In it for the long run

Rasmus H. Klokker

2024-08-14

Contents

1 Introduction

1800 words

Teachers are the cornerstone of the educational system, and their career trajectories significantly impact the quality of education that students receive. Understanding these career paths can provide critical insights into the factors that influence teacher retention, professional development, and overall job satisfaction. Such knowledge can help policymakers and stakeholders who aim to improve educational outcomes and ensure a stable and effective teaching workforce.

Research on teacher career choices is intrinsically linked to issues of educational inequality. Teachers' decisions regarding where to work, whether to stay in the profession, and how to advance their careers can exacerbate or alleviate disparities in educational opportunities(REF). For instance, schools in low-income or rural areas often face higher turnover rates, leading to a lack of experienced and qualified teachers in these communities(REF). This turnover can result in a cycle of educational disadvantage, where students in underprivileged areas receive lower quality education compared to their peers in more affluent regions(REF).

Moreover, teacher career paths are influenced by various factors such as working conditions, support systems, and professional development opportunities, which are often unevenly distributed across different schools and districts. Schools serving disadvantaged populations frequently lack the resources to provide the necessary support for teacher growth and retention, further perpetuating educational inequality(REF). By examining these career trajectories, researchers can identify key interventions and policies that can support teachers more effectively, particularly in underserved areas.

While there has emerged a scholarly literature on teacher career choices, particularly within the literature on teacher sorting(REF), knowledge on teacher career paths that span long time horizons is still very limited. As such, knowledge on teacher career paths does, in many cases, concern partial teacher careers rather than the full careers of teachers, spanning entry to the labor market all the way to retirement.

In this paper i will investigate patterns of teacher career paths using sequence analysis. Subsequently i will use cluster analysis to form types of teacher career paths and investigate how teacher characteristics, such as teacher credentials, are associated with different types of teacher career paths

Crucially, using high-quality danish register data, i am able to follow cohorts of teacher graduates from the year of graduation and onwards. Given that the earliest cohort graduated in 1980, i am able to investigate career paths of teachers across several decades, enabling me to map out the full teacher career paths of the earliest teacher cohorts.

2 Contribution

Thus, this paper will primarily contribute by expanding our knowledge on the late stages of teacher career paths. This includes the transitions from mid- to late-stage careers as well as how patterns of late-stage teacher careers compare to the early- and mid-stage teacher careers. Crucially, this will also contribute knowledge on the timing of teachers who opt out of teaching. If there exists a non-trivial share of teachers who opt out of teaching in the later stages of their career, earlier studies on teacher careers, and particularly teacher attrition, may underestimate the share of teachers that opt out of teaching. If, however, teachers who opt out of teaching in the earlier stages of their career choose to enter teaching again at a later stage in their careers, previous studies could have overestimated the share of teachers that choose to opt out of teaching indefinitely.

2.1 Findings

2.2 State of the art

In this section i will outline the current scholarly literature on teacher careers.

2.3 What do we know about how teachers change jobs over the course of their career?

In general, only few studies provide broad overviews of teacher career paths(also Lindqvist, Nordänger, & Carlsson, 2014). In Gubler, Biemann, and Herzog (2017) we find the most broad overview of teacher career trajectories. The authors find 6 different types of career paths, with the majority of teachers starting out as primary school teachers before they transition to childcare and part-time work as public school teachers. The second most frequent type of career path involves teachers that have stable employment in primary school teaching across the entire period. While the study of Gubler, Biemann, and Herzog (2017) is closed related to this paper in terms of method and overall research goal, there are some key differences. Firstly, while the authors construct overall patterns of teacher careers using sequence analysis, the authors do not focus on transitions between schools serving different types of students, e.g. low- and high-SES students. Secondly, while the study is able to track teachers across a relatively long period of time(15 years), the sample size is limited to 999 teachers.

The remaining literature on teacher career paths focuses on more specific elements of teacher career choices or are more loosely related to research questions on teacher career paths.

For instance Falch (2022), Goldhaber et al. (2022), and Boyd et al. (2002) all focus on the early careers of teachers. For Falch (2022) and Boyd et al. (2002) the focus is specifically on teacher attrition. Falch (2022) finds that attrition from teaching positions increases steadily as teachers gain experience and that teachers mainly pursue occupations outside teaching. Further, attrition was higher among teachers with higher degrees of academic ability, which is also among the results of Boyd et al. (2002). Goldhaber et al. (2022) find similar results with regards to attrition, but also find that a substantial proportion of newly graduated teachers enter employment outside public primary schools before they transition to teaching positions.

There is a large scholarly literature that does not address teacher career paths directly. However many of the results in this literature are relevant to the present paper, especially in regards to how school and teacher characteristics impact teachers decisions to stay at or leave the school at which they are currently employed.

2.4 What do we know about how teachers change jobs between schools that serve different kinds of students?

There is a considerable literature on how school characteristics, e.g. the composition of the student body, the physical environment of the school, the work environment, quality of school leadership etc., impact teacher attrition and retention. Three comprehensive literature reviews,Guarino, Santibanez, and Daley (2006), Borman and Dowling (2008), and Nguyen et al. (2020), summarize this literature. However, Nguyen et al. (2020), is the most recent review and also encompasses the studies included in the previous reviews. In Nguyen et al. (2020), comparisons of different types of schools, e.g. private vs. public schools, middle-school vs. elementary school etc., are consistently among the factors that increase the propensity for teachers to leave their current school the most. Higher degrees of student disciplinary problems is however is the third highest effect size among the factors that increase attrition, with higher degrees of student discipline problems increasing the probability of leaving ones current school by about 16%. Interestingly, there are more school characteristics that consistently lower teacher attrition. Here, an improved work environment reduced the odds of leaving the current school by nearly 50%, although the estimate is associated with a very high degree of uncertainty. While the effect sizes for providing induction/mentoring to newly hired teachers, opportunities for professional development and administrative support are smaller, the effects are estimated with considederably less uncertainty than the effect of improved work environment. Interestingly, student body characteristics do not seem to impact teacher attrition to any discernable degree. Only the estimated effect of student poverty is of noticeable size, but due to the width of the confidence interval we cannot reject that student poverty might actually decrease teacher attrition.

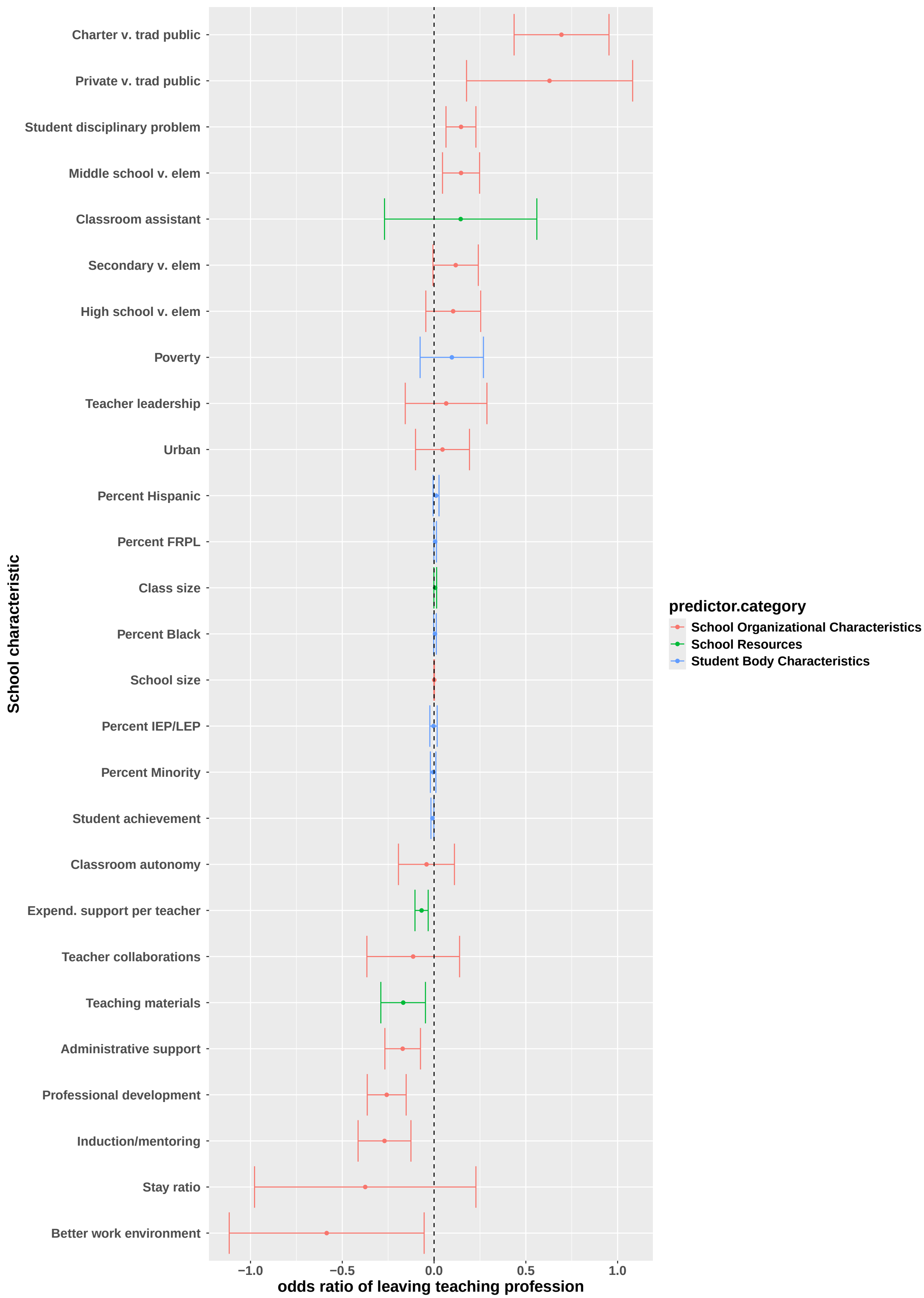


Figure 1: Associations between leaving the teaching profession and school characteristics - reproduced from Nguyen, 2020
(#fig:compare effs)

