Measuring All Students: An Alternative Method for Retention and Completion Rates

Jason Bryer
Lisa Daniels
Excelsior College

Abstract

Retention and completion rates are frequently reported as measures of an institutions' success. However, the current method defined by IPEDS is often insufficient for many institutions, especially those who serve transfer students, continuously enroll students, or have non-traditional academic calendars. This paper outlines an alternative method that includes a larger proportion of students (e.g. all first-time enrolled students and alumni). The method relies on warehouse data so to provide a reproducible and transparent approach to calculating retention and completion rates with an emphasis on visualizations. This method also allows for more timely indications of changes in these rates. Lastly, a third measure of persistence is introduced to indicate academic activity for institutions where enrollment does not necessarily indicate academic progress.

Keywords: retention, completion, persistence, student success

Retention and completion rates are important measures of student success within an institution. However, traditional measures are often insufficient or inappropriate for institutions that have continuous enrollment and/or do not serve first-time, full-time students. This is exemplified by the definition of retention rates provided by the National Center for Educational Statistics (NCES) as part of the Integrated Postsecondary Education Data System (IPEDS). It states that retention rate is:

A measure of the rate at which students persist in their educational program at an institution, expressed as a percentage. For four-year institutions, this is the percentage of first-time bachelors (or equivalent) degree-seeking undergraduates from the previous fall who are again enrolled in the current fall. For all other institutions this is the percentage of first-time degree/certificate-seeking students from the previous fall who either re-enrolled or successfully completed their program by the current fall.

Though for institutions such Excelsior College providing retention (and completion) rates according to the above definition is not possible, the concept of retention and completion rates are important measures that necessitate an alternative definition. This document proposes a framework for measuring retention and completion rates that maintain key features of traditional definitions but defines cohorts beyond first-time, full-time students. Specifically the features this framework maintains are:

- Define cohorts that can be followed through to 200% of normal time to completion.
- Provide a measure that is reproducible and transparent.
- Define a measure that is useful for both within and between institutions.

For institutions with continuous enrollment and/or graduation, defining cohorts provide a particular challenge. Whereas traditional institutions have very few discrete starting points (e.g. beginning of a semester), non-traditional institutions may have many starting points with overlapping semesters. It is then natural to define cohorts by a range of enrollment dates. It is important to define such ranges such that each cohort is large enough that measures of central tendency (e.g. mean, median, standard deviation) are reasonable, but also so that students at the extremes within each cohort do not have significantly different retention or completion rates. Excelsior College has approximately 966 (see figure 2) first-time enrollments[1] per month, as such defining a cohort as all students who enroll within a month is reasonable. For subpopulations where cell sizes are too small, cohorts may be combined.

Method

In order to develop an algorithm that provides reproducible results, the calculation of retention and completion rates is based upon warehouse data. Given that student information generally resides in a transactional student information system that may change as new information is obtained, warehouse data provides a "snapshot" of the student information system at a particular point in time. That is, warehouse data is merely a static copy of other data that may be transient. This method for calculating retention and completion rates relies on a warehouse table that is created once a month[2] to coincide with our cohort specification above. Specifically, the warehouse tables contain basic information about all students currently enrolled as of some particular point in time[3]. The resulting warehouse table that begins with the July 2002 cohort contains 3,427,355 records representing 102,428 unique students. Required variables in the table include: a unique student identifier (student id), student's enrollment date, and the degree the student is enrolled in as of that date. Other variables such as gender and ethnicity are only necessary if retention and completion rates broken down by these variables is desired. With these variables along with a list of graduates, the algorithm works as follows (see Figure 1):

---

[1]Cohorts are defined to include students enrolling in the institution for the first time. For example, students who complete more than one degree are only included in the cohort of their first degree enrollment. Moreover, students who enroll, withdraw, and re-enroll remain in their original cohort. That is, if they then complete their degree they contribute to the completion rate for that cohort.

[2]The creation of warehouse data began in February 2009. Warehouse data prior to February 2009 was created retroactively in early 2011 and therefore approximate.

[3]Excelsior College takes a "snapshot" of all currently enrolled students on the 15th of each month. This has the added benefit of coinciding with IPEDS reporting date (i.e. October 15th). Other reports are also

Students

ID	Month	Degree
1	2002-08	AAA
2	2002-08	BBB
3	2002-08	AAA
...		
2	2002-09	BBB
3	2002-09	AAA
4	2002-09	CCC
...		
2	2002-10	AAA
4	2002-10	CCC
5	2002-10	CCC
6	2002-10	AAA
...		
2	2002-11	AAA
3	2002-11	AAA
4	2002-11	CCC
6	2002-11	AAA

→ 1 →

Students

ID	Month	Degree
1	2002-08	AAA
2	2002-08	BBB
3	2002-08	AAA
...		
2	2002-09	BBB
3	2002-09	AAA
4	2002-09	CCC
...		
2	2002-10	BBB
5	2002-10	CCC
6	2002-10	AAA

→ 2 →

Students

ID	Month	Degree
1	2002-08	AAA
2	2002-08	BBB
3	2002-08	AAA
...		
2	2002-09	BBB
4	2002-09	CCC
5	2002-10	CCC
6	2002-10	AAA

→ 3 →

Students

ID	Month	Degree	Transfer
1	2002-08	AAA	FALSE
2	2002-08	BBB	TRUE
3	2002-08	AAA	FALSE
...			
2	2002-09	BBB	TRUE
4	2002-09	CCC	FALSE
...			
5	2002-10	CCC	FALSE
6	2002-10	AAA	FALSE

↓ 4

Graduates

ID	Graduated
5	2002-11

→ 4 →

Students

ID	Month	Degree	Transfer	Graduated
1	2002-08	AAA	FALSE	NA
2	2002-08	BBB	TRUE	NA
3	2002-08	AAA	FALSE	NA
4	2002-09	CCC	FALSE	NA
5	2002-10	CCC	FALSE	2002-11
6	2002-10	AAA	FALSE	NA

