Wallace, A. M. (2016)

A Typology of Delayed Graduation: Using Sequence Analysis of Enrollment Data to Uncover Heterogeneous Paths to a Degree

in G. Ritschard & M. Studer (eds), Proceedings of the International Conference on Sequence Analysis and Related Methods, Lausanne, June 8-10, 2016, pp 769-801.

A Typology of Delayed Graduation:

Using Sequence Analysis of Enrollment Data to Uncover Heterogeneous Paths to a Degree[1]

Andrew M Wallace

City University of New York, Office of Policy Research

Abstract Taking more than the "traditional" amount of time to graduate is an increasingly common path for undergraduate students at both the baccalaureate and associate levels. The expectation that these degrees can be earned in four and two years respectively is less likely to hold for students, especially those at public universities. This study uses transcript data from a large, urban public university system to examine patterns of enrollment among delayed completers. For the purposes of the analysis, delayed graduation is defined as more than six years for a baccalaureate degree and four years for an associate degree.

Traditional Time to Degree is Increasingly Unrepresentative

Large proportions of undergraduates in American colleges and universities fail to complete a degree within the "ideal" time frame: two years for an associate degree and four years for a baccalaureate degree. Even when allowing a six-year time frame for graduation, nationwide only 21.6% of entrants to associate programs and 61.8% of entrants to baccalaureate programs have finished a degree (NCES 2011). Many policy makers and researchers view these low numbers as indicating serious flaws in our system of higher education, harming both the students involved and our nation's economic competitiveness (Goldin and Katz 2009). Additionally, graduation rates differ between ethnicities, with Black and Hispanic students graduating at lower rates compared to their White and Asian counterparts (Massey et al., 2011).

This paper will present an analysis of patterns of enrollment of students who earn a degree, but not necessarily "on-time." For the purposes of this analysis, on-time graduation is defined as earning a certificate or associate degree within eight se-

[1] An earlier version of this paper appeared as a chapter in my dissertation *The Measure of a Man: The Role of Measurement in Shaping Our Understanding of College Graduation along Ethnic Lines*. This paper was made possible with generous support from the National Science Foundation (NSF).

mesters or a baccalaureate degree within 12 semesters. Sequence Analysis is used for this paper instead of a more traditional, regression-based method in order to attempt to uncover possible heterogeneity in the paths students take to their graduation. The intent is to construct a policy-relevant typology of student pathways that could provide the basis for further analysis and eventual interventions to improve student outcomes. While other methods exist for uncovering heterogeneity, they would either measure other aspects of a students pathway or they lack the potential interpretability of a typology arrived at by Sequence Analysis. For example, a Growth Mixture Model of student credit earning would allow a researcher to look for subgroups within student populations that earn credits at different rates. But because it focuses on credits instead of graduation, it does not allow us to compare patterns of students who earn different types of degrees or multiple degrees in the same way that the state space described below does. A type of model that would allow us to directly measure time to degree would be Survival Analysis. However, any subgroups identified by using a mixture model version of a Survival Analysis would not be as directly interpretable as those uncovered by Sequence Analysis.

Using Sequence Analysis on Educational Data

The way that the university semester system discretizes time means that longitudinal educational data is especially well suited to applications of Sequence Analysis. Sequence Analysis requires that time be measured in ordered, discrete units as opposed to continuously. At each point of measurement in the data under analysis, each subject is in one and only one of the states described below. The data used to create the sequences comprises 20 semesters of enrollment and graduation variables. The total number of observations is 125,515. Each semester is coded as one of the following states:

- Enrolled Full-Time (12 or more credits) without earning a degree
- Enrolled Part-Time (fewer than 12 credits) without earning a degree
- Stopped Out Before a Graduation Outcome
- Transferred Out to a Non-System Institution
- Earned a Certificate Degree
- Earned an Associate Degree
- Earned a Baccalaureate Degree
- Not Enrolled, Post-Graduation

While there are ways of dealing with missing data in sequences, for the purposes of this analysis it is not a problem. This is because not being enrolled in a given semester is an influential factor in student outcomes. It is a source of information, not a source of missingness. Further, there are effectively two types of non-enrollment in this study: expected and unexpected. After a student graduates, we would not expect them to still be enrolled in the university system. Many of the system's students come back for further education, but for the purposes of this study, those who have received at least one degree are a success. However, students who are not enrolled but have not yet received a degree are unexpectedly not enrolled.

A student may be enrolled in the semester in which they earn a degree, but for the purposes of this analysis, the type of degree they earn in a semester is far more important than the number of credits they were attempting in that semester. Because graduation is ultimately the outcome of interest, this aspect of a student's college career is the most salient feature of the state a student can be in during a given semester. For example, given three students observed in a particular semester:

- Student A: Attempted 15 credits, earned a baccalaureate degree
- Student B: Attempted 15 credits, but earned no degree
- Student C: Attempted 9 credits, earned a baccalaureate degree

I argue that student C is more similar at this point of observation to student A than is student B, regardless of the difference in credits attempted.

The fourth possible state, transfer to an outside institution, uses data collected by the university system from an outside source to measure whether or not a student who started within the system enrolled at a college or university outside of the system. The National Student Clearinghouse (NSC) provides enrollment records for students at participating colleges. For students who began college in the university system under study but did not earn a credential within the system, a query was made to the NSC to see if they turned up at another institution. This query was conducted by the institutional research office of the system and the results were provided in the data set used for analysis. The NSC also provides graduation data for students who earn a degree at participating colleges but the number of colleges that participates in the degree reporting is smaller than the number of colleges that report enrollment. Because their graduation coverage is not as comprehensive as their enrollment coverage, I am not including graduation outcomes at non-system colleges in the state space. This choice was made because the lack of equal coverage means that graduation at non-system colleges will necessarily be undercounted. For the purposes of this analysis, the fact that a student transferred to another college is sufficient information to differentiate them from students who remain within the system or who drop out of higher education entirely.

The effect of this on the coding schema that I use is that I don't have a way to differentiate between non-enrollment prior to earning a degree and non-enrollment after earning a degree for students that transfer. Because of this, I make the choice to code all semesters after which a student has an enrollment at a non-system college and does not subsequently return to the system as equivalent. Thus a semester after the observed point of transfer in which a student is not enrolled is treated as an equivalent state to a semester for which the NSC has an enrollment record. This has the effect of eliding some differences in patterns for transfer students but I argue that this is more desirable than mistakenly lumping in students who graduate from a college that participates in enrollment report but not degree reporting to the NSC with those students who transfer to a non-system college and do not earn a degree.

This choice has the side effect of making transfer cumulative thus affecting its interpretation. A student who transfers out of the system after two years will have a sequence that ends with sixteen semesters of transfer, regardless of how many enrollment records were found in the NSC. Thus an analysis of the distribution of states over time may show higher proportions of transfers (as well as stop outs and non-enrollment post-graduation). It is important for the interpretation of this distribution to keep in mind that, for cumulative states, an increase in later semesters may indicate a higher incidence of earlier experience of that state, not that students are experiencing that state for the first time later in their career. So the proportion of students who are in a cumulative state such as transfer in nineteenth semester will include a mixture of those who experience that state for the first time in that semester and those who experienced it earlier.

The Sequence Analysis is completed using the TraMineR package in R (Gabadinho et al., 2011). The data set I use for this analysis provides historical, longitudinal data on all first-time freshman, undergraduate students at the system's campuses. Students who attended another college prior to entering the university system are not included in this data set. The data set includes students who entered between September of 1999 and September 2002. These cohorts were chosen because they are the ones for whom a 10-year window of opportunity to graduate exists in the data. Admittedly, a shorter time frame would allow for more cohorts to be included, but as noted above, students often graduate in a longer time frame than what is considered traditional. A time frame longer than 10 years is not possible at this time due to the constraints of the data set.

Degree Bands

Most research on higher education distinguishes between community college students and baccalaureate students. Of this, there is research that investigates how well community college students who transfer to the baccalaureate level fare compared to either their community college compatriots or to those students who started off as baccalaureate students. Not enough research includes the reverse phenomenon, downward transfer, in its analyses. In order to fully capture the variation in student trajectories through higher education, the type of degree pursued must be measured both at entry and at exit.

To this end, I separate out these students into degree bands. Those students who ended up at the certificate level (as well as those who started there) are sufficiently few and outside of the analytic scope of this study that I am excluding them from analysis. Associate and baccalaureate degree attainment are the main focus of this analysis. The final analytic sample is separated into degree bands based on initial and final degrees sought. This allows me to include a more complete set of possible degree paths in the analysis. The degree bands used in this analysis are as follows:

- Baccalaureate at entry to and exit from the system (BA1)
- Baccalaureate at entry to and Associate at exit from the system (BA2)
- Associate at entry to and exit from the system (AA1)
- Associate at entry to and Baccalaureate at exit from the system (AA2)

As noted above, the total number of observations in the data set is 125,515. Of these, 69.01% entered the system initially seeking an associate degree. 30.99% initially sought a baccalaureate degree in the system. Of those who initially sought an associate degree, 63.66% were still pursuing an associate degree upon departure from the system, regardless of whether that departure was due to graduation, drop out, or transfer. 36.34% initially associate-seeking students transferred up to the baccalaureate level by the time of their departure from the system. Of those who initially sought a baccalaureate degree, 89.51% were still pursing a baccalaureate degree upon departure from the system and 10.49% transferred down to the associate level.

Describing the Patterns

Figure 1 shows that for initially baccalaureate students who stay at the baccalaureate level, the most common pattern is 7 semesters of full-time enrollment, followed by a semester in which they receive their baccalaureate degree. This pattern is followed by 10.4% of the BA1 population (3,618 students). 9 of the top 20 patterns for this degree band involve graduation. Further, all of those graduation patterns involve graduation within 6 years. Another 4 patterns in the top 20 involve transfer (the yellow blocks). Finally, there are 7 patterns in the top 20 that do not involve graduation or a transfer outcome. These are students who have dropped out of the higher education system during the window of analysis. The orange blocks represent semesters of non-enrollment before receiving some sort of degree. The reader will also note that there are two patterns that involve part-time enrollment (lavender). This type of enrollment is not very prevalent among the top 20 patterns for BA1 students.