2.5 What do we know about how teacher characteristics impact career choices?

Regarding teacher characteristics, the estimated effect of having more than 3 years of experience suggests that having more than 3 years of experience increases the odds of attrition by more than 48%. Having a professional specialization does also seem to increase the odds of teacher attrition, with teachers being specialized in other fields than STEM or special education being the most at risk of attrition. Interestingly, having better test scores, e.g. SAT-scores, only has a small effect on attrition of teachers. On the other end of the spectrum, having more children and having a teacher certification reduces the odds of attrition by about 50%. Having a full time teaching position, being in the Hispanic minority group and being older than 28 years of age also reduce the odds of attrition considerably, all reducing the odds of attrition by more than 30%. The effect sizes for the remaining factors are either too small to warrant any attention here or are not statistically significant, and thus we cannot reject that these effect sizes could be either positive or negative.

```
##                    Factor X..of.studies X..of.ES Odds.ratio Log.odds.ratio
## 8      Number of children             4        4     0.501         -0.692
## 21 Standard certification            19       19     0.536         -0.624
## 9             Young child             3        3     0.551         -0.596
## 5     Minority (Hispanic)            17       17     0.591         -0.525
## 11     Full time teaching             6        6     0.619         -0.480
## 26        Experience (<3)            12       12      1.484          0.395
##       SE Lower.bound Upper.bound     p     I2       Q    PQ
## 8  0.165   0.3624024   0.6914254 <.001  <.001   1.755 0.625
## 21 0.181   0.3760625   0.7633795 0.001  98.151 973.469 <.001
## 9  0.153   0.4081992   0.7437874 <.001  <.001   1.086 0.581
## 5  0.102   0.4838405   0.7232502 <.001  98.391 994.446 <.001
## 11 0.142   0.4681343   0.8179124 0.001  65.000  14.286 0.014
## 26 0.086   1.2523227   1.7576892 <.001  92.610 148.841 <.001
##          predictor.category review
## 8  Teacher Characteristics  Nguyen
## 21   Teacher Qualifications Nguyen
## 9  Teacher Characteristics  Nguyen
## 5  Teacher Characteristics  Nguyen
## 11 Teacher Characteristics  Nguyen
## 26   Teacher Qualifications Nguyen
```

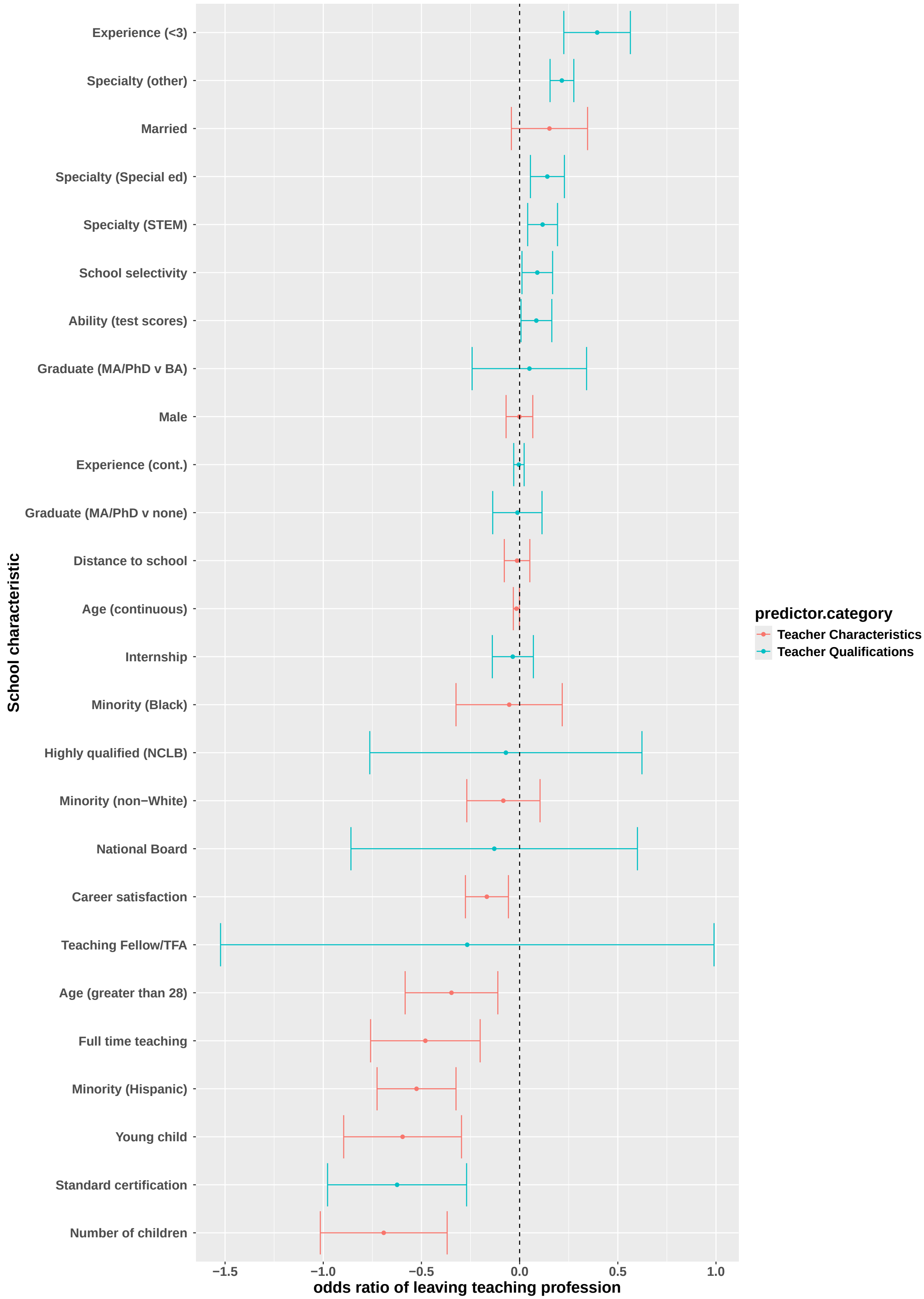


Figure 2: Associations between leaving the teaching profession and teacher characteristics - reproduced from Nguyen, 2020
(#fig:compare effs teacher chars)

As we can see, the scholarly literature suggests that there are a number of characteristics of both schools and teachers that impact the attrition and retention of teachers, and thereby also the career choices of teachers. Interestingly, most student body characteristics and indicators of teacher ability are not among the most impactful of these characteristics. However, it should be noted that the results in Nguyen et al. (2020) pertain to leaving or staying in the current job, and not entire career paths of teachers. As such, the results of Nguyen et al. (2020) cannot overrule choices in relation to the analysis of data that are founded in the theoretical framework of this paper.

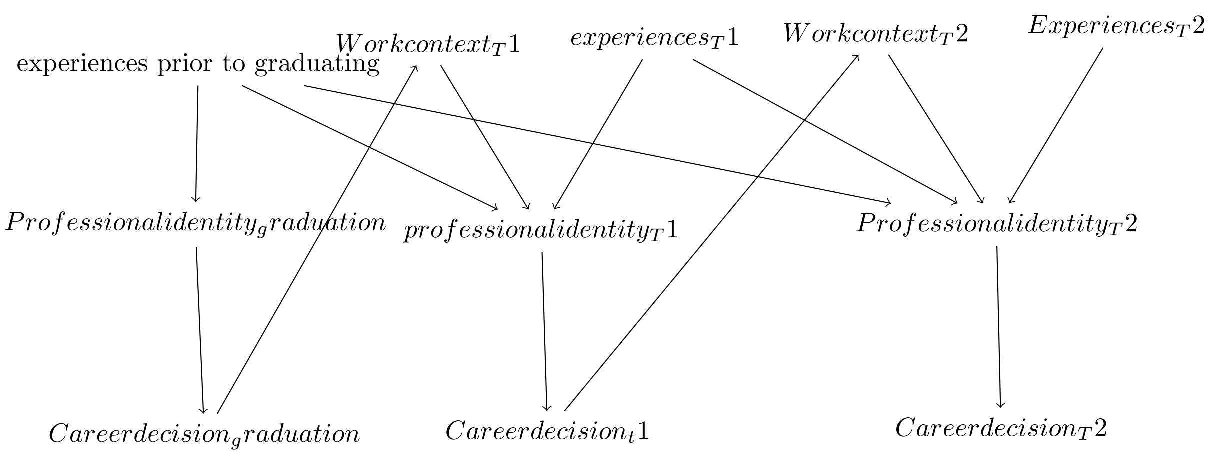
2.6 Research questions

1) How can individual teacher carreer choices be segmented to form distinct types of teacher carreer paths?

2) How are teacher characteristics associated with distinct types of teacher carreer paths?