↓ 5

Recent Month

ID	Month	Degree
2	2002-11	AAA
3	2002-11	AAA
4	2002-11	CCC
6	2002-11	AAA

→ 5 →

Students

ID	Month	Degree	Transfer	Graduated	Current
1	2002-08	AAA	FALSE	NA	NA
2	2002-08	BBB	TRUE	NA	2002-11
3	2002-08	AAA	FALSE	NA	2002-11
4	2002-09	CCC	FALSE	NA	2002-11
5	2002-10	CCC	FALSE	2002-11	NA
6	2002-10	AAA	FALSE	NA	2002-11

↓ 6 & 7

Students

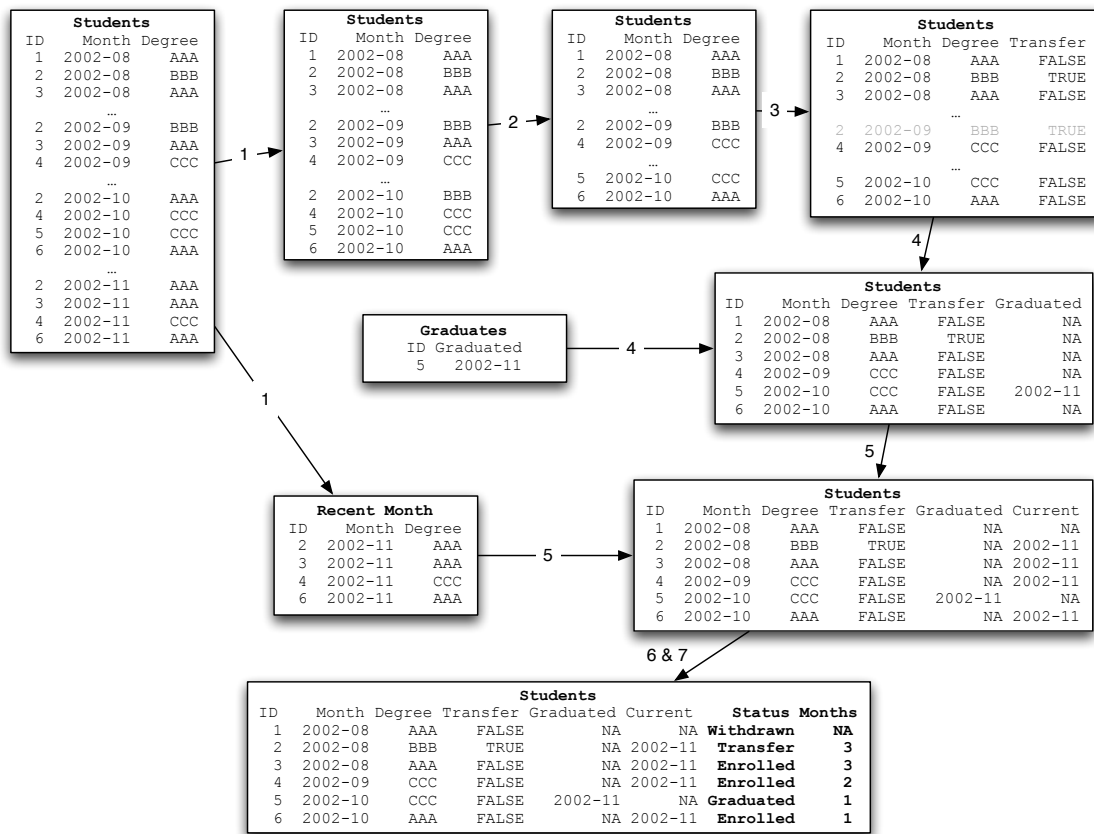| ID	Month	Degree	Transfer	Graduated	Current	Status	Months
1	2002-08	AAA	FALSE	NA	NA	**Withdrawn**	NA
2	2002-08	BBB	TRUE	NA	2002-11	**Transfer**	3
3	2002-08	AAA	FALSE	NA	2002-11	**Enrolled**	3
4	2002-09	CCC	FALSE	NA	2002-11	**Enrolled**	2
5	2002-10	CCC	FALSE	2002-11	NA	**Graduated**	1
6	2002-10	AAA	FALSE	NA	2002-11	**Enrolled**	1

Figure 1. Graphical Representation of the Retention Algorithm

     1. The most recent cohort is moved to a separate table. This will provide the basis of determining which students are still enrolled.

     2. All duplicated combinations of student id and degree are removed. This leaves the earliest instance of each student and degree combination (that is, students who switched degrees will be represented more than once in the table).

     3. A new variable `transferred` is calculated. For students that appear more than once in this table, the value of `transferred` is set to `true`, otherwise `false`.

     4. The table is merged with the graduate table (with a `graduation date` variable) ensuring that all student records remain. The resulting table will contain a new variable with `graduation date`. If that variable is not null, then the student graduated.

     5. The table is merged with the last cohort table (with a `created date` variable) created in step one above. The resulting table contains a new variable. If that variable is not null then the student is still enrolled, otherwise the student withdrew or graduated.

     6. Based on the variables now available, a new factor variable (`student status`)

———

generated based upon this warehouse data such as persistence and demographics.

is created that classifies each student as either still enrolled, enrolled in different degree, graduated, graduated with a different degree than originally enrolled in, or withdrawn.

7. A `months enrolled` variable is calculated based upon the difference between the warehouse date (i.e. cohort) and the reference month (i.e. the most recent warehouse data).

It should be noted that the warehouse data begins in July 2002. To ensure that a student's first a occurrence in the warehouse table is their first enrollment in the college, a baseline table was created that includes all students enrolled prior to July 2002. This baseline table is concatenated to the warehouse table with a null cohort specification.

Retention and completion rates are easily calculated by aggregating the `months enrolled` and `student status` variables. A cross tabulation, divided by the cohort size, provides the rates for each cohort. Retention rates are typically reported at 15 months[4] and completion rates at 150% and 200% percent of normal time-to-completion (36 and 48 months for associate and master's degrees, and 72 and 96 months for baccalaureate degrees, respectively). These rates are calculated using weighted means[5] across all cohorts. That is, the 15-month retention is the weighted mean of each cohort's retention rate at 15 months.