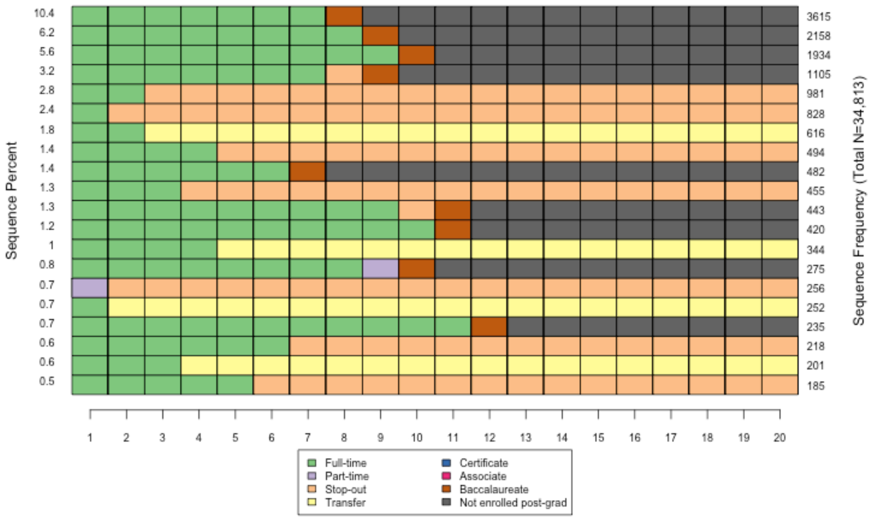


Fig. 1 Top Twenty Patterns, BA1

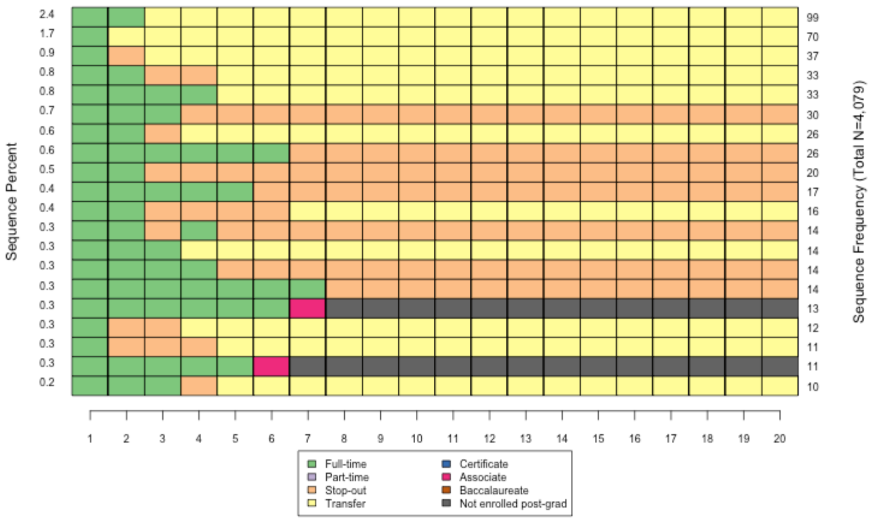


Fig. 2 Top Twenty Patterns, BA2

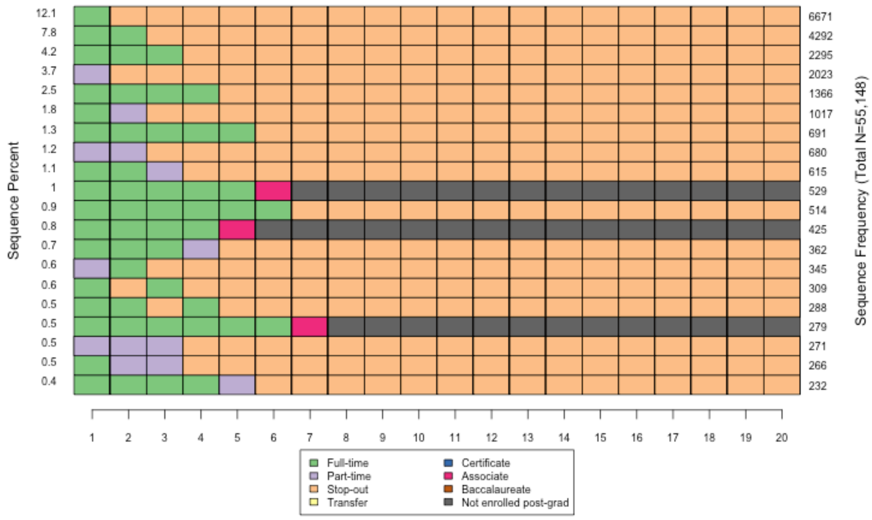


Fig. 3 Top Twenty Patterns AA1

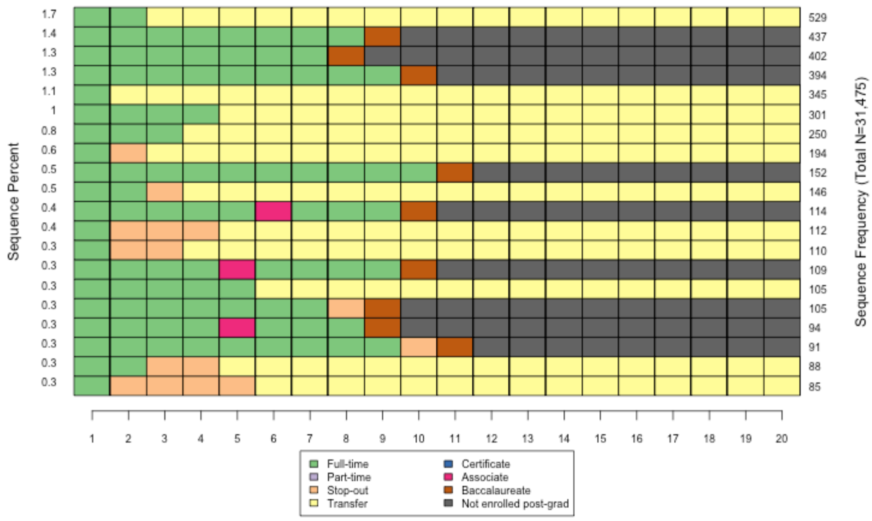


Fig. 4 Top Twenty Patterns, AA2

The top twenty patterns for the other degree bands told a rather different story. Figure 2 shows the top twenty patterns for BA2 students. Of the twenty patterns, only two involve graduation, seven involve dropping out, and eleven involve transfer. Only three of the top twenty patterns for AA1 students involve graduation and the rest involve dropping out (Figure 3). The patterns look better for AA2 students (Figure 4). Almost half of the patterns (nine) involve graduation and the remaining patterns involve transfer. None of these student's top twenty patterns involve dropping out.

As interesting as they are, the top 20 patterns for a degree band do not tell the whole story. The top 20 patterns for BA1, BA2, AA1, and AA2 only represent 44.52%, 12.75%, 42.56%, and 13.23% of their respective degree bands. These percentages indicate that students who stay at the degree level they started at are a lot more homogeneous than those who change level. This is evident from the fact that the top 20 patterns of those who stayed at the level they started at represent almost 45% and 43% of the students at the baccalaureate and associate levels respectively. On the other hand, of those who changed level, only around 13% of the students are represented by the top 20 patterns, regardless of starting level. The absolute number of students represented by these top 20 patterns is also worthy of note. For those who stayed at the same level that they started at, 15,516 and 23,520 students followed the top 20 baccalaureate and associate patterns respectively. For those who changed levels, 521 and 4,167 students are represented by the top 20 initially-baccalaureate and initially-associate patterns respectively.

Another way to examine the central tendencies of pattern data is to analyze the distribution of states over time. This distribution is not indicative of any sequence in the data much less the most common pattern. What the distribution can elucidate is general trends in state as the time window progresses. Figure 5 shows state distribution for the BA1 group of students (those who pursue a baccalaureate degree at entry to and exit from the system). In the figure, we can see that the proportion of students in the stop out state starts to grow in the second semester and peaks around the eighth semester. Around the same time as the peak of stop out, the baccalaureate graduation state (rust) starts to grow. Consequently, the proportion of non-enrollment post-graduation starts to grow, peaking at 57% in the twentieth semester.

The state distribution for the BA2 students shows smaller proportions of positive outcomes (Figure 6). The state of stopping out grows much more quickly and ends up being a larger proportion than it did for BA1 students. It peaks at 48% in the tenth semester. The proportion of transfer states is greater for this subgroup than the BA1 students, topping out at 41% in the twentieth semester. That said, there is some graduation and thereby non-enrollment post-graduation for these students starting after the seventh semester and growing slowly but steadily to 20% by the end of the analytic window.