3 Theory

Some ChatGPT and Powerdrill used here



1200 words

Why do teachers choose to work at the school at which they are employed? In this paper i will argue that this is essentially a question of teacher choosing to be employed at the schools that whey *want* and *can* be employed at, and that these two factors do not necessarily coincide. In other words, the career choices of teachers can be explained by the *preferences* of teacher and the *circumstances* at which they currently find themselves in. For instance, teachers may prefer to work at schools with high achieving students(want) but they are currently living in an area in which only one school has such a teacher composition and that school turned down their last application(can't). Thus the core of this section serves to build a theoretical framework that can explain how preferences of teachers are formed and how these preferences interact with current circumstances to shape the career choices of teachers

Preferences and circumstances are dynamic. While preferences may in large part be formed by, more or less, stable teacher characteristics, e.g. teacher skills, personality, professional, moral or cultural values, identity, etc., they are likely also influenced by experiences at the current or earlier schools of teachers, e.g. teachers figuring out that they have a strong preference for working with other motivated teachers or that lack of administrative support can overshadow other positive characteristics of schools. In other words, as teachers gain experience their preferences may develop. Circumstances may also be dynamic in the sense that schools can close, or new schools can be built in the neighboring areas, gentrification may change student compositions or teachers may have crossed a certain threshold of experience that allows them apply for jobs that are more aligned with their preferences.

When *want* and *can* do not coincide teachers leave their current school and search for employment at another school or find other job opportunities that can satisfy their needs. This is what creates the dynamic of teacher careers.

3.1 What shapes the preferences of teachers?

The professional identity of teachers is likely form the backbone of the preferences that teachers hold towards the kind of school they want to work in, or whether they'd prefer to opt out of teaching altogether. The professional identity of teachers comprise their aspirations, values as well as how they explain their decisions and actions(Forde and McMahon Citation2019). Previous studies highlight the importance of teachers' understanding of themselves and the way that teachers use their professional identity to make sense of themselves as teachers(REF).As such the professional identities of teachers likely guides the way that teachers make choices in terms of their careers, as well as how they relate to colleagues, students and their everyday practice as teachers.(REF)

Importantly, the professional identity of teachers are not immutable, but are the result of a dynamic process that constructs and alters teachers' professional identity during the course of their career(REF). Personal experience, professional practice and the external environment all interact in the shaping of teachers' professional identities. While some constituents of teachers' professional identity, e.g. experiences during childhood and adolescence,moral values, experiences during education and especially college education(Rinke), are strong and lasting components of teachers' professional identity, other components e.g professional practice, interactions with other teachers and students in different schooling context, are likely to contribute to the dynamic nature of teachers' professional identity. "external" factors, e.g. family formation, sickness, are also likely contributors to teachers' professional identity outside their experiences in schools.(REF, Tonna & Calleja)

As such, teachers' career choices can be seen as the *enactment* of their professional identity.

Professional identity, keeps teachers from opting out

-Identity(Already elsewhere e A study of (skilled) teachers' choice to leave teaching)

-life circumstances

Job satisfaction(autonomy, competency and coherence)

perception of self, do teachers perceive themselves to be skilled?

self-efficacy?

teacher biographies, previous experiences with school system

3.1.1 How does the professional identity of teachers vary across teachers with different characteristics?

How is the professional identity of teachers reflected in various, measurable, characteristics of teachers, e.g. their cognitive ability, sex, social background, geographic origin, ethnic origin etc.?

3.2 What makes some schools more attractive than others?

The attractiveness of some schools, as opposed to others, is not arbitrary. This is higlighted in the work of (**Ingersoll_2001?**) who points to the organizational features of schools as prominent "push-and-pull" factors for teachers.

The organizational features cover characteristics such as employee compensation structure, administrative support, degree of conflict, and strife, and employee influence over the organization policies(**Ingersoll_2001?**). Ingersoll's central premise is that these factors impact, e.g., the cohesion among teachers and professional motivation and thus influence how likely teachers are to stay at or leave schools. Further, Ingersoll argues that schools faced with external challenges, such as difficulties in hiring teachers due to being located in a deprived neighborhood, are forced to decrease organizational standards. For instance, principals facing trouble attracting teachers specialized in teaching biology may assign teachers with no specialization in biology to teach biology classes

Such organizational features are also present in the work of Rinke, often conceptualized as "extrinsic factors". Rinke's study highlights that teachers actively evaluate their professional direction based on their identities and the alignment of their values with their work environment. For instance, teachers who feel a strong emotional connection to their workplace are more likely to remain in the profession, while those whose experiences contradict their ethical beliefs may choose to leave(REF)

3.3 Why do teachers change from one school to another?

The decision to leave one school in favor of another comes down to an interplay between between life experiences and school context

Working conditions and life experiences interact to form the picture of what success looks like to the teacher, and thus ultimately decides the career choice of the teacher(Rinke, why do half leave). This choice, life experiences and previous work-experiences contribute to the ongoing construction of teachers professional identities, which they will draw on to make future career decisions. In the framework of Rinke, this interaction is depicted as a process of negotiation, in which experiences in the workplace context are "filtered" by teachers' life experiences. Past experiences, e.g. personal and moral values, formative experiences during childhood and adolescence and experiences throughout education and former workplaces, work as filter that aid the interpretation and management of experiences and challenges that teachers are exposed to in schools. It is in this negotiation between life experiences and experiences at the current workplace that teachers construct their perception of success(Rinke, why do half leave)

As such, the decision to leave the current school in favor of another school occurs when the current perception of success does not align with current work situation. One may imagine that a teacher aspires to make a difference for underprivileged students, but is currently employed in a school of relatively affluent students. In such a situation, the current situation is not well aligned with the teachers' current perception of success. In another situation one may imagine a teacher with a background in STEM who has trouble sustaining themselves economically on a teachers salary. At the same time, the teacher is disillusioned by feeling unprepared to teach primary school students, and thus finds it very difficult to make the positive change the teacher had hoped to make when beginning a career as a teacher. In such a situation, the current perception of success would likely be to find employment in STEM for a higher salary and thus fully leave teaching altogether

personal development

life circumstances(marital status, children, moving)

3.4 Why do teachers opt out of teaching?

Disillusionment("what does it even matter"),

burnout

discrepancy between expectations and reality

better "outside options"(e.g. physics teachers who can work in a lab)(Rinke, pusn and pull p. 372)

3.5 How do teacher characteristics impact teacher career choices?

Teachers with better credentials and more experienced teachers are more prone to get a job at schools with academically high-performing students

3.6 How does the theory predict that teachers will shape their carreer paths?

Aplying the theoretical framework of Rinke, leads to a range of predicted findings

- Career paths of teachers will vary between teachers due to the differences between teachers with respect to professional identity
- Career paths of teachers will exhibit dynamic patterns, i.e. teachers will switch between schools, different types of schools, e.g. between schools serving different types of student populations, or opt in and out of teaching. This is to be expected due to the dynamic process that shapes teachers' professional identity throughout their career. The propensity to switch between schools, types of schools or opting in and out of teaching is also expected to vary between teachers
- Teacher career paths will not be completely heterogenous between teachers. Teachers who share experiences, or who have had similar experiences, e.g. having had similar types of childhoods, or similar experiences within the educational system or sharing characteristics such as gender, ethnic origin or geographic origin, will likely also share aspects of their professional identity. As such, it is to be expected that teachers with similar experiences or characteristics will have similar career paths, and that these similarities in career paths can be used to construct a typology of teacher career paths.

- Measurable characteristics of teachers will be associated with different types of career paths. While past experiences and personal characteristics feature prominently in the work of Rinke, there are no precise and coherent descriptions of how certain experiences or characteristics would affect teacher career paths. Parts of her work however does touch upon certain experiences and characteristics, chiefly the educational background and educational performance of teachers.

  - Teachers' educational performance: In the work of Rinke, teachers' with good educational performance, or better/special credentials, may show higher aspirations in terms of career. In particular, such teachers may have more opportunities for employment outside teaching, and will thus be more inclined to leave teaching(REF). As such, teachers with more opportunities outside teaching may also react more strongly to dissatisfactory work environments, and may be more prone to switch to schools with better work environments or with class environments in which they can navigate more easily. As such, teachers who are academically gifted or who have attained better credentials may be more prone to switch to schools with better resources or schools who serve more affluent student populations, if they do not opt out of teaching entirely(REF). Additionally, the comparatively low perceived status of teaching may also push the most academically able teachers out of teaching. Even if the teachers themselves may prefer teaching due to, e.g. moral values, teachers may not find support in their decision, or even face resistance, from their close ties who may purport the idea that academic ability is "wasted" on a career in teaching(REF, why half of teachers)
  - Gender: The teachers' gender likely has an impact on the career decision of teachers. Given that gender has likely been an influence on previous life experiences and is a part of the identity of teachers prior to their work as teachers(REF) it is therefore also likely that teachers' gender will exert an impact on teachers' professional identity. The likely impact of gender on teachers' career decisions is not clear however. Women face challenges in a number of profession due to gendered norms surrounding work, in which they are expected to e.g. bear and care for children and bear the brunt of unpaid labor in the home(REF). These gendered norms likely hold women back from both participation in the labor market and professional advancement(REF). With respect to gendered norms in teaching, these may work against the decision of men to either chose or stay in teaching, as teaching is traditionally seen as a profession associated with women. In that sense, male teachers may feel a cultural deterrence to chose or stay in teaching, as they are expected to chose a career that is more aligned to classical views of male occupations(REF, why half of teachers leave, Why Not Become a Police Officer?: Challenges in the Recruitment and Retention of Men in Early Childhood Education). The participation of women in the danish labor market rose sharply in the 1960's and 1970's, aided in large part by maternal leave benefits and universal daycare(REF). While the institutional context certainly doesn't nullify the aforementioned challenges women face in the workplace, i will argue that these challenges are currently present to a much lesser extent in the teaching profession than in other professions that typically require very long working hours for career advancement, or professions in fields that are still heavily dominated by men, e.g. executive positions in private companies or positions in manual labor. To the extent that teaching in primary schools have become a profession in which women make up the majority, gendered norms may work more strongly to deter men from entering or staying in teaching than challenges traditionally faced by women deters women from entering or staying in teaching.