Results

The overall 15-month institutional beginning retention rate as of October 2011 is 74% (this includes 89,515 students in 92 cohorts from August 2002 to March 2010). Table 1 summarizes the beginning retention and completion rates by degree level. Cell sizes are provided in parenthesis and rates are only reported in instances where the cell size is greater than 10.

		Completion Rate			
Degree Level	Retention Rate	36-Months	48-Months	72-Months	96-Months
Associate	70.89 (57210)	27.83 (46944)	32.77 (36675)	34.94 (22949)	38.20 (7418)
Bacc-Master's	54.11 (429)	1.94 (283)	5.63 (224)	4.57 (84)	3.03 (12)
Baccalaureate	80.63 (30693)	50.68 (22722)	55.60 (18376)	60.81 (10995)	65.01 (3460)
Master's	72.28 (1183)	27.35 (720)	38.68 (546)	45.23 (225)	35.66 (71)

Table 1

Aggregated Beginning Retention & Completion Rates by Degree Level

Table 2 provides a summary of cohorts from 15, 36, 48, 72, and 96 months ago. *These are the most recent cohorts that have reached the respective milestones for beginning retention and completion rates.* These rates will be much less stable then the overall, aggregated rates reported above. This is due to the smaller cell sizes as well as the variation between cohorts over time. However, these rates allow for better detection of institutional shifts in beginning retention and completion rates. Moreover, comparing cohorts over time provide some longitudinal perspective. As such, Table 2, as well as other cohort summary tables in this report, include sparklines (Tufte, 2006). Sparklines are "intense, simple, word-sized graphics" (p. 47) that provide an overall sense of the longitudinal changes in data. For our purposes here, each sparkline represents a change in rates across cohorts from older[6] on

---

[4]Excelsior measures retention at 15 months due to its unique enrollment model.

[5]Means are weighted by cohort size.

[6]Sparklines are limited to two years.

the left to most recent on the right. To provide context, a grey band is used to represent the overall results. Specifically, for the retention sparklines, the grey band is the overall weighted mean beginning retention rate +/- (plus and minus) two standard deviations[7]. Similarly, for the completion sparklines, the grey band is the overall weighted mean beginning completion rate at 48-months +/- (plus and minus) two standard deviations. These figures are similar to smaller versions of Figure 3 as discussed below.

| Category | Retention Rate | | Completion Rate | | | | Past Two Years |
		36-Months	48-Months	72-Months	96-Months	
Associate Bacc-Master's	69.02 (510) (6)		21.13 (426) (6)	34.77 (604) (7)	39.68 (499) (3)	40.15 (675) (3)	
Baccalaureate	77.85 (298)		39.10 (266)	56.69 (284)	63.24 (272)	65.74 (324)	
Master's	53.85 (13)		40.00 (20)	50.00 (12)	(7)	(8)	

Table 2

Recent Cohort Beginning Retention & Completion Rates by Degree Level
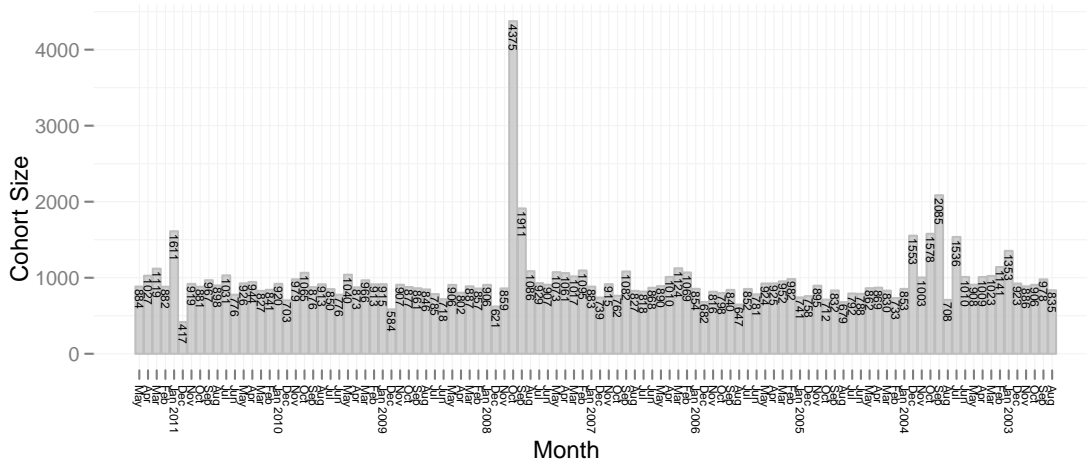
Figure 2. Cohort Sizes: First-Time Enrollments by Month

Visualizing Retention and Completion Rates

This methodology lends itself well to two approaches for visualizing retention and completion rates, namely by individual cohorts and by aggregating cohorts. Figure 3 is a cohort graphic. The x-axis corresponds to each cohort and the y-axis to percentages. That is, values on the y-axis correspond to the percentage of students within that cohort.

---

[7]Under normal distributions approximately 95% of values will lie within two standard deviations of the mean.

Dark-blue bars represent graduates, light-blue represents graduates but of a different degree than they first enrolled, green represents students still enrolled, light-green of students still enrolled but in a different degree than they first enrolled, and pink represents students who withdrew. Vertical lines at 15, 36, 48, 72, and 96 months highlight key milestones for retention and completion. Note that these rates correspond to a single cohort. Figure 2 is a histogram of cohort sizes and provides some context to the cohort graphic.