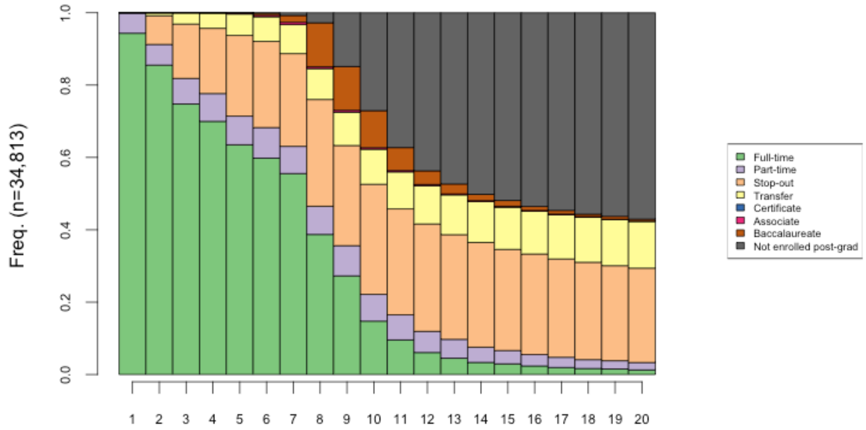


Fig. 5, State Distribution, BA1

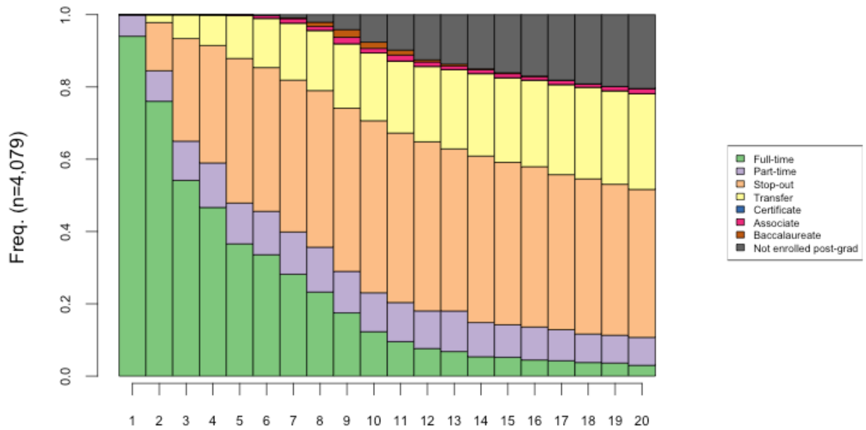


Fig. 6 State Distribution, BA2

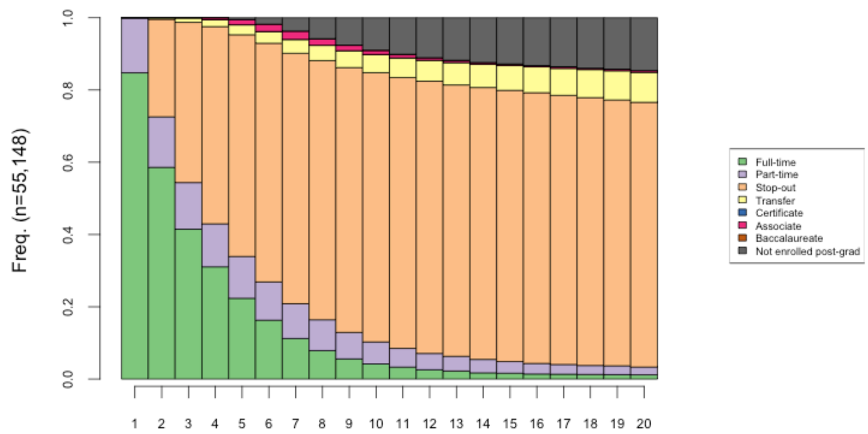


Fig. 7, State Distribution AA1

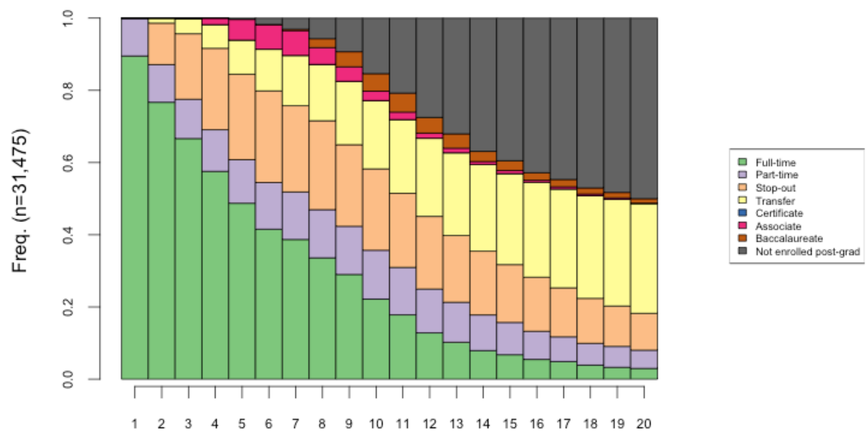


Fig. 8, State Distribution, AA2

Figures 7 and 8 show the opposite trend in comparison for those who start at the associate level. Those who stay at the associate level have worse outcomes than those who transfer up to the baccalaureate level. This echoes findings in the literature that notes that students who manage to make it to the baccalaureate level from the associate level do as well or better than those who began at four-year colleges and stayed there. For AA1 students (those who were at the associate level at entry to and exit from the system), the proportion of students in the stop out state rises steeply beginning in the second semester then leveling out after the seventh semester. As a proportion of all states, stopping out peaks for these students at 75% in the tenth semester and remains around that level until the seventeenth semester where it declines slightly until the end of the analytic tracking window.

By contrast, AA2 students have increasing proportions of graduation, transfer, and non-enrollment post-graduation states starting in the fourth semester. Half of AA2 students are in a state of non-enrollment post-graduation by the end of the tracking window and a further 30% have transferred outside of the system.

Taken together, the top twenty patterns for a degree band, as well as the distribution of states over time gives a better indication of the central tendencies of the data than either measure alone. The top twenty patterns show commonalities in whole patterns while the distributions show changes in the composition of states in the subpopulation over time. The top twenty patterns point to the most common outcomes for students in a given group. The state distributions allow for comparisons at a given time point between groups. The differences between degree bands at a given time point as well as the distribution trends over time show how BA1 and AA2 students do systematically better than BA2 and AA1 students.

The average time spent in a given state is another useful measure of the central tendency of sequence data. Figures A.1 – A.4 in the appendix represent these measures calculated by degree band. On average, BA1 students spend slightly more than six semesters enrolled full-time and not quite six semesters not enrolled after graduating. BA2 students spend much less time enrolled full-time, more time stopped out, and much less time in post-graduation non-enrollment. AA1 students spend a distressingly long time in the state of being stopped out, which is likely due to dropping out early. In contrast, AA2 students average a similar amount of time enrolled full-time to BA1 students. They average less time not enrolled post-graduation but this is probably due to the effect of transferring programs (if not colleges) on delaying graduation.

While mean time in state tells us the proportion of a student's career that they spend in a given state, it does not tell us how long they spent in that state in a single spell. That means that a student who is enrolled full-time for ten consecutive semesters and then stops out for ten consecutive semesters will have the same average time in the full-time and stopped out states as a student who alternates between the two every other semester. Nonetheless, the average amount of time that a student spends in a state is still useful information.

Table 1 presents this information in another form: the proportion of a sequence that students average in a state. While Figure A.3 tells us that an AA1 student spends 12.88 semesters out of 20 stopped out, this table shows that accounts for 64.41% of a student's career. Compared to the 38.72% that BA2 students spend stopped out, the 23.97% that BA1 students spend stopped out, and the 17.77% that AA2 students spend stopped out, this paints a dismal picture for AA1 students. This table also shows that BA1 students spend the largest proportion of their time enrolled full-time and AA2 students have the largest proportion of part-time enrollment.

Table 1. Mean Time Spent in State by Degree Band

State	BA1	BA2	AA1	AA2
Full-time	30.97	23.80	15.06	29.01
Part-time	5.61	9.96	7.06	10.42
Stop Out	23.97	38.72	64.41	17.50
Transfer	8.70	17.36	4.85	17.77
Certificate	0.00	0.01	0.01	0.01
Associate	0.21	0.90	0.77	2.05
Baccalaureate	2.87	0.43	0.02	1.98
Non-Enrollment Post-Grad	27.66	8.83	7.81	21.26

Clustering the Patterns

It is important to note that, while there are generally accepted techniques for doing a cluster analysis, a certain amount of subjectivity is involved with a cluster analysis because of the necessity of choosing one algorithm over another. Further, the choice of a final number of clusters is ultimately based on theory and interpretability as much as it is upon any objective measure of cluster quality. Cluster analysis of sequence data requires that a matrix of distances be calculated to tell the researcher how close (by whatever measure) any two sequences are. The distance measure chosen for this analysis is the Optimal Matching distance as implemented by TraMineR. There is no consensus in the literature as to the single best way to weight the substitution, insertions, and deletions, but it is generally acknowledged that theory should be a guiding force in any weighting schema (Abbott and Tsay 2000; Gauthier et al., 2009; Lesnard, 2010). I chose to base the substitution cost on transition rates in order to have the difficulty of exchanging one state for another at any given time point be tied into how often this transition occurs from one time point to another. While this makes the weights less generalizable because the clusters are more dependent on the transitions that occur in this particular data set, I argue that the benefit of not artificially imposing a substitution cost based on intuition outweighs the chance that the transitions occurring in this rather large data set are systematically different than those that might occur with a different data set. Tables A.5 – A.8 in the Appendix present the substitution cost matrices by degree band.

Once a distance matrix is calculated, a clustering algorithm needs to be chosen. In this analysis, I use Ward's method to cluster patterns because of its wide usage (Murtagh and Legendre, 2014). Table 2 presents a variety of measures of cluster quality. While the cluster quality measures produced by TraMineR suggested a two-, three-, or four-cluster solution, these clusters were not very informative and were certainly not policy-relevant. Essentially those cluster solutions tell us that students graduate, drop out, and transfer and not much more. In order to find more interesting patterns, I look at larger numbers of clusters to see what groups would emerge from the data. In these clusters, I find interesting patterns of degree completion, transfer, and dropping out. Looking at the various cluster solutions, I arrive at the set of cluster solutions presented in Table 3 as the best balance of interpretability and sample size. That is, I looked at representative sequences from each possible cluster solution (up to 20 clusters) and interpreted the story of the members of that cluster based on the representative sequences. Different numbers of cluster solutions were chosen for each degree band because the different degree bands had differing amounts of heterogeneity and with some of the degree bands (AA1 in particular), it took a higher number of clusters for interesting sequences to separate out from the larger clusters present in cluster solutions with fewer numbers of clusters.