  - Geographic origin: While the work of Rinke does not mention how geographic origin specifically impacts teachers' career decisions, it is mentioned as a part of the past experiences that shape teachers' professional identity(REF). Previous studies have shown that teachers prefer to teach in close proximity to the geographic origin. As such, teachers' originating from, e.g. rural areas are likely more prone to find employment in the types of schools found in such areas. Given that individuals with either high amounts of economic resources or those with college degrees tend to live in, or around, metropolitan areas(REF), teachers' from rural origins are likely less prone to be teaching in schools serving students with a high-SES background than teachers' from more urban or sub-urban origins.
  - Socio-economic background: The socio-economic background of teachers may also impact teachers' career decisions. Specifically, teachers from affluent, and especially, highly educated homes may be perceived as engaging in "downward mobility" and "wasting" their opportunities(Dworkin, A. G. (1980). The changing demography of public schools teachers: Some implications for faculty turnover in urban areas.; Why half of teachers leave, John Rury (1989), ref in "why half leave"). Conversely, teachers from homes in which parents hold no college degrees may perceive the attainment of a position as a teacher as a perpetual success, in the framework of Rinke. For such teachers, current workplace situations may not fulfill their wishes, but remaining in teaching may be perceived as favorable compared to opting out, regardless of the current work environment. Teachers from more affluent backgrounds may however be more prone to perceive opting out of teaching to pursue other occupations as more aligned with their perception of success.

- Age: While most teachers' are similarly aged when they graduate, other teacher graduates may have opted into teaching from different occupations or educational tracks, and will likely be older than their peers. Such teachers may have a more developed perception of success with regards to career than their peers, due previous experiences in e.g. the labor market or other educational tracks. As a result, these teachers may be more determined and motivated to make a career out of teaching than their younger peers, whose first experiences in the labor market will be as teachers(REF).
- Marital status and children: Marriage and children are both events in teachers' personal lives that likely lead to a wish for more stability with regards to career. Married couples are likely to live, or wish to live together, which means that decisions with regards to residential location is now a joint decision of married couples. This likely restricts the geographic mobility of both individuals in the couple, which in turn restricts the choices of occupation in terms of geography. This restriction is further exacerbated by couples who are homeowners, as this adds additional economic restrictions, such as mortgage payments. Having children entails many of the same restrictions and demands for stability, with respect to residential location. In addition, caring for children adds additional constraints on parents' time and economic resources. Additionally, marriage and family formation likely adds a dimension of fulfillment in teachers' lives, which likely impacts teachers' perception of success in their careers, e.g. teachers' may wish to balance career decisions such that future career decisions will leads to improvement in terms of work context but do not negatively impact either teachers' marriage or family e.g. by having to relocate. As such, marriage and having children likely leads to more stable career patterns.(REF)

-Some will drop out, perhaps those that are the academically most able due to better outside-options

-Some will prefer schools and can find jobs

- Variation across career paths is expected due to the dynamic process of constructing teachers' professional identity.

-Due to differences in identity between teachers, teachers will seek different career paths. Schools with different characteristics will appeal to teachers' with different types of professional identities. Teachers with similar types of professional identities will therefore attempt to find employment in the same type of schools, and will therefore have similar types of career paths when career paths are defined on the basis of different types of schools.

- While Rinke stresses that school characteristics work as either push- or pull-factors in conjunction with intrinsic factors of teachers, e.g. personal values, beliefs and experiences, some school characteristics, e.g. student discipline problems, do seem to be a prominent push-factor, on average. As such, while differences between teachers, with regards to teachers' professional identity, may make some teachers more resilient towards student discipline problems, teachers do seem, on average, to prefer schools with lower amounts of student discipline problems.

-Measurable teacher characteristics, e.g. teachers' sex, geographical origin, academic ability and social background are tied are expressions of and influences on teachers' professional identity will be predictive of teacher career paths

- Teachers own experience within the educational system shapes how filter workplace experiences when teaching(Rinke, the apprenticeship. . . .)

4 Data

700 words

In this paper i use Danish administrative data covering two populations: All danish primary school teachers in the years 1980-2020 and all danish primary school students, and their parents, in the years 1980-2020.

In this paper the population of teachers is further restricted. The population of danish primary school teachers is restricted to only include teachers that graduated with a college degree in teaching, in the years 1980-1985. There two reasons for this restriction. Firstly, i chose to limit the years of graduation to the years 1980-1985, as this would allow me to investigate teacher careers over the course of 35 years, while also yielding a population of considerable size as well as also giving me the opportunity to investigate how teacher career paths vary between cohorts of teacher college graduates. The population is secondly restricted to only include teachers that graduated with a college degree in teaching. Career paths for teachers with a college degree in teaching has a natural starting point, which is the year that these teachers attained their degree in teaching. For teachers that have not attained traditional qualifications as a teacher through a college degree, it is very hard to define a natural starting point. While one possible starting point might be the first job as a teacher, it is very hard to precisely pinpoint when individuals were working as primary school teachers for the first time without having the full labor market history of individuals. As data on labor market participation does not cover years earlier than 1980, it is very hard to precisely measure if individuals entered the labor market in 1980-1985, as the full labor market history of most individuals in the labor market in 1980-1985 likely begins prior to 1980. I therefore chose to limit the population to teachers that had attained a college degree in teaching in order to work with a population that is well defined and comprehensively covered in danish administrative data. In sum, *the population of teachers in this paper cover teachers that graduated with a college degree in teaching in the years 1980-1985.* The starting point of the career sequences of teachers are defined by the year of graduation, and sequences extend to 35 years after graduation.

The population of students in primary schools in 1980-2020, and their parents, is not restricted in any way. The data on students and their parents primarily serve to construct aggregated measures of students' SES background and students' educational attainment following primary school. Students are directly linked to schools in the danish administrative data.

4.1 Linking teachers and schools

While the link between teachers and their workplaces is obtained directly from Statistics Denmark, the link between workplaces and primary schools needs to be constructed in order to link teachers' to schools in which they are employed. To do this, public administrative data on all educational institutions(https://data.stil.dk/instreghistorik/) is linked to administrative data concerning teachers' workplace via the "unit of production" identifier. The result is a combined dataset, in which teachers are linked to the schools in which they are employed, on a yearly basis, through the link between teachers and their workplaces. The link between teachers' workplaces and primary schools is described in greater detail in appendix XX

Danish register data

All teachers that graduated between 1980 and 1985. Teachers are defined as individuals who have completed a teacher education at the college level. There are individuals who work as teachers without having graduated from a teacher college, however these teachers are not included as sequences are defined on the basis of years since graduation from a teacher college.

All public primary schools

All students, their parents and siblings for aggregate school characteristics

Teachers linked to schools on a yearly basis

4.2 Teachers' career sequences

In this paper teachers career paths are comprised of sequences of career decisions, that are recorded on a yearly basis. Teachers career decisions are categorized into two primary categories: Working in different types of public primary schools or not working in public primary schools. These two primary categories are divided into different sub-categories, which are detailed below

4.2.1 Types of public primary schools

In this paper i will use the following school characteristics to construct a typology of schools.

- Average education of parents, at school level, measured in months
- Average income of parents, at school level
- Average high school enrollment rate: The proportion of students graduating in year t that are enrolled in high-school in year $t+1$

All school characteristics are time-varying, i.e. averages have been computed for each school s in each year t

Parents educational level and income provide measurements of the average SES background of students at each school. The proportion of students enrolled in high school in the following provide an indirect measurement of student academic ability and academic aspirations. In the literature on teacher retention and attrition, both student SES background and academic ability prominently feature as factors that impact teachers' propensity to stay at or leave their current place of employment(REF). For this reason, i expect that teachers will transition between schools vary on the chosen characteristics.

4.2.2 Constructing school types

In order to construct the different types of schools that teachers can be employed at, i aggregate the characteristics of each school in a linear scale, in which i scale and take the rowsum of the characteristics of each school, i.e.

$$School\ characteristics\ scale_{st} = \sum_{k=1}^{k} school\ characteristics_{kst}$$

All school characteristics are scaled within each year t, i.e. for each school characteristic, each value is scaled by the mean and standard deviation of the school characteristic within the current year instead of the mean and standard deviation across the entire sample. This is to avoid the scaling being dependent on trends in time that may be present for each of the characteristics. For instance, parental educational level increases quite dramatically from 1980-2020. On average, the average length of mothers education increased by nearly 3 years. As consequence, there would be very few schools in 1980 with a positive z-score on length of mothers education.

While this approach is simple compared to other approaches of constructing such a scale, e.g. factor analysis or principal components analysis, performing a principal components analysis on the school characteristics suggests that the variables are correlated in highly unidimensional manner. In table 1 we see that the first principal component accounts for 72% of the variance among the chosen school characteristics. As such, expanding the scale to include additional principal components would likely substantially increase complexity in order to account for a relatively small amount of variance.