Figure 4 summarizes beginning retention and completion rates across cohorts. The x-axis represents months since enrollment and the y-axis, like the cohort graphic described above, corresponds to percentages. Each color represents a cohort. The solid lines correspond to beginning retention over time and the dashed lines to completion over time. The bars (dark grey for completion and light grey for retention) then represent the weighted means across all cohorts. That is, these bars provide a summary of institutional beginning retention and completion rates based upon time since enrollment. Vertical lines at 15, 36, 48, 72, and 96 months highlight key milestones for retention and completion but unlike the cohort graphic, these values are weighted means across all cohorts. Note that the instability to the right of the graph is an artifact of the decreasing number of cohorts, and therefore decreasing number of students, contributing to the average rates.

Longitudinal Rates

This methodology lends itself to multiple ways of examining longitudinal trends. Figures 8, 9, and 10 depict longitudinal trends using individual cohorts by degree level (note that the black line in Figure 8 represents the institutional beginning retention rate). However, it is typical to examine trends by fiscal year. Table 4 and Figure 6 provide 150% completion rates by degree level across fiscal years. Note that the Baccalaureate-Master's programs have been ommitted from Figures 8, 9, and 10 due to the relatively small cohort sizes and the resulting appareant volitality of rates.

	Associate	Bacc-Master's	Baccalaureate	Master's
2003	79.80 (7337)	(13)	83.13 (3598)	70.42 (71)
2004	75.52 (9493)	86.21 (29)	83.40 (3897)	76.36 (55)
2005	75.28 (5951)	55.32 (47)	84.39 (3965)	74.49 (98)
2006	71.89 (6527)	(52)	78.48 (3787)	75.54 (139)
2007	70.30 (7223)	59.00 (100)	81.13 (3758)	76.09 (230)
2008	65.29 (10347)	45.00 (60)	79.81 (4299)	71.20 (184)
2009	63.98 (5764)	54.84 (62)	75.51 (4243)	66.21 (219)
2010	65.08 (4568)	56.06 (66)	79.88 (3146)	66.31 (187)

Table 3

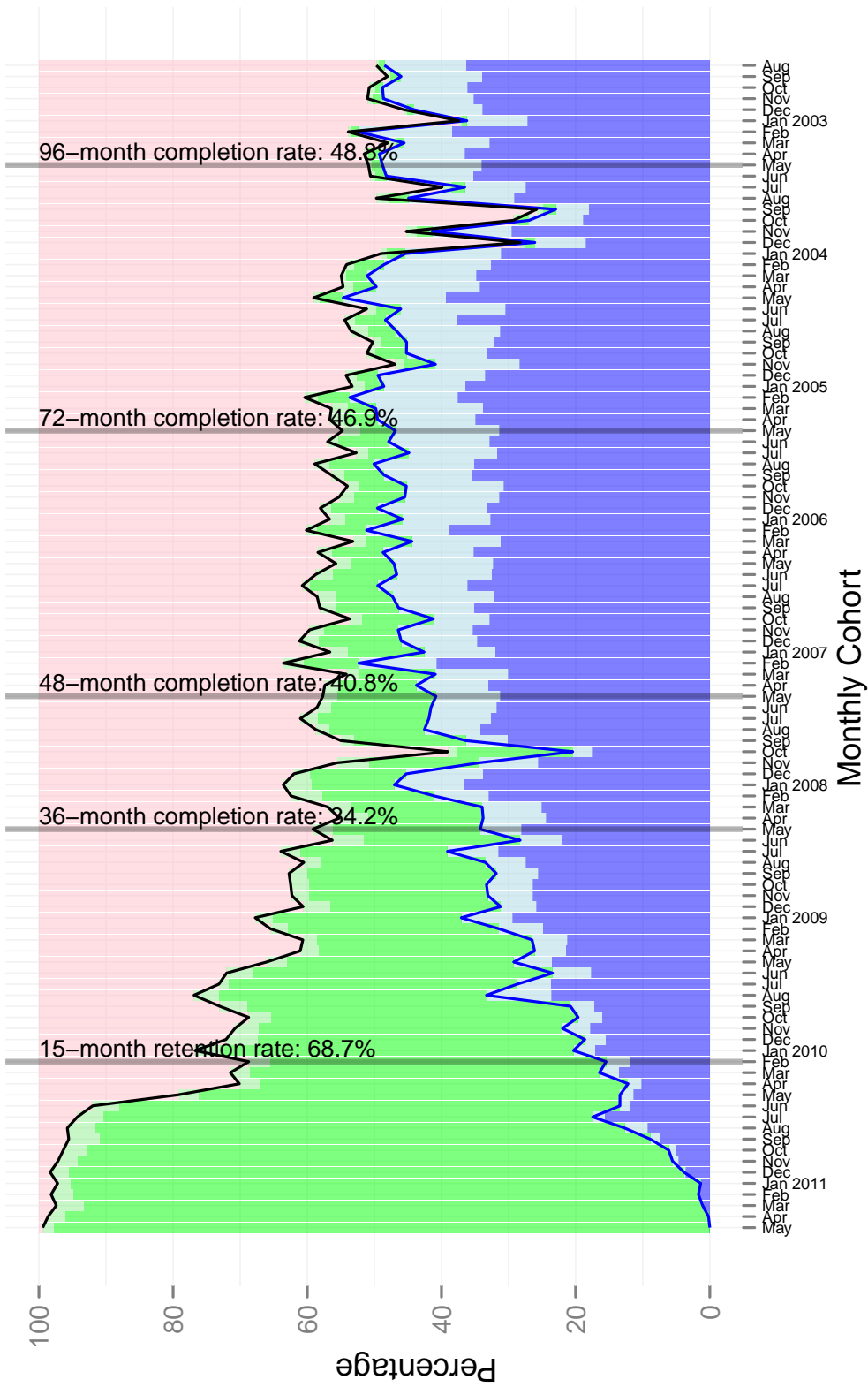*Beginning Retention Rates by Fiscal Year and Degree Level*

