysis: Methods and applications. New York, NY: Cambridge University Press.

Crosnoe, R., & Benner, A. D. (2016). Educational pathways. In J. T. Mortimer, M. J. Shanahan, & M. K. Johnson (Eds.), Handbook of the life course (pp. 179–200). Cham: Springer.

Czaika, M., & Toma, S. (2017). International academic mobility across space and time: The case of Indian academics. Population, Space and Place, 23(8), 1–19.

Deil-Amen, R. (2011). The "traditional" college student: A smaller and smaller minority and its implications for diversity and access institutions. Paper presented at the mapping broad-access higher education conference at Stanford University.

Duta, A., Wielgoszewska, B., & Iannelli, C. (2021). Different degrees of career success: Social origin and graduates' education and labour market trajectories. Advances in Life Course Research, 47, 100376.

Gabadinho, A., Ritschard, G., Mueller, N. S., & Studer, M. (2011). Analyzing and visualizing state sequences in R with TraMineR. Journal of Statistical Software, 40(4), 1–37.

Gauthier, J.-A., Widmer, E. D., Bucher, P., & Notredame, C. (2010). Multichannel sequence analysis applied to social science data. Sociological Methodology, 40(1), 1–38.

Goldin, C., Katz, L. F., & Kuziemko, I. (2006). The homecoming of American college women: The reversal of the college gender gap. The Journal of Economic Perspectives, 20(4), 133.

Gross, C., & Hadjar, A. (2021). Institutional characteristics of education systems and inequalities—Introduction I. International Journal of Comparative Sociology, 61(6), 381–388.

Halpin, B. (2012). Sequence analysis of life-course data: A comparison of distance measures. University of Limerick department of sociology working paper series. Department of sociology, University of Limerick.

Halpin, B. (2016). Cluster analysis stopping rules in Stata. Department of sociology working paper series. University of Limerick.

Halpin, B. (2017). SADI: Sequence analysis tools for Stata. STATA Journal, 17(3), 546–572.

Hillmert, S., & Jacob, M. (2005). Institutionelle Strukturierung und inter-individuelle variation. Kölner Zeitschrift für Soziologie und Sozialpsychologie, 57(3), 414–442.

Hout, M. (2012). Social and economic returns to college education in the United States. Annual Review of Sociology, 38, 379–400.

Hurtado, S. (2007). The study of college impact. In P. J. Gumport (Ed.), Sociology of higher education: Contributions and their contexts (pp. 94–112). Baltimore, MD: Johns Hopkins University Press.

Korom, P. (2020). How do academic elites March through departments? A comparison of the most eminent economists and sociologists' career trajectories. Minerva, 58(3), 343–365.

Kosmützky, A. (2016). The precision and rigor of international comparative studies in higher education. In Theory and method in higher education research (pp. 199–221).

Kröger, H. (2015). Newspell—Easy management of complex spell data. STATA Journal, 15(1), 155–172.

Lacy, T. A. (2015). Event history analysis: A primer for higher education researchers. In Theory and method in higher education research (pp. 71–91). Bingley: Emerald Publishing Limited.

Lassibille, G., & Navarro Gómez, Ma L. (2009). Tracking students' progress through the Spanish university school sector. Higher Education, 58(6), 821–839.

Lörz, M., & Mühleck, K. (2019). Gender differences in higher education from a life course perspective: Transitions and social inequality between enrolment and first post-doc position. Higher Education, 77(3), 381–402.

Mare, R. (1980). Social background and school continuation decisions. Journal of the American Statistical Association, 75(370), 295–305.

Marginson, S. (2016). The worldwide trend to high participation higher education: Dynamics of social stratification in inclusive systems. Higher Education, 72(4), 413–434.

Martin, P., & Wiggins, R. D. (2011). Optimal matching analysis. In M. Williams & W. P. Vogt (Eds.), The Sage handbook of innovation in social research methods (pp. 385–408). London: SAGE.

Meggiolaro, S., Giraldo, A., & Clerici, R. (2017). A multilevel competing risks model for analysis of university students' careers in Italy. Studies in Higher Education, 42(7), 1259–1274.

Milesi, C. (2010). Do all roads lead to Rome? Effect of educational trajectories on educational transitions. Research in Social Stratification and Mobility, 28(1), 23–44.