Table 1: Summary of principal components analysis of school characteristics

	PC1	PC2	PC3	PC4	PC5
Standard deviation	1.90	0.96	0.54	0.34	0.25
Proportion of Variance	0.72	0.18	0.06	0.02	0.01
Cumulative Proportion	0.72	0.91	0.96	0.99	1.00

After having computed the scale-scores for each school, i then divide schools into 4 groups.

1: Schools that among the 0th and 25th percentile of the distribution of the “school characteristics scale”

2: Schools that above the 25th percentile and at maximum the 50th percentile of the distribution of the “school characteristics scale”

3: Schools that above the 50th percentile and at maximum the 75th percentile of the distribution of the “school characteristics scale”

4: Schools that above the 75th percentile and at maximum the 100th percentile of the distribution of the “school characteristics scale”

4.2.3 Career decisions outside public primary schools

Outside public primary schools teacher can make X different career decision:

1: Work in private primary schools

2: Pursue post-graduate education

3: Work in other education

4: Work outside education

For detailed descriptions of how each of these categories are defined with respect to the administrative data, readers are referred to appendix XX

4.3 Teacher characteristics

In addition to investigating the distribution of teachers’ career sequences, i will also investigate whether characteristics of teachers are associated with the different types of teacher career sequences.

Following the theoretical framework i will include the following teacher characteristics as predictors of membership in types of teacher career sequences.

Teachers’ gender: A binary variable indicating whether teachers are registered as female or male in the danish administrative data

Teachers’ high-school GPA: A continuous variable measuring the overall high-school GPA of teachers. While it would have been more relevant to include the teachers’ college GPA, teachers’ college GPA is covered in the danish administrative data earlier than 2004.

Number of children in household: A continuous, integer-valued, variable measuring the number of children living in the household of the teacher

Teachers’ marital status: A discrete variable indicating whether teachers are unmarried, married or divorced.

Type of municipality: A discrete variable indicating whether the teacher lived in one of X types of municipality:. . . .

Age at graduation: A continuous, integer-valued, variable measuring the age of teachers, in years, at the time teachers attained their college degree in teaching

Year of graduation: A discrete variable indicating in which year the teacher graduated all covariates are measured at the time of graduation, and are thus time-invariant. As such, teacher marital status may change during the course of teachers' careers, but in this paper teachers' marital status indicates the marital status of teachers in the same year that they attained their college degree in teaching. This choice may seem overly restrictive, but it is not trivial to include time-varying covariates in sequence analysis(REF)

5 Methods

1000 words

5.1 What are we trying to look at with sequence analysis?

E.g. "what is the estimand?"

We are interested in distribution(s) of teacher career trajectories

5.1.1 What is a teacher career trajectory?

A sequence of career choices for each teacher e.g.:

low-SES pub school -> medium SES pub school -> not teacher -> high SES pub school

5.1.2 What is a distribution of teacher career trajectory?

a description of multiple teacher career trajectories

eg.

Trajectory 1: low-SES pub school -> medium SES pub school -> not teacher -> high SES pub school

Trajectory 2: low-SES pub school -> medium SES pub school -> high SES pub school -> high SES pub school

in period 1, 100% is low-SES pub school, in period 2 100% is medium SES pub school, in period 3 50% is not teacher and 50% is high SES pub school

5.2 Sequence analysis

In this paper i will investigate sequences of teacher career decisions using sequence analysis. Sequence analysis aims to apply a more holistic approach to e.g. analysis of career decisions or transitions between life course events such as marriage, family formation etc., as opposed to e.g. regression approaches that aim to model the average probability of transition between one state to another. As such, sequence analysis is also a more descriptive technique(REF). In this paper i will primarily use sequence analysis for three purposes.

1) Investigating the overall distribution of teacher carer sequences

2) Constructing a typology of teacher careers by applying optimal matching and cluster analysis to teacher sequences

3) Investigating associations between teacher characteristics and membership of teacher career types

I will use the R package “TraMineR”(REF) to carry out much of the analysis

5.2.1 Distribution of sequences

In this paper sequences of teacher career decisions are constructed as consecutive teacher career, during the years 1-35 since graduation, decisions within the state space defined by the data. For instance a sequence of career decision in the first 5 years since graduation may look like *school type 1->school type 1->school type 1->private school->private school* for one teacher and may look like *school type 1->school type 2->school type 2->school type 4->school type 4* for another teacher. Taken together across all teachers we get a distribution of sequences, where we can asses the share of teachers in each type of career decision across years since graduation, e.g. in year 2 since graduation, 10% of teachers may be employed in school type 1 and 8% are employed in private schools, while in year 3 since graduation 7% of teachers may be employed in school type 1 and 11% are employed in private schools.

Short intro to sequence analysis

Mostly descriptive

What is a sequence

will use the R package TraMineR(REF)

#How will sequences be clustered?

After having constructed the sequences, i will investigate if similar sequences can be clustered together to form different types of career sequences. To do this i will:

1) Using optimal matching, measure the similarity between sequences of teacher career sequences

2) Use hierarchical clustering to construct types of career sequences. Additionally i will use range of statistics describing the “fit” of the clustering procedure when the number of clusters vary. This is a data-driven procedure that aims to aid in deciding which number of clusters will lead to the “optimal” clustering of similar sequences of career decisions.

5.3 sequence similarity

The similarity of sequences is measured using optimal matching(REF). Optimal matching is a which attempts to align sequences such that they are identical using 3 different operations: Substitution, insertion and deletion. For each of these operations a “cost” assigned, which will ultimately determine how similar sequences are measured to be, following the optimal matching procedure. As an example we can see optimal matching would determine the similarity of the two previous sequences.

Sequence 1: school type 1 -> school type 1 -> school type 1 -> private school -> private school

Sequence 2: school type 1 -> school type 2 -> school type 2 -> school type 4 -> school type 4

In this case 4 operations would be needed to align the sequences:

1: substitute the second element in sequence 1, school type 1, for school type 2

2: substitute the third element in sequence 1, school type 1, for school type 2 '

3: substitute the fourth element in sequence 1, private school, for school type 4

4: substitute the fifth element in sequence 1, private school, for school type 4

As such, the distance between sequence 1 and 2 is four operations according to OM

In the example above either of the three operations had the same "cost", which was 1. However insertions and deletions can be more or less costly than substitutions and not all substitutions need to have the same cost. For instance substituting school type 4 for school type 1 can be more costly than substituting school type 1 for school type 4. The costs for for substitutions,deletions and insertions can be set by the user following e.g. theoretical considerations. In this paper however, i have chosen a data-driven strategy for setting these costs, by relying on the inverse of the observed transition rates between states. As such, the substitution costs between states is given by

$SC(i, j) = cval - P(i|j) - P(j|i)$

Where i and j represent states, while $P(i|j)$ and $P(j|i)$ is the probability of transitioning from state i to j and from state j to i respectively. $cval$ is a constant with the value of 2

In table XX the observed transition probabilities between states is shown, and in table XX the cost associated with each type of substitution is shown

Optimal matching

empirically derived transition- and substition costs

5.4 Hierarchical clustering

ChatGPT used here

Hierarchical clustering is a method used to group similar sequences into clusters based on their dissimilarity. It builds a hierarchy or a tree of clusters, known as a dendrogram, which can be cut at different levels to form various numbers of clusters. One commonly used method for hierarchical clustering is Ward linkage, which minimizes the variance within each cluster.

Briefly, hierarchical clustering iterates by merging observations into clusters until all observations belong to one single cluster. The process is depicted in figure XX.

Clustering sequences using Ward linkage implies that, at each step the total within cluster variance is computed between each cluster, is an agglomerative hierarchical clustering technique that minimizes the total within-cluster variance. It does so by selecting the pair of clusters to merge at each step that leads to the smallest possible increase in the sum of squared differences within all clusters. How it Works: At each step of the algorithm, Ward's method looks at all possible pairs of clusters and calculates how much the within-cluster variance would increase if those two clusters were merged. It then merges the two clusters that result in the smallest increase in variance.

$$\Delta(A,B) = \frac{n_A n_B}{n_A + n_B} ||\vec{m}_A - \vec{m}_B||^2 \tag{1}$$

Once the merger resulting in the smallest within-cluster variance has been identified, the distances between clusters are "updated" in the following manner:

Source Wikipedia

$$d_{PC}^2 = \frac{n_A + n_P}{n_C + n_P} d_{PA}^2 + \frac{n_B + n_P}{n_C + n_P} d_{PB}^2 - \frac{n_P}{n_C + n_P} d_{AB}^2 \tag{2}$$

where n_A and n_B denote the number of sequences in clusters A and B, which are the clusters that will be merged, n_C denotes the number of sequences in the cluster resulting from the merge of clusters A and B and n_P is the number of sequences in cluster P, which is a cluster that will not be included in the new cluster C

To use figure XX as an example, we see that in step 1, the sequences A and B are merged into one cluster. In such a case, the distances to sequences C, D and E are updated following 2, prior to merging sequences A and B. As such, in step 2, the clustering process will have access to the distances between the cluster AB and the sequences C, D and E, rather than the distances between each of the individual sequences.