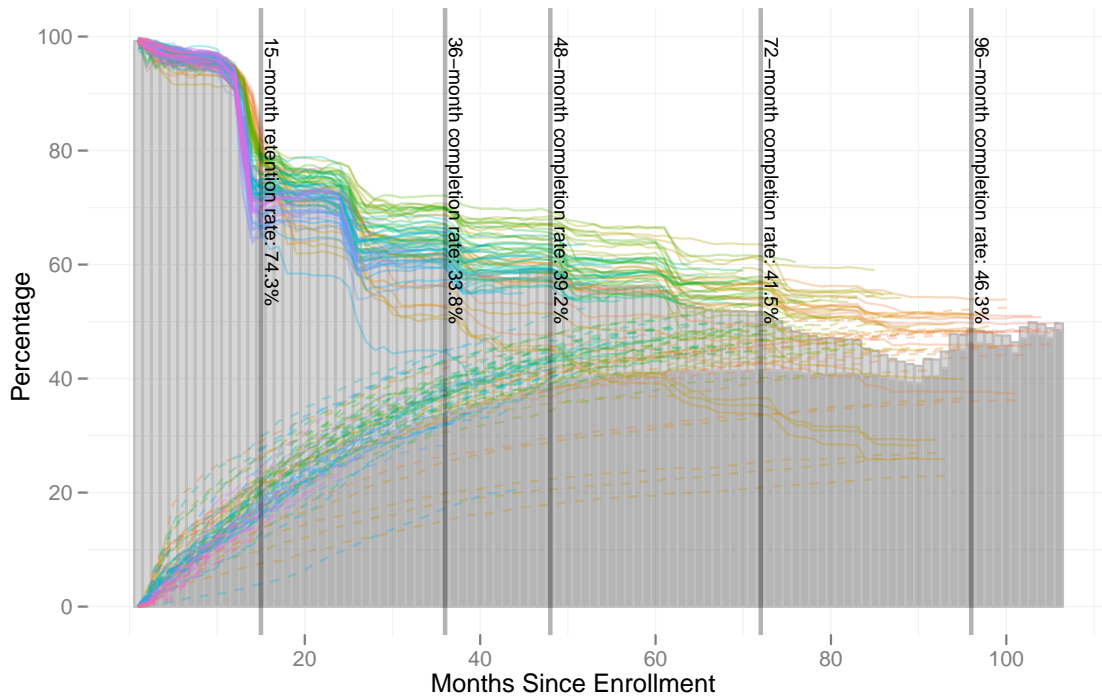*Figure 3.* Institutional Beginning Retention by Cohort
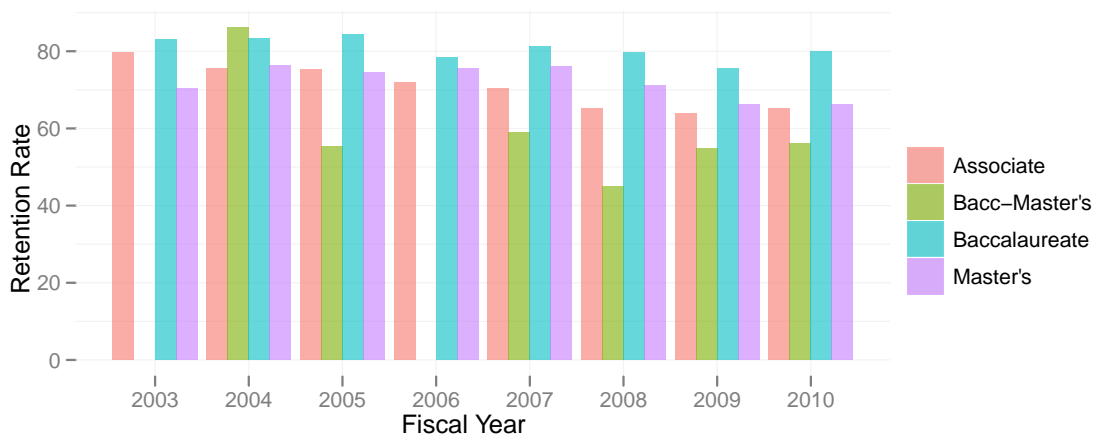
Figure 4. Institutional Beginning Retention

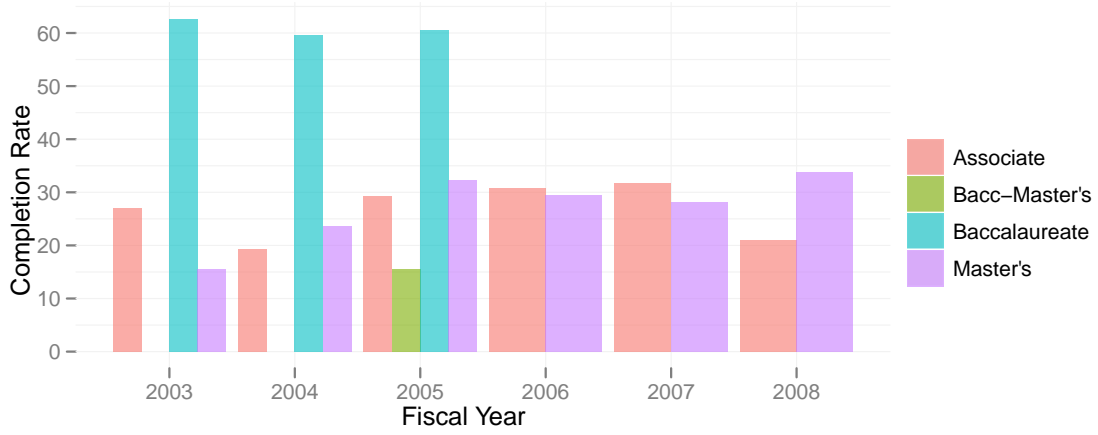*Figure 5*. Beginning Retention Rates by Fiscal Year & Degree Level

Figure 6. 150% Beginning Completion Rates by Fiscal Year & Degree Level

	Associate	Bacc-Master's	Baccalaureate	Master's
2003	27.12 (7339)	(12)	62.64 (3472)	15.49 (71)
2004	19.31 (9498)	(27)	59.66 (3748)	23.64 (55)
2005	29.32 (5983)	15.56 (45)	60.53 (3775)	32.32 (99)
2006	30.87 (6556)			29.55 (132)
2007	31.82 (7222)			28.19 (188)
2008	21.04 (10346)			33.71 (175)