Table 2 Quality Measures of Best Cluster Solutions by Degree Band[2]

	BA1		BA2		AA1		AA2	
	Clusters	Statistics	Clusters	Statistics	Clusters	Statistics	Clusters	Statistics
PBC	3	0.88	3	0.79	3	0.90	3	0.81
HG	3	0.98	3	0.91	3	0.99	4	0.94
HGSD	3	0.98	3	0.91	3	0.99	4	0.94
ASW	3	0.68	3	0.52	3	0.72	3	0.51
ASWw	3	0.68	3	0.52	3	0.72	3	0.51
CH	2	23492.5	3	1654.17	2	26219.23	2	13959.77
CHsq	3	66154.18	3	4511.72	3	85479.95	3	35738.92
HC	3	0.02	3	0.05	3	0.01	4	0.03

[2] The cluster quality measures are as follows: Point Biserial Correlation (PBC), Hubert's Gamma (HG), Hubert's Somer's D (HGSD), Average Silhouette Width (ASW), Average Silhouette Width – weighted (ASWw), Calinski-Harabasz index (CH), Calinski-Harabasz index squared (CHsq), and Hubert's C (HC). For details, see Studer (2013).

Table 3 Final Number of Clusters Chosen

Degree Band	Number of Clusters
BA1	10
BA2	8
AA1	12
AA2	8

Figures A.9 – A.19 in the Appendix present the representative sequences for all degree bands. For example, the subfigures in Figure A.9 show the ten representative sequences chosen by TraMineR to represent the ten clusters for the BA1 degree band in which graduation was the defining outcome. Unlike the top twenty sequences in Figures 1 – 4, the height of the representative sequences is proportional to how many students in the cluster followed a given sequence. The color scheme for this figure is the same as it was in the top twenty pattern figures.

Each subfigure within the larger figure represents a cluster that I have named based on the sequences presented. For example, in one cluster of BA1 students, the top ten sequences that TraMineR chose to represent this cluster all involved earning a baccalaureate degree by the sixth year from entry. As this was the characteristic that seemed to differentiate this cluster from the others, I labeled this cluster "On-time Graduates." Figure A.9(a) shows the top ten representative sequences for this cluster.

I have arranged the clusters in the overall figure in order of how long it took the students to reach the graduation outcome. The biggest difference between (a) and (b) is the how many semesters it took the students to get there (4-6 years instead of 6-7 years). They are both otherwise characterized by mostly full-time attendance (green) with some part-time attendance (lavender) and some stop out (orange). The clusters represented by subfigures (c) and (d) represent graduation outcomes that occur in the 6-8 year range. They are differentiated from each other by the manner in which the students got to this outcome in that time frame. Students in subfigure (c) took a leave that lasted 3-6 semesters and graduated immediately upon returning to the system. Students represented by subfigure (d) were enrolled for a significant amount of time attempting fewer than twelve credits (part-time) on their way to graduation. This shows a difference in the manner in which students achieved their outcome, not simply in the duration to that outcome as we saw between (a) and (b). Subfigure (e) shows patterns that represent students who stopped out relatively early (within 2.5 years) and took a long break (2 - 6.5 years) before coming back to finish their degree.

Table 4 Cluster Description and Frequencies, BA1.

Outcome	Cluster Name	N	Percent within Degree Band	Percent Coverage
	On-time Graduates	16,269	46.73	81.6
	6-7 year Graduates	2,176	6.25	52.4
	7-8 year Graduates with Gap	881	2.53	21.0
Graduates	6-8 year Graduates with lots of part-time	804	2.31	10.0
	7-10 year Grads with a long break	513	1.47	12.2
Transfers	Early Transfers	3,710	10.66	75.7
	Late Transfers	567	1.63	41.4
Drop Outs	Early Drop Outs	7,490	21.51	77.2
	Late Drop Outs	1,920	5.52	49.9
Other	Characterized by a lot of part-time	483	1.39	14.5

Table 5 Cluster Description and Frequencies, BA2

Outcome	Cluster Name	N	Percent within Degree Band	Percent Coverage
Graduates	Associate Degree in 3-5 years	398	9.76	39.4
	Associate Degree in 6-9 years	158	3.87	13.9
	Earn Baccalaureate then Seek Associate	317	7.77	32.2
Transfers	Early Transfers	727	17.82	69.7
	Late Transfers	304	7.45	52.3
Drop Outs	Early Drop Outs	1,049	25.72	41.3
	Late Drop Outs	563	13.80	26.6
	Porpoising Enrollment	563	13.8	8.0

Table 6 Cluster Description and Frequencies, AA1

Outcome	Cluster Name	N	Percent within Degree Band	Percent Coverage
	2-4 year Associate	5,776	10.47	64.9
Graduates	5-7 year Associate because of break	1,001	1.82	13.8
	5-7 year Associate because of part-time	794	1.44	22.1
	7-10 year Associate because of long break	565	1.02	12.9
	Early Transfers	2,168	3.93	71.6
Transfers	Middle Transfers	1,221	2.21	60.3
	Late Transfers	1,042	1.89	59.1
	Drop out within 2 years	27,541	49.94	100.0
	Drop out in 2-4 years	9,090	16,48	71.0
	Drop out in 3-5 years with part-time	3,327	6.03	49.3
Drop Outs	Drop out in 4-5 years mostly full-time	1,813	3.29	56.4
	Drop out in 6-7	810	1.47	14.7

Table 7 Cluster Description and Frequencies, AA2

Outcome	Cluster Name	N	Percent within Degree Band	Percent Coverage
	2–4.5 year Associate, 4–5.5 year Baccalaureate	5.585	17.74	60.2
	2.5–4 year Associate, 5–6.5 year Baccalaureate	8.476	26.93	28.0
Graduate	3–5 year Associate, 5.5–8 year Baccalaureate, mostly part-time	2,123	6.75	5.5
	3–5 year Associate, 8–9.5 year Baccalaureate	1,062	3.37	4.0
Transfers	Early Transfers	6,821	21.67	64.7
	Late Transfers	2.446	7.77	49.6
Drop Outs	Drop out by year 5, staggered	3,578	11.37	36.4
Other	Mostly characterized by a large break	1,384	4.40	4.3

Tables 4 – 7 present the cluster solutions in detail. They are arranged by degree band, general outcome (graduation, transfer, drop out, or other). The within-degree-band percentage of students who fall into the cluster is presented along with the percentage of coverage for the representative sequences in the Figure A.9 – A.19. Alexis Gabadinho and Gilbert Ritschard define the level of coverage for a set of sequences within a cluster as "the percentage of cases that are within the neighbourhood of at least one of the patterns in the set" (2013).

In these tables we can see that earlier onset of outcome tends to be more common than later onset in terms of the number of students in a cluster. As an example, the most common graduation outcome for AA1 students is receiving an Associate degree in two to four years, with 10.47% of students in the degree band falling within this cluster (Table 6). However, only a further 4.27% of AA1 students are categorized in the other three graduation clusters. Despite this, it is still important to take note of these students as they are largely ignored by metrics that only measure "traditional" time to degree. For this degree band, however, the biggest group of concern is the 27.541 students (49.94%) who are in the "Drop out within two years" cluster.

As the two to four cluster solutions suggested by the measures of cluster quality mentioned above indicate, there are three general trends in outcome across all four degree band: Graduation, Transfer, and Dropping Out. This is significant variation within these group and some differences across degree bands, but overall most clusters fell into one of these three categories.

However, within each outcome there is a definite distinction to be made between clusters based on the number of semesters it took students to reach that outcome. As an example in the BA2 cluster solution there are two clusters that are characterized by transfer outside of the university system under study. The difference between the clusters appears to be that one group of students transfers rather quickly after leaving the system while the other group takes a long break before transferring. The Early Transfers time to outcomes ranged from one to three semesters after their last enrollment in the system while the Late Transfers had a gap in enrollment lasting between five and fifteen semesters.

In some of the degree bands, there was more heterogeneity among those clusters that involve taking longer to reach the eventual outcome. In the AA1 cluster solution there were 4 graduation clusters. Figure A.14 presents these clusters ordered by time to outcome. Similar to the graduation clusters for the BA1 degree band, there is an on-time cluster, two delayed-graduation clusters, and a very delayed graduation cluster. As with the BA1, the difference between the delayed graduation clusters is between students who mostly attend part-time and those who take a break. Also similar to the BA1 cluster solution, the extremely delayed graduation cluster is due to a lengthy break in enrollment (as opposed to taking a number of short breaks).

Implications of the Cluster Solution

The clusters indicated by the analysis presented here have a number of potential implications both for our understanding of graduation and for what policies might be implemented to improve student outcomes.

That there are students who take longer than what is traditional to graduate is not surprising given the previous work done in this area. However, I was expecting that there would be a larger variety of differences in how students got to their delayed graduation than was found in the analysis. I expected that there might be students who came in and out of the higher education system, taking breaks in order to work or for some other reason. However, I find in general that students who took more than six years to graduate did so because they were part-time for a large portion of their career in the university system or they took a multi-semester break, after which they returned and finished the degree. The length of the break for delayed students surprised me.

This second group of delayed graduates is potentially interesting from a policy standpoint because the difference between them and a dropout is that they came back. This implies that further work can be done to explore the systematic differences between dropouts and those students who succeed despite an absence from higher education.

The cluster solution for the BA1 degree band along with those for the other degree bands support the typology of delayed graduation that I am proposing:

- Those who have no delay
- Those who have a short delay
- Those who have a medium delay with a break
- Those who have a medium delay with part-time enrollment
- Those who have a long delay with a long break

This typology is potentially useful as the basis for further research that would inform a variety of interventions aimed at helping student graduation faster. For example, following up with students who had a long or short break before graduation could help administrators understand the reasons that ultimately successful students were not able to complete a degree in a contiguous manner.