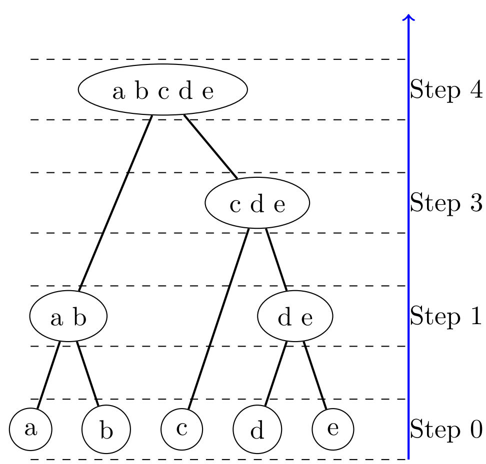


Figure 3: adapted from http://www.sthda.com/sthda/RDoc/images/hierarchical-clustering-agnes-diana.png

5.4.1 Determining optimal number of clusters

As shown in figure XX, the clustering process continues until all individual sequences are clustered into a single cluster, but it does not suggest at which step, the optimal clustering of sequences has been achieved. While, there is no clear consensus on how one decides which number of clusters will be "optimal" there are several measures that can aid the decision. In this paper i have chosen to apply the average silhouette width(ASW) as a measure of fit across solutions with different numbers of clusters, as this measure has several favorable properties(REF). The silhouette width of each sequence can be though of a measure of well each sequence fits into the cluster to which it is assigned. Positive values of silhouette width indicates that a given sequence fits better into the cluster to which it is assigned than to the nearest "neighboring" cluster, i.e. the cluster in which the sequence would have the second best fit. Conversely, a negative value silhouette width indicates that the given sequence would actually fit better in the nearest neighboring cluster than the one to which it is currently assigned(REF).

#How will sequence analysis be applied to current data?

5.5 Predicting cluster membership using teacher characteristics

Regressing teacher characteristics on cluster membership using multinomial logistic regression

In this paper i will investigate the association between a range of teacher characteristics and the types of career paths that teachers are assigned to via the clustering procedure. In order to account for unspecified non-linear functional forms, interactions between predictors or other deviations constraints of classical regression frameworks, e.g. linearity in parameters, i will apply flexible machine learning(ML) techniques. Many of these techniques are known to "learn" the functional form of associations between outcomes and predictors from data. Additionally tree-based techniques, such as decision trees, random forests and gradient boosted trees, are known to also learn unspecified interactions between predictors. I will also compare the predictive performance of the ML techniques with the performance of multinomial regression, which is the traditional estimation technique used to investigate relations between predictors and typologies of sequences(REF)

I will use the teacher characteristics detailed in section X as predictors of membership to types of career sequences

I will use the following approaches to investigate the associations between teacher characteristics and types of career sequences:

ChatGPT used here

- Multinomial logistic regression, implemented via the function "multinom" from the package "nnet" in R

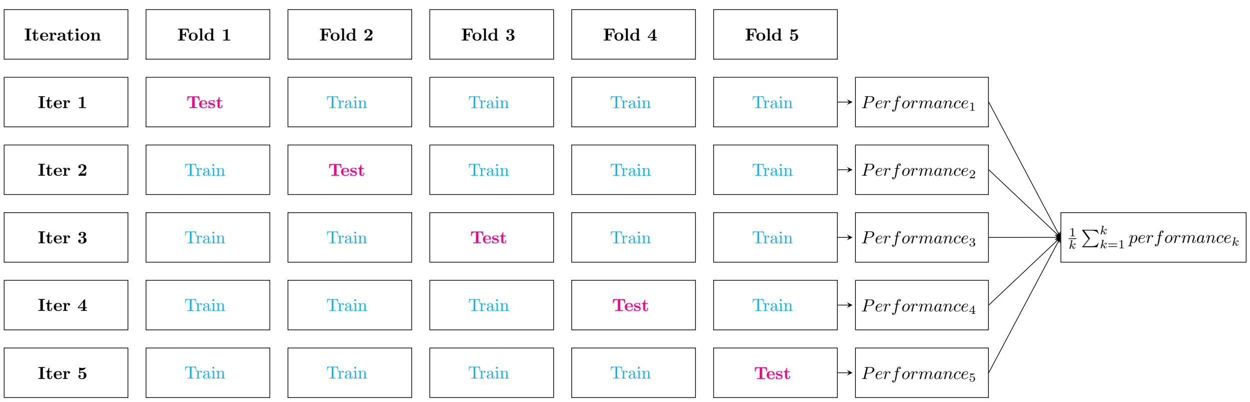
- Elastic net regression: Elastic net regression is a type of penalized regression, where regression coefficients are "penalized" according to an amount specified by the user. Elastic net combines both L1- and L2-penalization according to a "mixing weight" which controls the proportion of penalization that is attributed to either L1 or L2 penalization. Elastic Net can handle scenarios with highly correlated features and can perform better than linear regression by reducing overfitting and improving generalization. Elastic net is implemented via the "glmnet" package in R

- Decision Tree: A decision tree is a flowchart-like model that splits data into subsets based on variable values, leading to a tree structure. RPART (Recursive Partitioning and Regression Trees) is an implementation in R. Decision trees can capture non-linear relationships and interactions between features, regardless of whether these are specified or not. While this ability of decision trees is not guaranteed to automatically find the true functional form of relationships or all interactions between variables, this ability may still lead to improvements over multinomial regression(REF).Decision trees are implmented via the "rpart" package in R

- Conditional Inference Tree (ctree): A conditional inference tree is a type of decision tree that uses statistical tests to select splits rather than greedy algorithms. It controls for overfitting by stopping the tree when splits are not statistically significant. By focusing on statistically significant splits, ctrees reduce the risk of overfitting and can model complex interactions. This may be an improvement over traditional implementations of decision trees, which are prone to overfitting(REF)
- Support Vector Machine (SVM): SVM is a classification and regression algorithm that finds the hyperplane that best separates data points into different classes. It can handle both linear and non-linear data using kernel functions. SVM is particularly effective in high-dimensional spaces and can capture complex, non-linear relationships, offering better performance in cases where linear regression would be insufficient. SVM has proven highly effective in other practical applications(REF)
- K-Nearest Neighbors (KNN): KNN is a simple, non-parametric algorithm that predicts the outcome by averaging the values of the k nearest data points in the feature space. KNN can model non-linear relationships without making assumptions about the data distribution, potentially outperforming multinomial regression, especially in cases with complex patterns.
- Multivariate Adaptive Regression Splines (earth): Known as MARS, this method builds flexible models by creating piecewise linear regressions over different intervals of the data. Similar to decision trees, MARS has the ability to capture unspecified interactions between variables(REF). MARS is a regression approach at its core however, and may therefore be better suited to model continous predictors than decision trees.
- Gradient Boosted Trees (GBM, H2O): GBM is an ensemble technique that builds multiple decision trees sequentially, each trying to correct the errors of the previous ones(REF) GBM can model intricate patterns and interactions in the data by combining the strengths of multiple weak learners (decision trees). As such, gradient boosted trees improve upon approaches that use a single decision tree.
- Extreme Gradient Boosting (XGBoost): XGBoost is an optimized implementation of gradient boosting that is highly efficient, flexible, and portable, often used in machine learning competitions. XGboost is an extension upon GBM, which is intitially due it's optimized implementation, which reduces computing time. However XGboost also features other extensions, e.g. additional hyperparameters, that may lead to better predictive performance.
- Random Forest: Random Forest is an ensemble learning method that builds multiple decision trees and combines their predictions to improve accuracy and control overfitting. By averaging the predictions from various decision trees, Random Forest reduces variance, can handle large datasets with higher dimensionality, and can capture non-linear relationships. While random forests are initially an improvement upon single decision trees, random forests have shown to deliver good predictive performance even when compared with more advanced machine learning techniques, such as GBM and XGboost.

The list of machine learning techniques i have chosen here are often among the techniques that deliver the best predictive performance when dealing with ordered and tablular data, as is the case here. Conversely, while neural networks have shown to outperform all the techniques listed above in the presence of unordered data, e.g. text or images, neural networks often show, at best, a predictive performance that is comparable to the techniques presented above. As such, i have chosen not investigate whether neural networks may provide better predictive performance than multinomial regression, as these types of machine learning techniques typically require a large amount of computing time.

In order to asses the predictive performance and find the optimal combination of hyperparamter configurations for each ML technique i apply repeated K-fold cross-validation. Briefly, cross-validation involves splitting data into K partitions, usually denoted “folds”, where each partition divides the full data into “training” and “testing” parts. The model is then fitted on the training portion of the data and the fitted model is then used to predict the outcome in the test-portion of the data. Cross-validation, or similar procedures, are crucial when assessing the predictive performance of ML techniques. For many of these techniques, the flexibility and complexity can be substantial and the techniques may overfit to data as a result, which skews measures of predictive performance in a positive direction. As such, CV functions as a general and non-parametric way of giving an unbiased estimate of predictive performance(REF). In that sense CV also serves as a standardized way of comparing techniques for which the internal workings differ greatly.