Table 4

Beginning Completion Rates by Fiscal Year (Three-year rate for Associate and Master's, Six-year rate for Baccalaureate and Bacc-Master's)
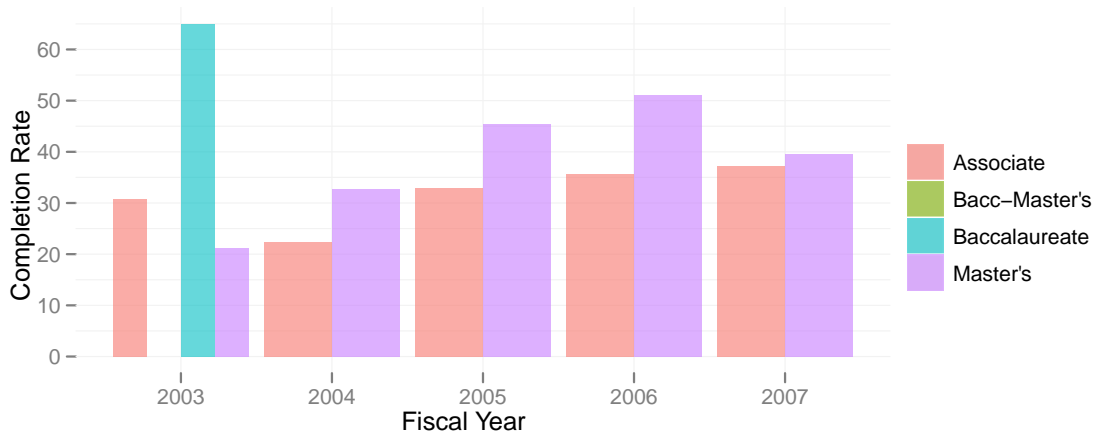
Figure 7. 200% Beginning Completion Rates by Fiscal Year & Degree Level

	Associate	Bacc-Master's	Baccalaureate	Master's
2003	30.86 (7353)	(12)	65.00 (3460)	21.13 (71)
2004	22.36 (9527)			32.73 (55)
2005	32.90 (5978)			45.45 (99)
2006	35.63 (6561)			51.15 (131)
2007	37.16 (7256)			39.47 (190)

Table 5

Beginning Completion Rates by Fiscal Year (Four-year rate for Associate and Master's, Eight-year rate for Baccalaureate and Bacc-Master's)
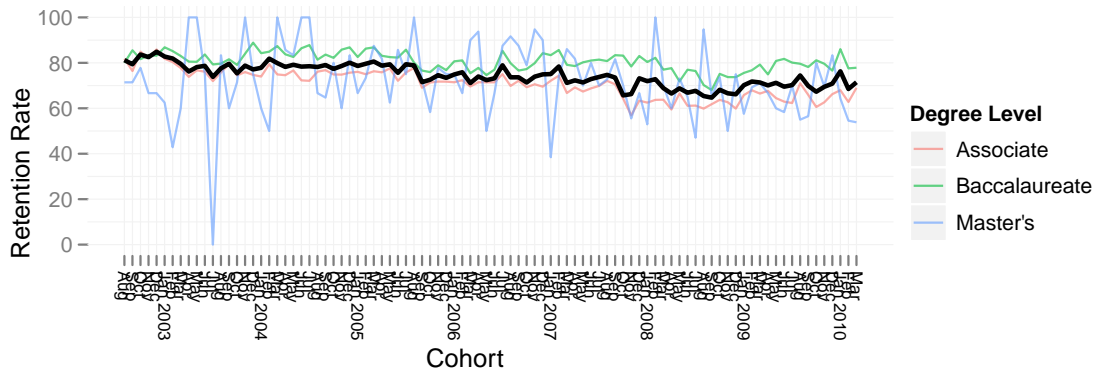
Figure 8. Beginning Retention Rates Across Cohorts
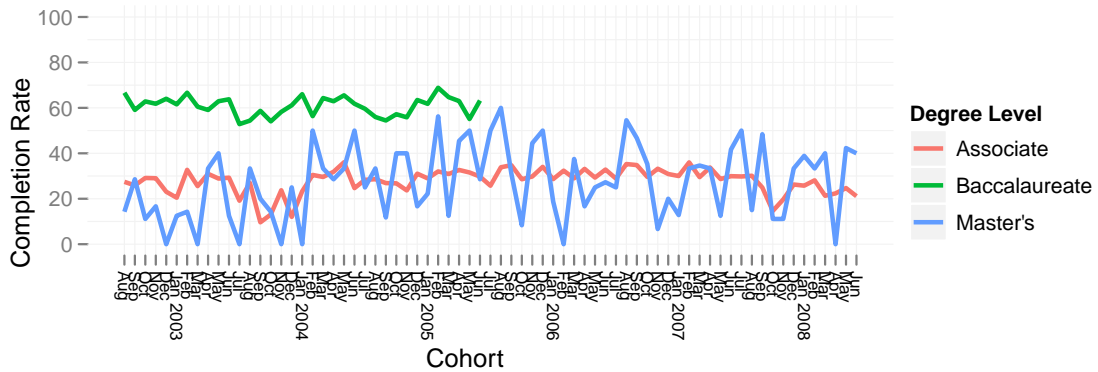
Figure 9. 150% Beginning Completion Rates Across Cohorts
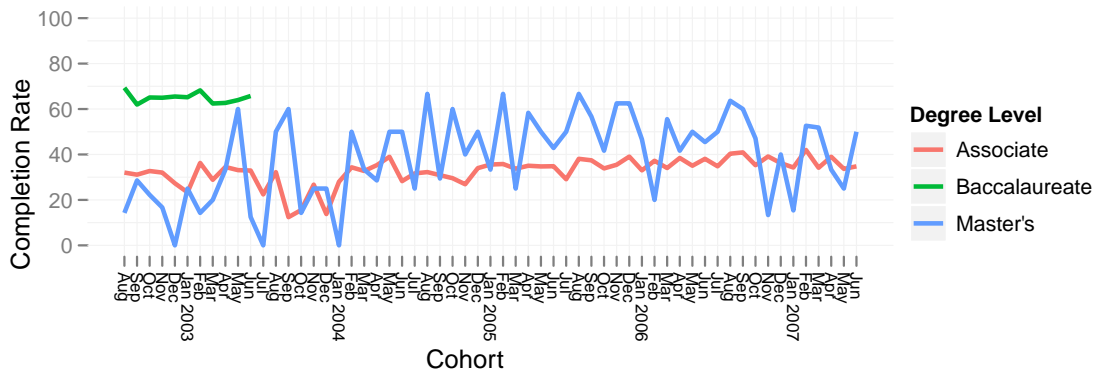
Figure 10. 200% Beginning Completion Rates Across Cohorts