Further, the existence of these delayed graduates whose pathway to degree includes a break implies a potential intervention for those students who have stopped out of the system and have not yet returned to complete their degree. Checking in on a student who nearly has enough credits to graduate but is not currently enrolled would reveal a wealth of actionable information. It could allow administrators to discover if a student has stopped out because of family obligations, financial considerations, or some other factor outside the control of the institution. If, however, the reason a student stops out just shy of the necessary credits for a degree is something within the scope of influence of the institution (class availability, tutoring, etc.), then the institution can attempt to provide the necessary services to help the student see their degree to completion.

On the other hand, students who drop out early in their career (a large proportion of AA1 students) would be helped by interventions earlier in the process. If a comprehensive survey of the reasons a student drops out early could be found by following up with those students who drop out early, the institution could tailor supports to students who face similar obstacles at or before entry to college.

The difference between paths to degree for delayed graduates and the similarity between a delayed graduate and a drop out who has nearly enough credits to graduate implies the need for interventions to bring these students back into the fold. It is important to note that a single type of intervention will be less likely to work than a variety of interventions targeted at students based on their pathway through college. A student who struggles early on and is in danger of dropping out is likely to need different supports than a student who is nearly done with their degree but is having trouble crossing the finish line. The typology proposed in this paper has the potential to act as a framework for understanding students' path to degree and where in their careers students might need additional supports to complete in a more timely manner. While it is important not to judge students who take longer to earn their degree if that is their choice or is the best possible solution given their life circumstance, the sooner a student completes their degree the sooner they can reap the economic benefits it will afford them.

References

Abbott, A. and Tsay, A. (2000). Sequence analysis and optimal matching methods in sociology review and prospects. *Sociological Methods and Research*, 29(1), 3–33.

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

Gabadinho, A., Ritschard, G. (2013). Searching for typical life trajectories applied to childbirth histories in in R. Levy and E. Widmer (eds) Gendered life courses – Between individualization and standardization. A European approach applied to Switzerland, 287 – 312.

Gauthier, J. A., Widmer, E. D., Bucher, P., and Notredame, C. (2009). How much does it cost? optimization of costs in sequence analysis of social science data. *Sociological Methods and Research*, 38(1), 197–231.

Goldin, C. D. and Katz, L. F. (2009). *The race between education and technology*. Cambridge, MA: Harvard University Press

Lesnard, L. (2010). Setting cost in optimal matching to uncover contemporaneous socio-temporal patterns. *Sociological Methods and Research*, 38(3), 389–419.

Massey, D. S., Charles, C. Z., Lundy, G., and Fischer, M. J. (2011). *The source of the river: The social origins of freshmen at America's selective colleges and universities*. Princeton, NJ.: Princeton University Press

Murtagh, F. and Legendre, P. (2014). Ward's hierarchical agglomerative clustering method: Which algorithms implement ward's criterion? *Journal of Classification*, 31(3), 274–295.

NCES (2011) Six-Year Attainment, Persistence, Transfer, Retention, and Withdrawal Rates of Students Who Began Postsecondary Education in 2003–04. U.S. Department of Education. Institute of Education Sciences, National Center for Education Statistics. https://www2.ed.gov/rschstat/research/pubs/prinindicat/prinindicat.pdf Accessed January 9, 2016

Studer, M. (2013). Weightedcluster library manual: A practical guide to creating typologies of trajectories in the social sciences with R. LIVES Working Papers, 24.

Appendix

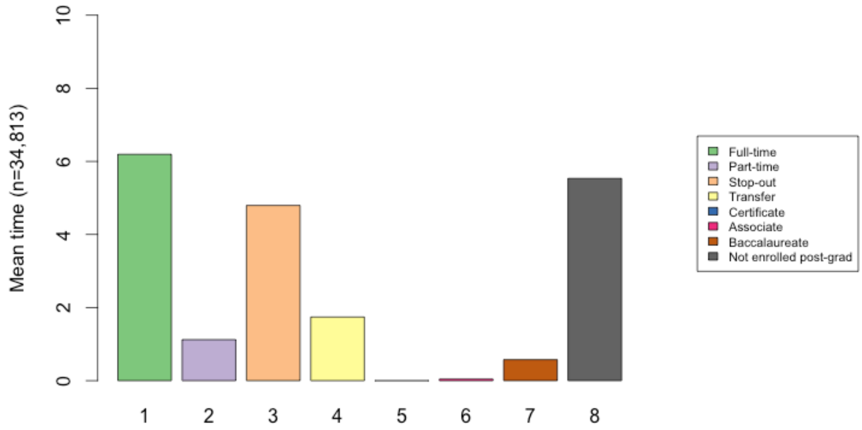


Fig. A.1 Mean Time in State, BA1

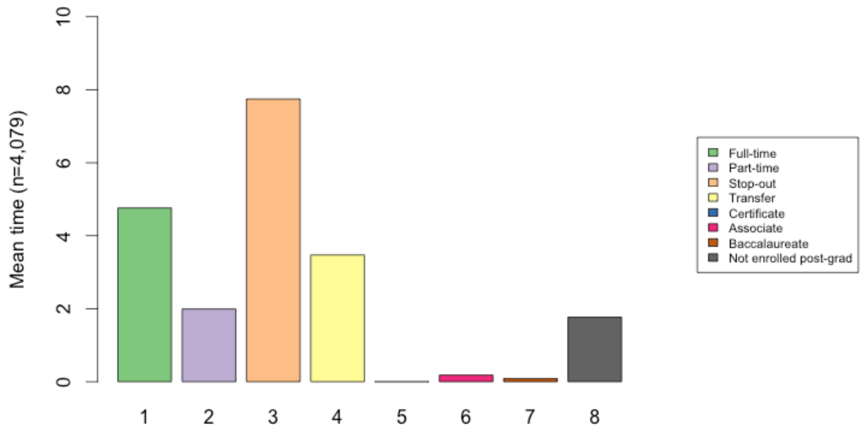


Fig. A.2 Mean Time in State, BA2.

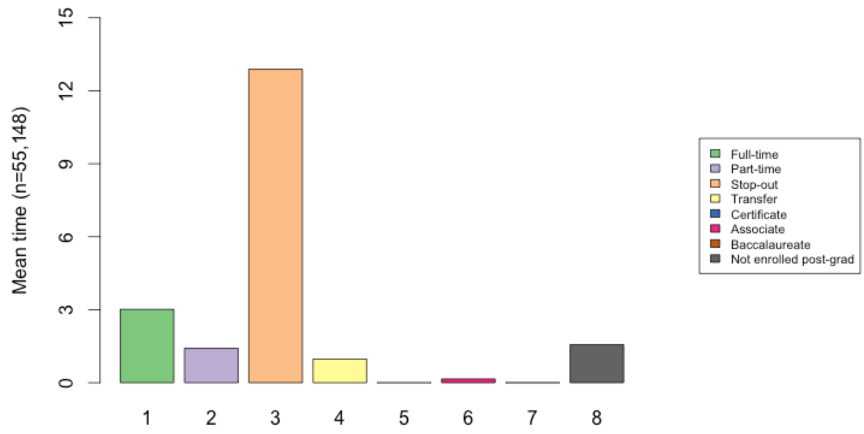


Fig. A.3 Mean Time in State, AA1.

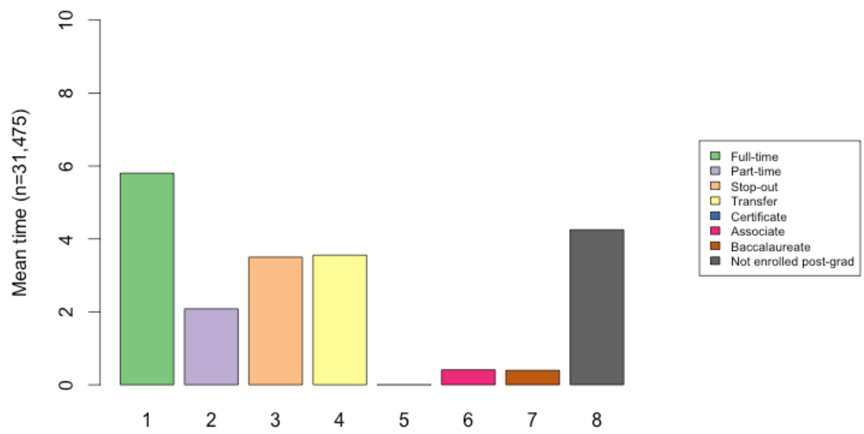


Fig. A.4 Mean Time in State, AA2.

Table A.5 Substitution Cost Matrix, BA1

	1	2	3	4	5	6	7	8
1	0.00	1.69	1.87	1.99	1.86	1.50	1.93	2.00
2	1.69	0.00	1.74	1.99	1.93	1.83	1.94	1.97
3	1.87	1.74	0.00	1.99	2.00	2.00	1.98	2.00
4	1.99	1.99	1.99	0.00	2.00	2.00	2.00	2.00
5	1.86	1.93	2.00	2.00	0.00	2.00	1.93	1.29
6	1.50	1.83	2.00	2.00	2.00	0.00	1.92	1.74
7	1.93	1.94	1.98	2.00	1.93	1.92	0.00	1.01
8	2.00	1.97	2.00	2.00	1.29	1.74	1.01	0.00

Table A.6 Substitution Cost Matrix, BA2.