Many of the ML techniques have several hyperparameters that impact the predictive performance, e.g. the degree of penalization of elastic net or the number of trees to aggregate in random forest. To pick the optimal combination of these hyperparameters for each technique, each technique will be fit, using CV, on a range of hyperparameter combinations. The hyperparameter combination with the best predictive performance will then be chosen. For models with few adjustable hyperparameters i will use a “grid search” in which all combinations of hyperparameter-values are assessed. For techniques with more adjustable hyperparameters, in this case GBM, XGBoost and random forest, i will pick 200 random hyperparameter combinations. Using random hyperparameter combinations is often a effective way to cover the search space of hyperparameter combinations, as opposed to doing a grid search which quickly results in thousands of hyperparameter combinations to asses in order to cover the hyperparameter search space adequately(REF)

To measure the predictive performance i will use Cohens Kappa, which is given by

In which p_0 is the proportion of agreements between predictions and observed values, while p_t is the hypothetical proportion of agreement between predictions and observed values purely by chance. The kappa value ranges between 0 and 1, where 0 indicates that predictions and observed values are in no better agreement than what could be expected by chance. I have chosen Cohens Kappa as a measure of predictive performance, given that this measure is robust to “class imbalance”, i.e. some categories of the outcome being much more prevalent than others. For instance, if the typology of career sequences is dominated by a single type in which 90% of teachers are assigned, then it would be easy to get 90% of predictions correct simply by predicting all teachers to be in the most prevalent type. The kappa value is robust to such a situation however(REF)

$$\kappa = \frac{p_0 - p_t}{1 - p_t} \tag{3}$$

Cross-validation and fitting of all ML techniques were implemented using the "caret" package in R

6 Imputing school characteristics

Data on the school characteristics that comprise the scale used to construct the types of schools are in many cases plagued by missing data. This complicates analysis, as school types then cannot be determined for some schools, in some years. To alleviate this issue, i opted to impute data on school characteristics using linear regression. Data is imputed by fitting a model in which the school characteristic in question is predicted by an indicator for year and school fixed effects, to account for variance between schools. To accommodate for a non-linear time trend, the relation between $year_s$ and $School\ characteristic_s$ is fitted using a 5th degree polynomial.

$School\ characteristic_s = \beta_0 + \beta_1 year_t + \beta_2 year_t^2 + \beta_3 year_t^3 + \beta_4 year_t^3 + \beta_5 year_t^4 + \beta_6 year_t^5 + \alpha_s$

7 Results

1400 words

7.1 Descriptive statistics

In the following i will present descriptive statistics for both teachers and the types of schools

7.1.1 Teachers

In table XX, descriptive statistics are presented for teachers in the first year since graduation. Statistics are presented across teacher "cohorts", which indicate the year at which each cohort graduated from teacher college. The majority of teachers are unmarried, female and do not have any children. Further, nearly all teachers are from the ethnic majority.

Characteristic	**Overall** N = 1,579[1]	**1980** N = 317[1]	**1981** N = 280[1]	**1982** N = 280[1]	**1983** N = 263[1]	**1984** N = 255[1]	**1985** N = 184[1]
age	29 (5)	28 (4)	29 (4)	29 (5)	29 (5)	29 (5)	29 (5)
unmarried	916 (58%)	177 (56%)	154 (55%)	157 (56%)	150 (57%)	156 (61%)	122 (66%)
female	931 (59%)	172 (54%)	164 (59%)	177 (63%)	144 (55%)	159 (62%)	115 (63%)

number of children in household							
0	240 (55%)	0 (NA%)	0 (NA%)	0 (NA%)	0 (NA%)	136 (53%)	104 (57%)
1	112 (26%)	0 (NA%)	0 (NA%)	0 (NA%)	0 (NA%)	59 (23%)	53 (29%)
2	66 (15%)	0 (NA%)	0 (NA%)	0 (NA%)	0 (NA%)	47 (18%)	19 (10%)
3	14 (3.2%)	0 (NA%)	0 (NA%)	0 (NA%)	0 (NA%)	9 (3.5%)	5 (2.7%)
4	5 (1.1%)	0 (NA%)	0 (NA%)	0 (NA%)	0 (NA%)	3 (1.2%)	2 (1.1%)
5	2 (0.5%)	0 (NA%)	0 (NA%)	0 (NA%)	0 (NA%)	1 (0.4%)	1 (0.5%)
Unknown	1,140	317	280	280	263	0	0
Ethnic minority status							
Born in Denmark	1,553 (98%)	313 (99%)	270 (96%)	274 (98%)	263 (100%)	252 (99%)	181 (98%)
Immigrant	19 (1.2%)	4 (1.3%)	6 (2.1%)	3 (1.1%)	0 (0%)	3 (1.2%)	3 (1.6%)
1st generation descendant of immigrants	7 (0.4%)	0 (0%)	4 (1.4%)	3 (1.1%)	0 (0%)	0 (0%)	0 (0%)

[1] Mean (SD); n (%)

7.1.2 Schools

In table XX i present descriptive results for the different types of schools. The average score of the school characteristics scale for each school type is reported in the bottom row. Here we can clearly see how school characteristics vary across school-types, with schools above the 75th percentile of the school characteristics scale having more than 20 months of additional education, compared to schools between the 0th and 25th percentile, for both mothers and fathers. The pattern is similar for the remaining school characteristics.

Characteristic	**0-25%** N = 6,043[1]	**>25%-50%** N = 6,308[1]	**>50%-75%** N = 6,468[1]	**>75%-100%** N = 5,975[1]
Average length of mothers education, in months	145 (14)	153 (12)	157 (12)	167 (13)
Average length of fathers education, in months	150 (9)	157 (8)	162 (8)	172 (10)
Average yearly disposable income of mothers in DKK	141,670 (64,940)	156,400 (62,192)	155,941 (68,106)	176,209 (90,303)

Average yearly disposable income of fathers in DKK	149,503 (71,036)	175,468 (70,830)	184,227 (79,791)	227,598 (121,897)
Proportion of students enrolled in secondary education the following year	0.18 (0.07)	0.22 (0.07)	0.26 (0.08)	0.36 (0.13)
Proportion of students who are immigrants or descendants of immigrants	0.15 (0.19)	0.08 (0.10)	0.06 (0.07)	0.04 (0.05)
Z-score of summed school characteristics	-4.3 (1.6)	-1.6 (0.6)	0.6 (0.7)	5.0 (3.0)

[1] Mean (SD)

7.2 Sequence analysis

In figure 4 we see the overall distribution of sequences for all teachers. While it is not surprising that the four different types of schools occupy roughly equal proportions of teacher career choices in each year, it is somewhat surprising that the proportion of teacher graduates that are not working in a public primary school is very high. While the proportions of teachers outside public primary schools continues to decrease until the very late stages of teacher careers, the proportion is consistently above 50% of teachers until the late stages of teacher career paths.

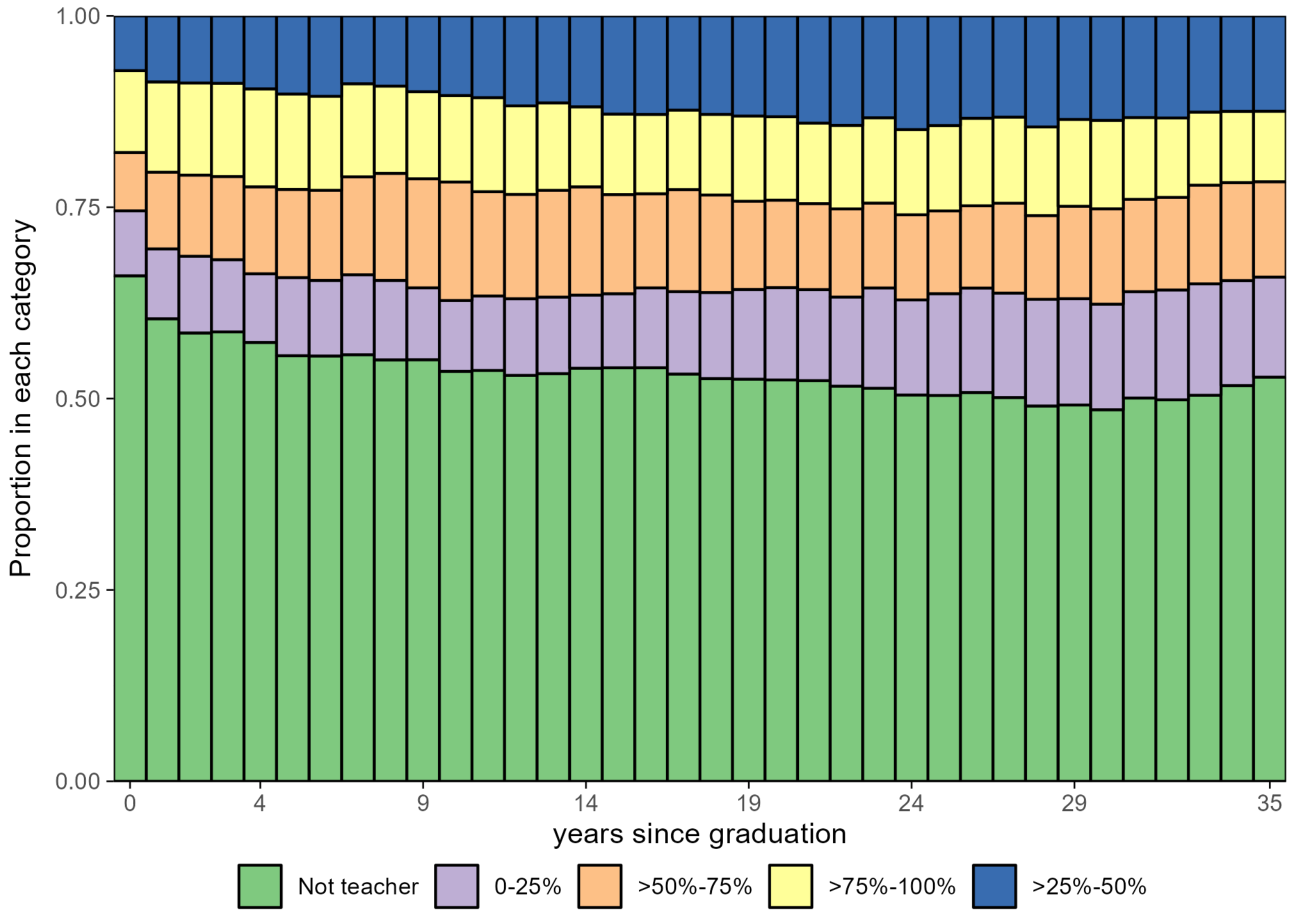


Figure 4: overall distribution of sequences

7.2.1 Clustering sequences

In the following section i will investigate how teacher career trajectories can be separated into different types of career trajectories using cluster analysis. To construct the clusters i compute the similarity between each sequence using optimal matching, using substition costs derived from observed transition rates between states. Next i use hierarchical clustering, specifically Wards method(REF), to construct the clusters.