Academically Active Students

Beginning retention and completion rates are important measures of student activity within an institution, however for institutions where enrollment does not necessarily imply academic progress towards a degree, these measures may be insufficient on their own. Persistence rates are a common approach to measuring the relative academic activity of students. At Excelsior College, an enrolled student is considered persistent if they have engaged in at least one of the following activities within the past six months:

- Attempted an Excelsior College course
- Attempted an Excelsior College exam (this includes portfolio assessment and military students)
- Transferred in credit
- Participated in an online conference or CPNE Workshop (Nursing)
- Participated in a preceptorship (select portion of Nursing)
- Registered for the CPNE (within 9 months)
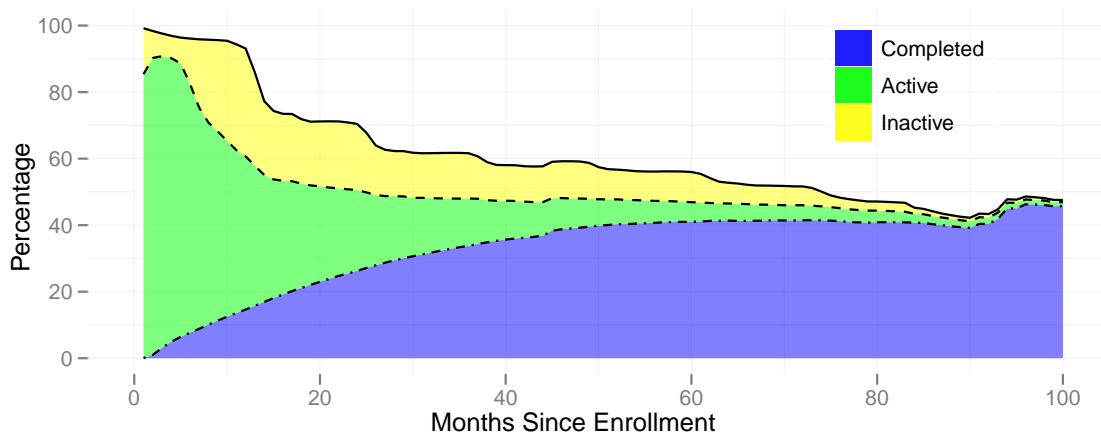- Completed the CCS100 Excelsior College Student Experience

Figure 11. Beginning Completion, Retention, & Academically Active Rates

Persistence rates are calculated based upon all enrolled students at a particular time using the same warehouse data as described above. What is missing with this approach is how student persistence changes over the course of their enrollment. Figure 11 is a modified version of Figure 4 that separates retained students based upon whether they are academically active or not. Specifically, the blue region corresponds to students who have completed their degree, the solid black line corresponds to the retention rate, the shaded green area corresponds to students who are still enrolled and academically active, and the shaded yellow area corresponds to students who are enrolled but not academically active. Figure 12 provides an alternate view of these data. The solid line corresponds the percentage of students still enrolled, the green line is percentage of students who are academically active of those still enrolled, the light gray line also represents academically active students but is based upon the full cohort size, and lastly, the current months institutional persistence rate

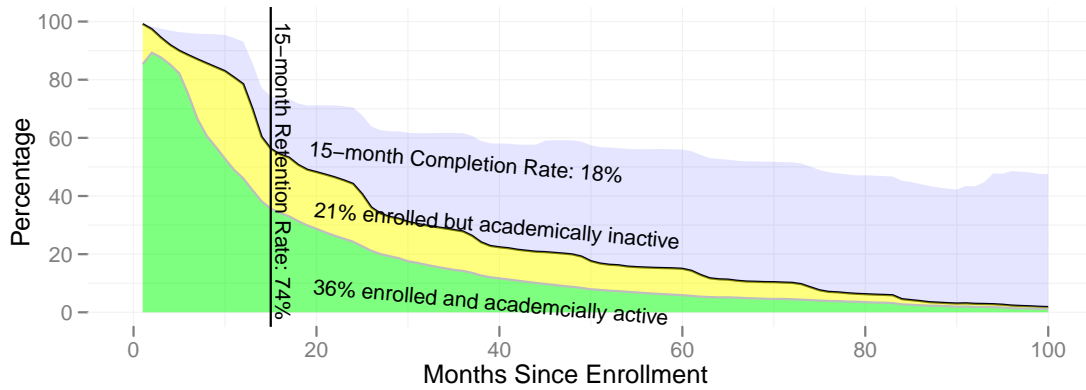is represented by the horizontal line to provide context.