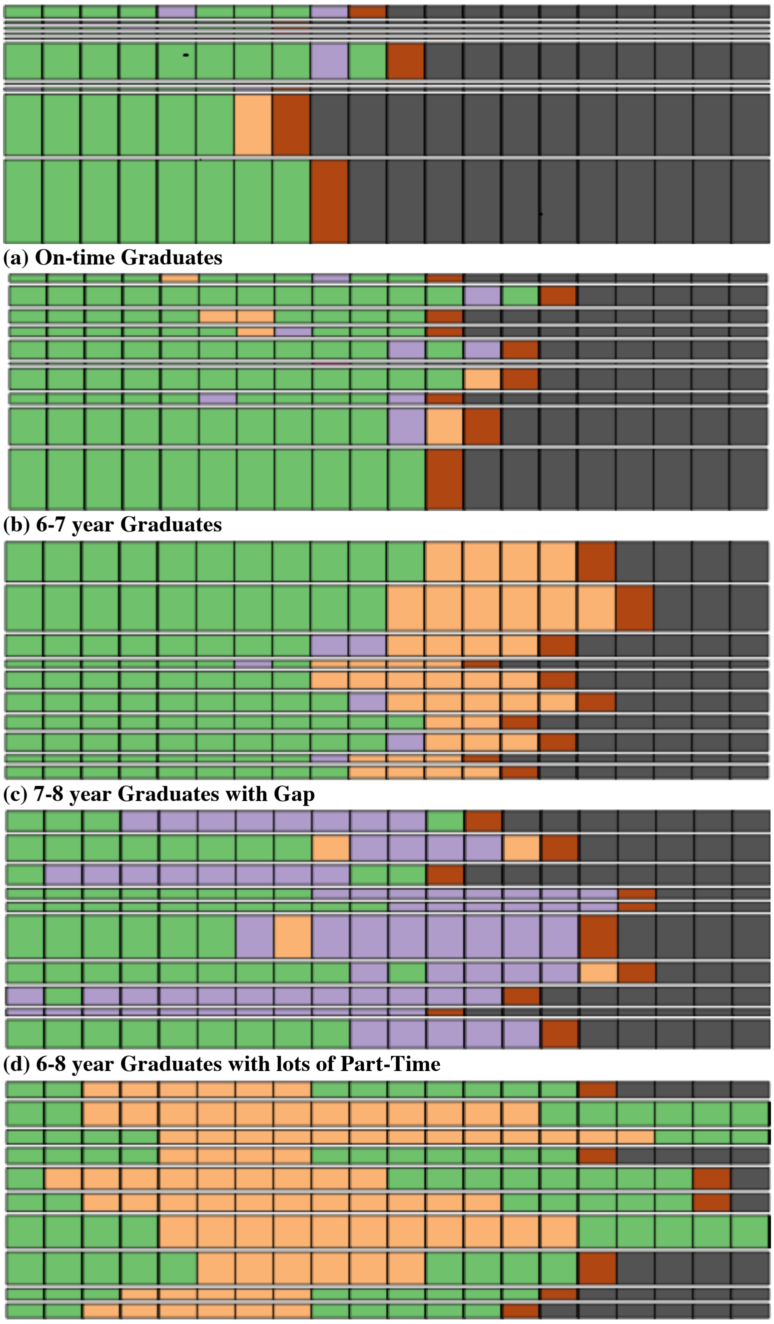
	1	2	3	4	5	6	7	8
1	0.00	1.70	1.72	1.98	1.89	1.95	1.93	1.99
2	1.70	0.00	1.62	1.99	1.67	1.92	1.87	1.92
3	1.72	1.62	0.00	1.96	2.00	1.99	2.00	2.00
4	1.98	1.99	1.96	0.00	2.00	2.00	2.00	2.00
5	1.89	1.67	2.00	2.00	0.00	2.00	2.00	1.44
6	1.95	1.92	1.99	2.00	2.00	0.00	1.99	1.09
7	1.93	1.87	2.00	2.00	2.00	1.99	0.00	1.18
8	1.99	1.92	2.00	2.00	1.44	1.09	1.18	0.00

Table A.7 Substitution Cost Matrix, AA1

	1	2	3	4	5	6	7	8
1	0.00	1.71	1.70	2.00	1.82	1.94	1.96	2.00
2	1.71	0.00	1.60	2.00	1.83	1.93	1.82	1.98
3	1.70	1.60	0.00	1.99	2.00	2.00	2.00	2.00
4	2.00	2.00	1.99	0.00	2.00	2.00	2.00	2.00
5	1.82	1.83	2.00	2.00	0.00	1.96	2.00	1.39
6	1.94	1.93	2.00	2.00	1.96	0.00	2.00	1.08
7	1.96	1.82	2.00	2.00	2.00	2.00	0.00	1.22
8	2.00	1.98	2.00	2.00	1.39	1.08	1.22	0.00

Table A.8 Substitution Cost Matrix, AA2

	1	2	3	4	5	6	7	8
1	0.00	1.67	1.82	1.99	1.73	1.47	1.96	1.96
2	1.67	0.00	1.78	1.99	1.73	1.79	1.95	1.89
3	1.82	1.78	0.00	1.94	2.00	1.97	1.99	2.00
4	1.99	1.99	1.94	0.00	2.00	2.00	2.00	2.00
5	1.73	1.73	2.00	2.00	0.00	1.96	1.98	1.61
6	1.47	1.79	1.97	2.00	1.96	0.00	1.97	1.70
7	1.96	1.95	1.99	2.00	1.98	1.97	0.00	1.00
8	1.96	1.89	2.00	2.00	1.61	1.70	1.00	0.00



(a) On-time Graduates

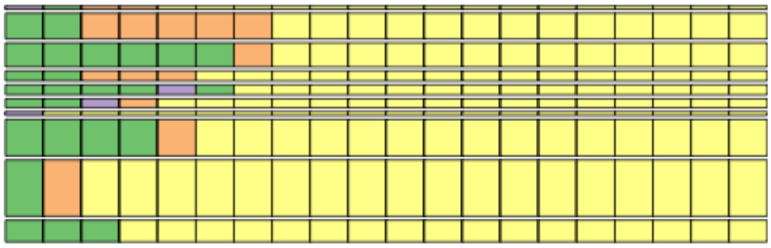
(b) 6-7 year Graduates

(c) 7-8 year Graduates with Gap

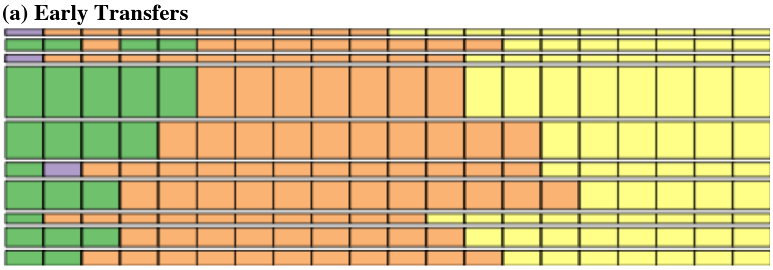
(d) 6-8 year Graduates with lots of Part-Time

(e) 7-10 year Graduates with a long break

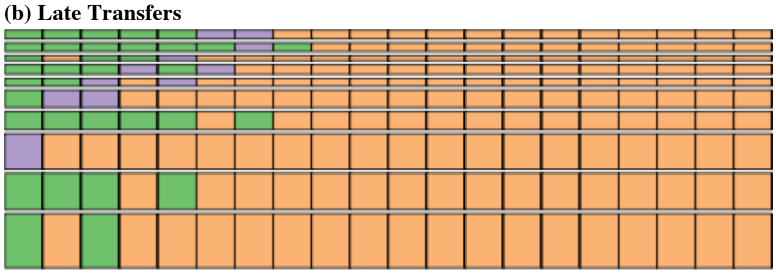
Fig. A.9 Representative Sequences, Graduation Group, BA1



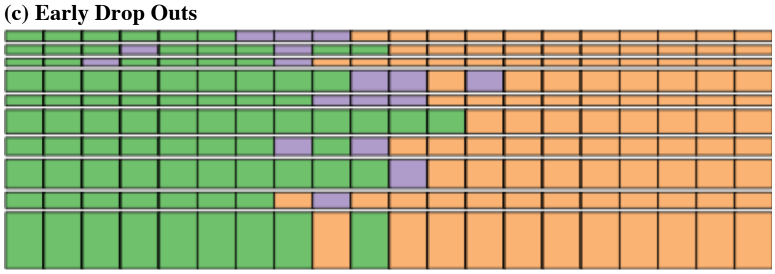
(a) Early Transfers



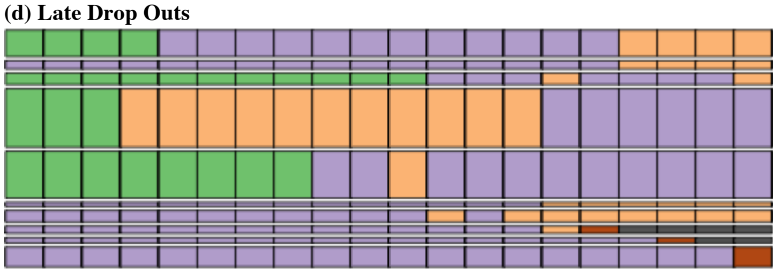
(b) Late Transfers



(c) Early Drop Outs

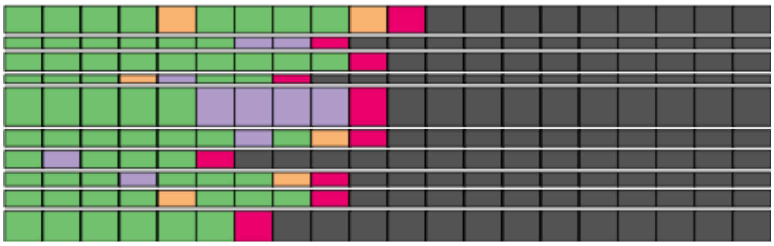


(d) Late Drop Outs

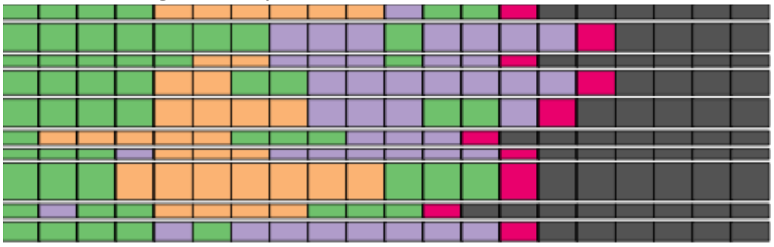


(e) Other, Characterized by a lot of part-time

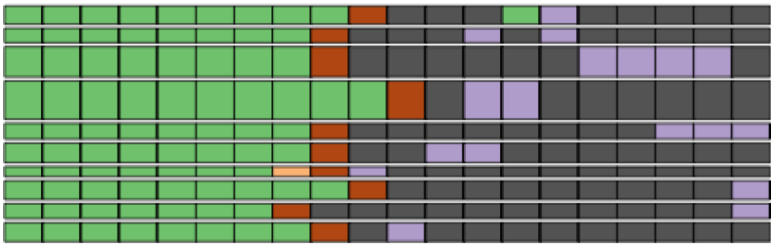
Fig. A.10 Representative Sequences, Transfer, Drop Out, and Other Groups, BA1