In order to select the optimal number of clusters i use the "Average Silhouette Width"(ASW), which measures the coherence of the clusters. Higher scores on the ASW measure indicates that observations in each cluster are more similar to other observations in the same cluster, than they are to observations in the other clusters.

In table 4 we see that the ASW measure favors a solution with four clusters

In figure 5 we see the distributions of sequences across each of the four clusters. Following figure 4 it is perhaps not surprising to see that the decision to not work in a public primary school features prominently. However in figure 5 we see that the decision to not work in public primary school features heavily in a large group of teacher graduates who only rarely work in any type of public primary school, during their entire career. The two smallest clusters, cluster 2 and cluster 4 contrast each other, given that teachers in cluster 2

Table 4: optimal number of clusters across measures

	Optimal number of clusters for measure	Value
PBC	6	0.74
HG	6	0.90
HGSD	6	0.90
ASW	4	0.39
ASWw	4	0.39
CH	2	540.43
R2	7	0.46
CHsq	2	1006.42
R2sq	7	0.67
HC	6	0.05

almost exclusively work at schools in upper end of the distribution of the school characteristics scale, while the opposite is true of teacher in cluster 4. Interestingly, teachers in cluster 2 are more homogeneous in their choice of school, given that teachers in cluster 2 almost exclusively work in the most advantaged schools. While teachers in cluster 4, in contrast work in the least advantaged schools they also work in the schools between the >25th percentile and 50th percentile. The last cluster, cluster 3, is more balanced, although the majority of teachers work in the "middle range" of school types. Interestingly, while clusters 2,3 and 4 all suggest that a proportion of teachers start their career outside public primary schools, the group of teachers in cluster 2 is the only group in which nearly all teachers "opt in" to public primary schools and choose to stay. In clusters 3 and 4, there is a persistent portion of teachers that never "opt in" to public primary schools. Further, teachers in clusters 3 and 4 also start to opt out of public primary schools in the later stages of their career, with this pattern being more pronounced in cluster 3.

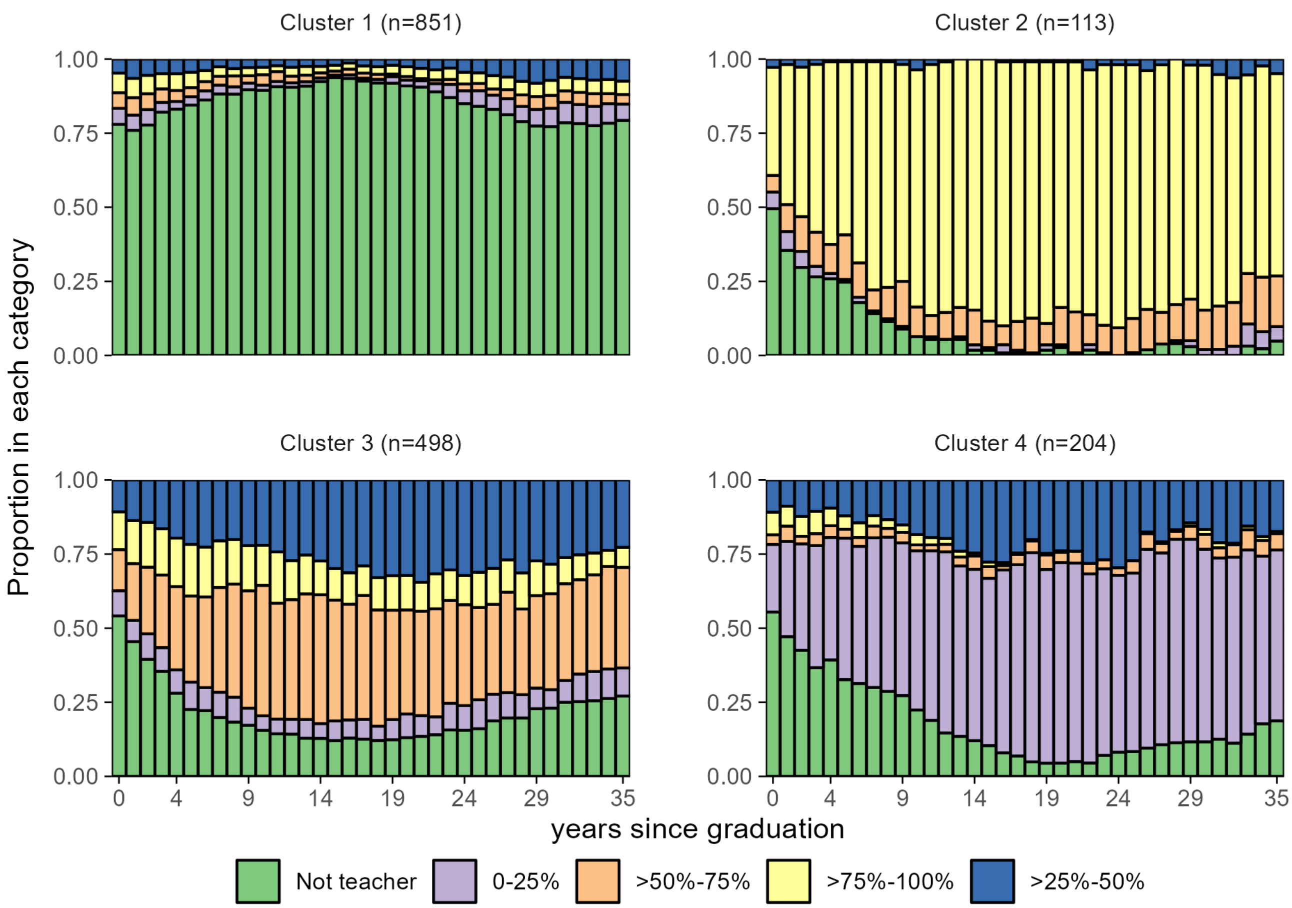


Figure 5: distribution of sequences across clusters

7.3 Predicting clusters

7.3.1 Performance of ML models

7.3.2 PDP plots

7.4 Robustness checks

- Missing data
  - unbalanced panel
  - attrition
  - missing school data
  - imputing school data
  - imputing sequences
- Confounders

- sensitivity -sensitivity to distance metrics -sensitivity to clustering method

8 Discussion

900 words

8.1 How might patterns of teacher career paths impact the distribution of teachers among high- and low-SES students?

Given the patterns of carreer paths, how would teachers be distributed among different types of schools at different stages of teacher careers? -How do these distributions differ among teacher with different characteristics?

9 External validity

Teacher career paths and their typologies differ substantially across teacher graduate cohorts. This likely suggests that teacher career paths are influenced by general societal conditions, e.g. fertility rates, the state of the economy, the state of the labor market for teachers etc., at play during the time the teachers are participating in the labor market and, perhaps especially so, when teachers graduate. As such the results presented here may not generalize to societal contexts that are very different from the one that is featured in this paper.

9.1 Policy implications

Should efforts to retain teachers be focused on early- or late-carreer teachers?

10 Conclusion

700 words

Borman, Geoffrey D, and N Maritza Dowling. 2008. "Teacher Attrition and Retention: A Meta-Analytic and Narrative Review of the Research." *Review of Educational Research* 78 (3): 367–409.

Boyd, Donald, Hamilton Lankford, Susanna Loeb, and James Wyckoff. 2002. "Initial Matches, Transfers, and Quits: Career Decisions and the Disparities in Average Teacher Qualifications Across Schools." *Unpublished Manuscript, University of Albany, SUNY.*

Falch, Torberg. 2022. "Teacher Education and Early Teaching Career." *Teachers and Teaching* 28 (8): 943–63.

Goldhaber, Dan, John Krieg, Roddy Theobald, and Stephanie Liddle. 2022. "Lost to the System? A Descriptive Exploration of Teacher Candidates' Career Paths." *Educational Researcher* 51 (4): 255–64.

Guarino, Cassandra M, Lucrecia Santibanez, and Glenn A Daley. 2006. "Teacher Recruitment and Retention: A Review of the Recent Empirical Literature." *Review of Educational Research* 76 (2): 173–208.

Gubler, Martin, Torsten Biemann, and Silvio Herzog. 2017. “An Apple Doesn’t Fall Far from the Tree—or Does It? Occupational Inheritance and Teachers’ Career Patterns.” *Journal of Vocational Behavior* 100: 1–14.

Nguyen, Tuan D, Lam D Pham, Michael Crouch, and Matthew G Springer. 2020. “The Correlates of Teacher Turnover: An Updated and Expanded Meta-Analysis of the Literature.” *Educational Research Review* 31: 100355.