Figure 12. Academically Active Rates
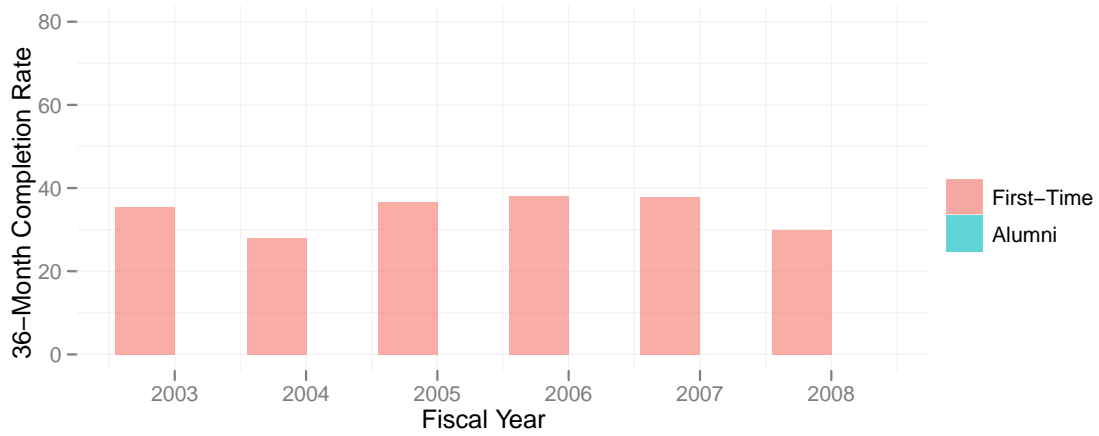
Retention & Completion Rates for Alumni

The primary focus of this analysis has been on first-time to Excelsior College students. However, alumni are an important population to consider. Of the 142084 students who have graduated from Excelsior College, 7392 have returned since July 2002 to attempt a second degree. In total, 7663 students have earned more than one degree. Table 6 provides 36-month completion rates for first-time to Excelsior College students compared with alumni by fiscal year.

FY	First-Time	Alumni
2003	35.47 (10937)	(357)
2004	27.95 (13396)	(437)
2005	36.67 (10000)	(490)
2006	38.02 (10423)	(575)
2007	37.78 (11086)	(712)
2008	29.93 (14827)	(815)

Table 6
36-Month Completion Rates for First-Time & Alumni Students by Fiscal Year

Conclusions

Retention and completion are important measures for an institution to consider. However, traditional measures potentially exclude substantial portions of an institutions' student population, especially for non-traditional institutions with continuous enrollment and/or enroll students with transfer credit. The method outlined here provides an approach that includes virtually all students who attend an institution while retaining the original spirit of

Figure 13. 36-Month Completion Rates for First-Time & Alumni Students

the traditional measures. Moreover, through the use of modern graphic techniques further insights can be achieved beyond a single numeric metric.

References

R Development Core Team. (2011). *R: a language and environment for statistical computing.* ISBN 3-900051-07-0. R Foundation for Statistical Computing. Vienna, Austria. Retrieved from R Foundation for Statistical Computing: `http://www.R-project.org/`

Tufte, E. R. (2006). *Beautiful evidence.* Cheshire, CT: Graphics Press.

Wickham, H. (2011). *Devtools: tools to make developing r code easier.* R package version 0.4. Retrieved from `http://CRAN.R-project.org/package=devtools`

Appendix
Implemented for Reproducible Research

As indicated at the outset, a major goal for the method outlined here is to provide transparency and reproducibility. The algorithms outlined in this paper are implemented as a package for the open source statistical program R (R Development Core Team, 2011) and are available at `http://github.com/jbryer/irutils`. The latest development version of the `irutils` package can be installed from Github using the `devtools` (Wickham, 2011) package:

```
> library(devtools)
> install_github('irutils', 'jbryer')
```

Once installed, the `irutils` package can be loaded with the `library` command. There are two functions that perform the retention and completing rate calculations. The `cohortRetention` function returns the rates for each cohort based upon the most recent cohort whereas the `retention` function returns the rates aggregated across all cohorts.

```
> library(irutils)
> str(irutils::cohortRetention)

function (students, grads, gradColumn = "START_DATE", grouping = NULL)

> str(irutils::retention)

function (students, grads, ...)

> ecCohortRetention = cohortRetention(students, graduates)
> retention = retention(students, graduates)
```

The structure of the returned data frames are provided below.

```
> str(ecCohortRetention)

'data.frame':          106 obs. of  10 variables:
 $ Cohort         : Factor w/ 106 levels "2002-08","2002-09",..: 1 2 3 4 5 6 7 8 9 10 ...
 $ Graduated      : num  36.3 33.9 36.1 35.2 33.9 ...
 $ Graduated Other: num  12.2 12.1 12.7 13.4 10.2 ...
 $ Still Enrolled : num  0.838 1.534 1.104 1.58 0.975 ...
 $ Transferred    : num  0.359 0.511 0.883 0.79 0.542 ...
 $ Withdrawn      : num  50.3 51.9 49.2 49 54.4 ...
 $ GraduationRate : num  48.5 46 48.8 48.6 44.1 ...
 $ RetentionRate  : num  49.7 48.1 50.8 51 45.6 ...
 $ PersistenceRate: num  80 70 61.1 61.9 64.3 ...
 $ Enrollments    : int  835 978 906 886 923 1353 1141 1023 1009 908 ...

> str(ecRetention)
```

```
'data.frame':          5671 obs. of  11 variables:
 $ Cohort         : Factor w/ 106 levels "2002-08","2002-09",..: 1 2 3 4 5 6 7 8 9 10 ...
 $ GraduationRate : num  48.5 46 48.8 48.6 44.1 ...
 $ RetentionRate  : num  49.7 48.1 50.8 51 45.6 ...
 $ PersistenceRate: num  80 70 61.1 61.9 64.3 ...
 $ Enrollments    : int  835 978 906 886 923 1353 1141 1023 1009 908 ...
 $ Graduated      : num  36.3 33.9 36.1 35.2 33.9 ...
 $ Graduated Other: num  12.2 12.1 12.7 13.4 10.2 ...
 $ Still Enrolled : num  0.838 1.534 1.104 1.58 0.975 ...
 $ Transferred    : num  0.359 0.511 0.883 0.79 0.542 ...
 $ Month          : num  106 105 104 103 102 101 100 99 98 97 ...
 $ Comparison     : chr  "2011-06-15" "2011-06-15" "2011-06-15" "2011-06-15" ...
```