(a) Associate Degree in 3-5 years

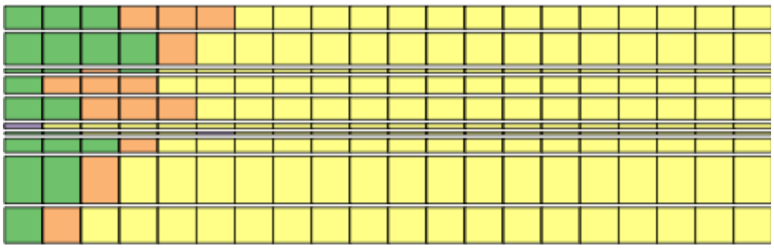


(b) Associate Degree in 6-9 years

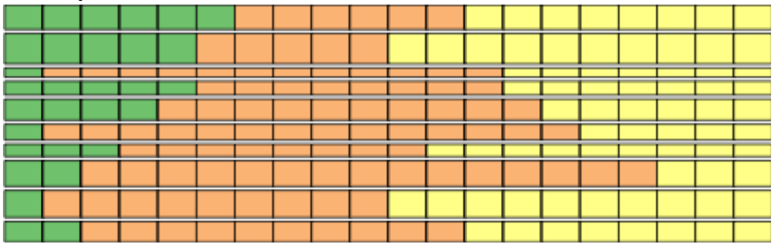


(c) Earn Baccalaureate then Seek Associate

Fig. A.11 Representative Sequences, Graduation Group, BA2

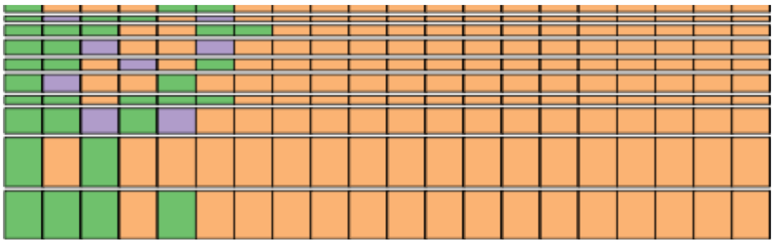


(a) Early Transfers

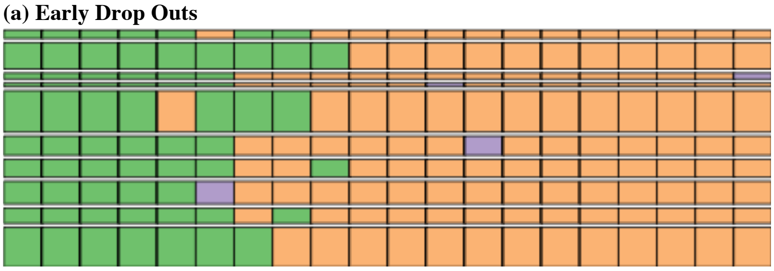


(b) Late Transfers

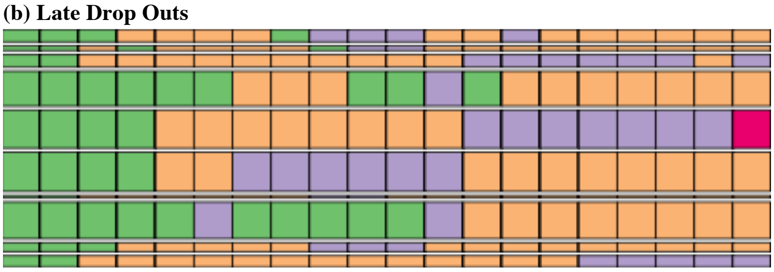
Fig. A.12 Representative Sequences, Transfer Group, BA2



(a) Early Drop Outs

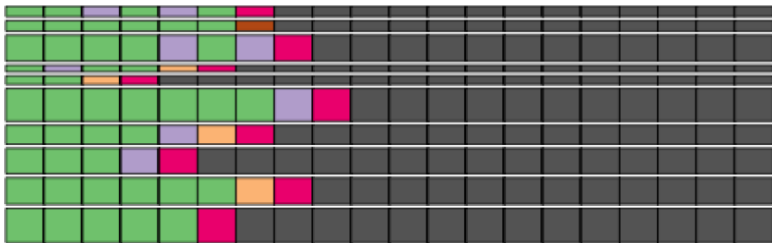


(b) Late Drop Outs

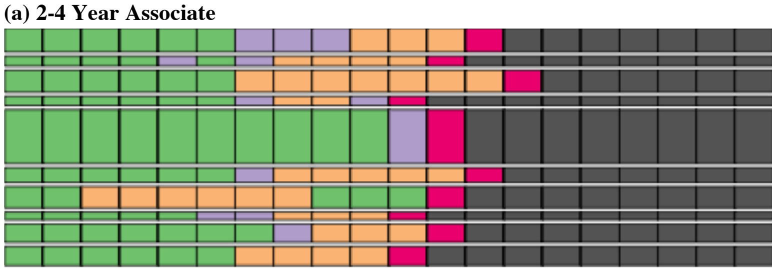


(c) Porpoising Enrollment

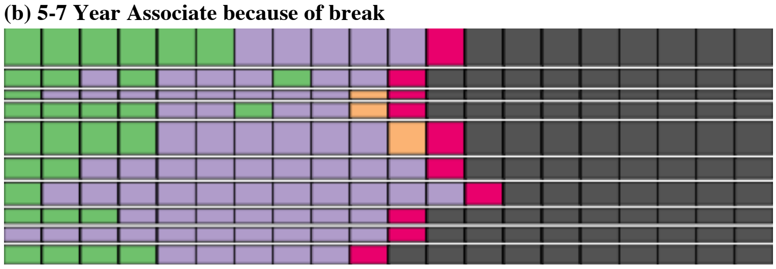
Fig. A.13 Representative Sequences, Drop Out Group and Other Group, BA2



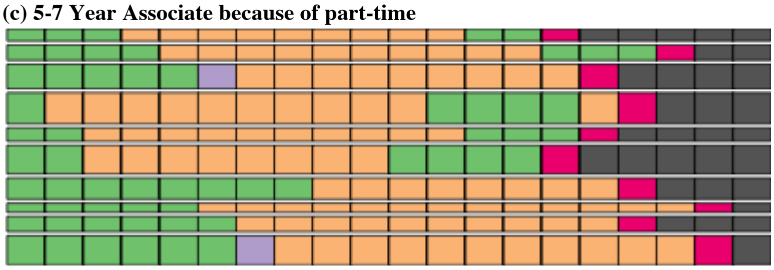
(a) 2-4 Year Associate



(b) 5-7 Year Associate because of break

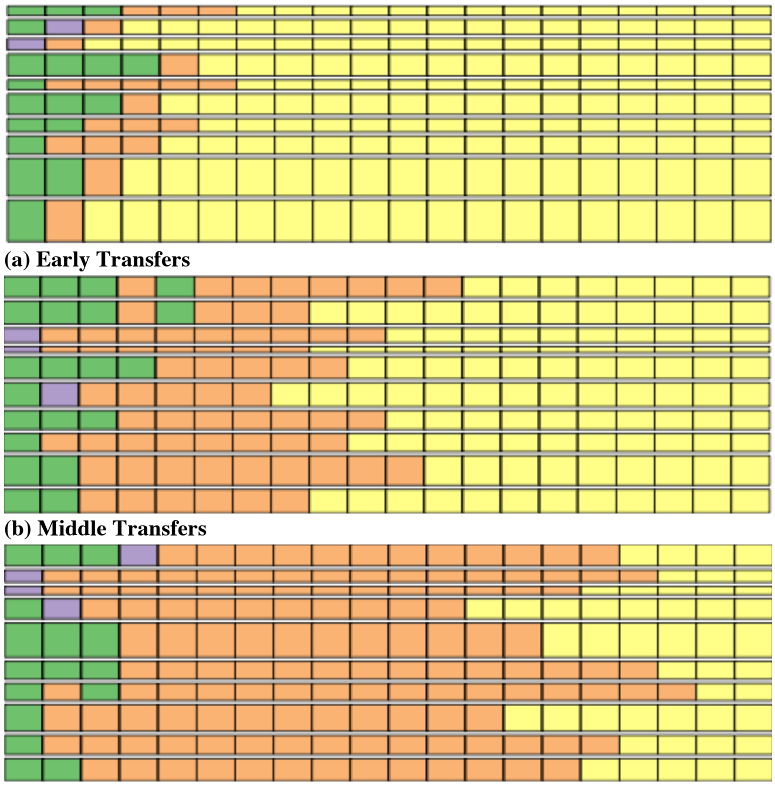


(c) 5-7 Year Associate because of part-time



(d) 7-10 Year Associate because of long break

Fig. A.14 Representative Sequences, Graduation Group, AA1.