Monaghan, D. B. (2020). College-going trajectories across early adulthood: An inquiry using sequence analysis. The Journal of Higher Education, 91(3), 402–432.

NEPS-Netzwerk. (2021). Nationales Bildungspanel, scientific use file der Startkohorte Studierende. In Leibniz-Institut für Bildungsverläufe (LIfBi). Bamberg. doi:10.5157/NEPS:SC5:15.0.0

OECD. (2019). Education at a glance 2019. Paris: OECD Publishing.

Pallas, A. M. (2003). Educational transitions, trajectories, and pathways. In J. T. Mortimer & M. J. Shanahan (Eds.), Handbook of the life course (pp. 165–184). New York, NY: Kluwer Academic Publishers.

Pfeffer, F. T., & Goldrick-Rab, S. (2011). Unequal pathways through American universities. Institute for Research on Poverty discussion paper.

Ritschard, G., & Studer, M. (2018). Sequence analysis and related approaches: Innovative methods and applications. Cham: Springer.

Robette, N., & Bry, X. (2012). Harpoon or bait? A comparison of various metrics in fishing for sequence patterns. Bulletin of Sociological Methodology/Bulletin de Méthodologie Sociologique, 116(1), 5–24.

Rossier, T. (2020). Accumulation and conversion of capitals in professorial careers. The importance of scientific reputation, network relations, and internationality in economics and business studies. Higher Education, 80(6), 1061–1080.

Rossignon, F., Studer, M., Gauthier, J.-A., & Le Goff, J.-M. (2018). Sequence history analysis (SHA): Estimating the effect of past trajectories on an upcoming event. In G. Ritschard & M. Studer (Eds.), Sequence analysis and related approaches: Innovative methods and applications (pp. 83–100). Cham: Springer.

Sackmann, R., & Wingens, M. (2003). From transitions to trajectories. Sequence types. In W. R. Heinz & V. W. Marshall (Eds.), Social dynamics of the life course. Transitions, institutions, and interrelations (pp. 93–115). New York, NY: Aldine de Gruyter.

Sánchez-Gelabert, A. (2021). Non-traditional students, university trajectories, and higher education institutions: A comparative analysis of face-to-face and online universities. Studia Paedagogica, 25(4), 51–72.

Schofer, E., Ramirez, F. O., & Meyer, J. W. (2021). The societal consequences of higher education. Sociology of Education, 94(1), 1–19.

Schuetze, H. G., & Slowey, M. (2002). Participation and exclusion: A comparative analysis of non-traditional students and lifelong learners in higher education. Higher Education, 44(3–4), 309–327.

Settersten, R. A., Jr. & Mayer, K. U. (1997). The measurement of age, age structuring, and the life course. Annual Review of Sociology, 23(1), 233–261.

Studer, M., & Ritschard, G. (2014). A comparative review of sequence dissimilarity measures. LIVES Working paper 33. doi: 10.12682/lives.2296-1658.2014.33

Studer, M., & Ritschard, G. (2016). What matters in differences between life trajectories: A comparative review of sequence dissimilarity measures. Journal of the Royal Statistical Society, 179(2), 481–511.

Tieben, N. (2020). Non-completion, transfer, and dropout of traditional and non-traditional students in Germany. Research in Higher Education, 61, 117–141.

UNESCO. (2022). UNESCO key figures. Retrieved from https://www.unesco.org/en/education/higher-education. Accessed on March 17, 2022.

Vila, L. E. (2000). The non-monetary benefits of education. European Journal of Education, 35(1), 21–32.

Vinkenburg, C. J., Connolly, S., Fuchs, S., Herschberg, C., & Schels, B. (2020). Mapping career patterns in research: A sequence analysis of career histories of ERC applicants. PLoS One, 15(7), e0236252.

Wu, L. L. (2000). Some comments on ‘Sequence analysis and optimal matching methods in sociology: Review and prospect’. Sociological Methods & Research, 29(1), 41–64.

Zinn, S., Steinhauer, H. W., & Aßmann, C. (2017). Samples, weights, and nonresponse: The student sample of the national educational panel study (wave 1 to 8). NEPS survey paper, Leibniz Institute for Educational Trajectories, Bamberg.