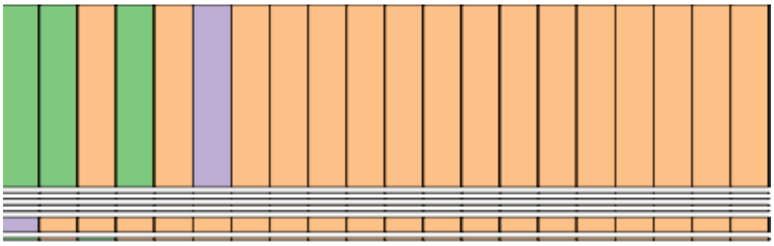


(a) Early Transfers

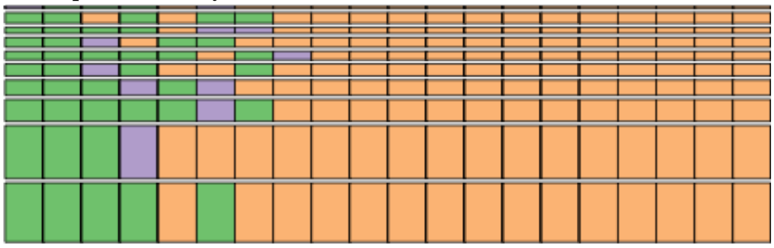
(b) Middle Transfers

(c) Late Transfers

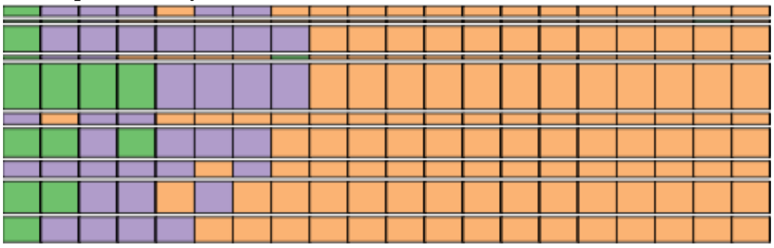
Fig. A.15 Representative Sequences, Transfer Group, AA1



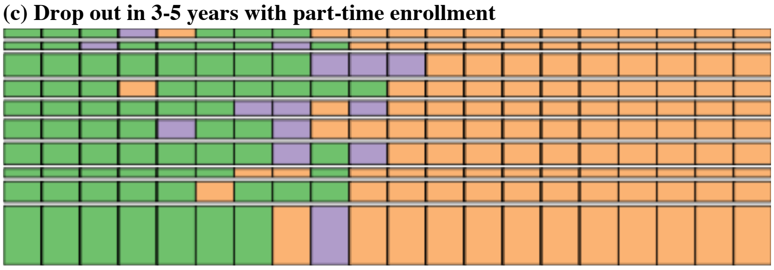
(a) Drop out within 2 years



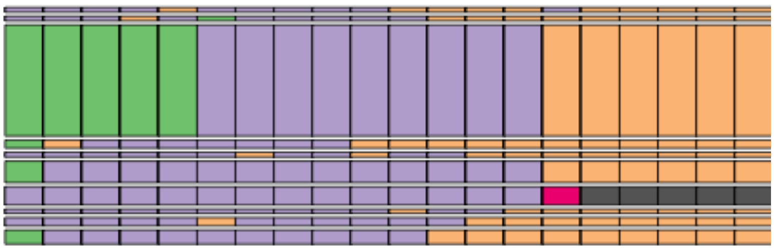
(b) Drop out in 2-4 years



(c) Drop out in 3-5 years with part-time enrollment

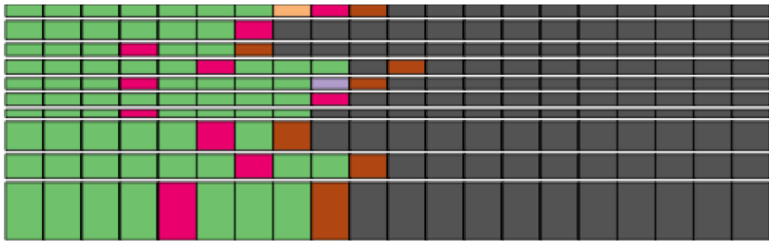


(d) Drop out in 4-5 years mostly full-time enrollment

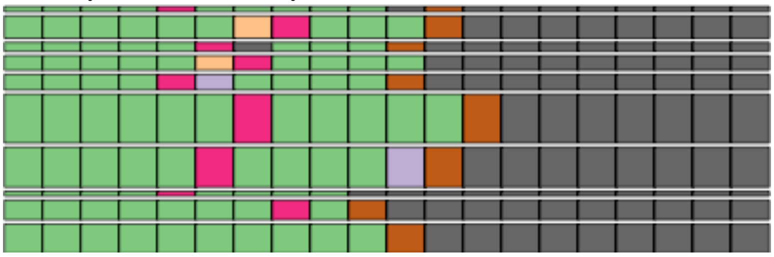


(e) Drop out in 6-7 years

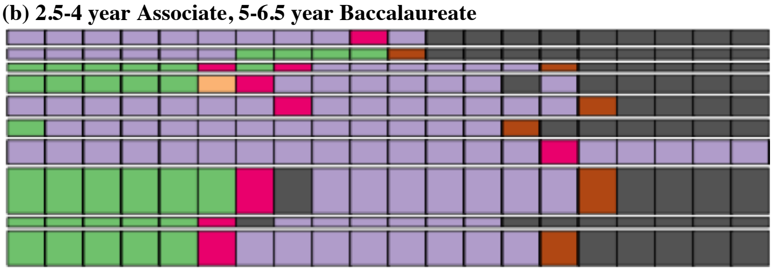
Fig. A.16 Representative Sequences, Drop Out Group, AA1



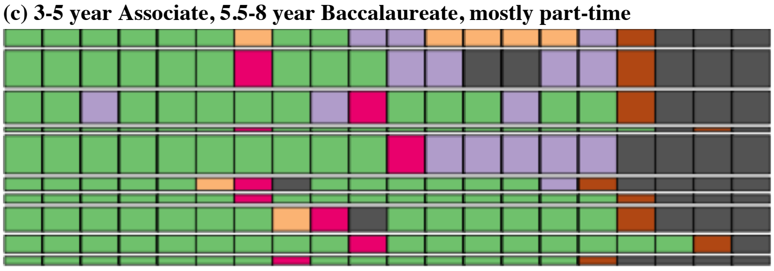
(a) 2-4.5 year Associate, 4-5.5 year Baccalaureate



(b) 2.5-4 year Associate, 5-6.5 year Baccalaureate

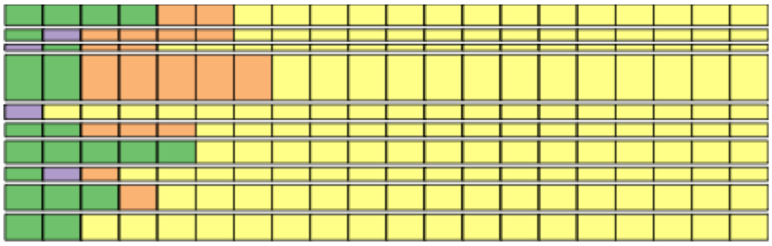


(c) 3-5 year Associate, 5.5-8 year Baccalaureate, mostly part-time

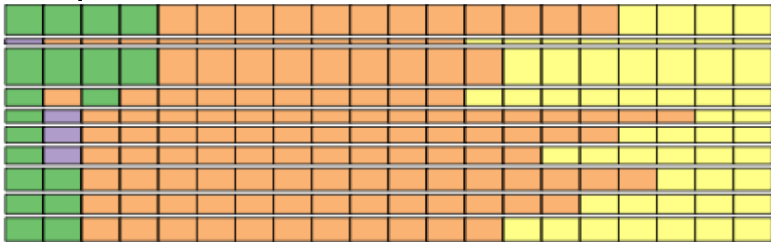


(d) 3-5 year Associate, 8-9.5 year Baccalaureate

Fig. A.17 Representative Sequences, Graduation Group, AA2.

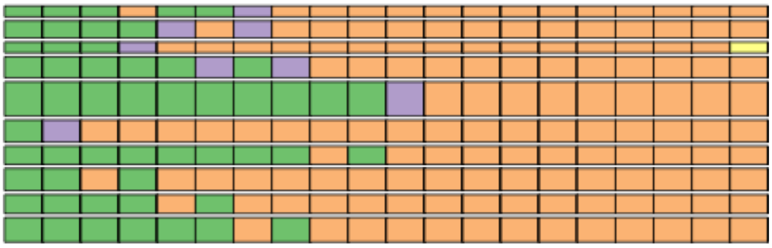


(a) Early Transfers

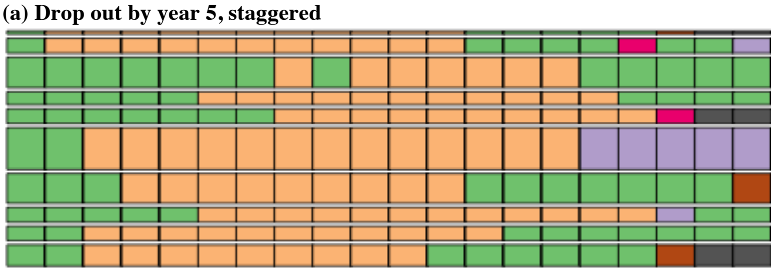


(b) Late Transfers

Fig. A.18 Representative Sequences, Transfer Group, AA2



(a) Drop out by year 5, staggered



(b) Other, Mostly characterized by a large break

Fig. A.19 Representative Sequences, Drop Out and Other Group, AA2.