



# SEGMENTATION MODEL FOR ASSESSING COURSE-TAKING PATTERNS

RESEARCH METHODOLOGY  
AND DISCUSSION GUIDE

*This research guide was written by Peter Riley Bahr, Kathy Booth and Terrence Willett, with support from Rob Johnstone and Gregory Stoup.*

*It was underwritten by LearningWorks. For more information, visit [www.rpgroup.org/projects/completion](http://www.rpgroup.org/projects/completion)*

In 2010, Peter Riley Bahr (Assistant Professor of Education at the University of Michigan) conducted a cluster analysis for the California Community Colleges Chancellor's Office that examined the course-taking behavior of first-time students over an eight-year period.<sup>1</sup> The research revealed a number of interesting issues, including under-reporting successes that do not result in "completion" (meaning transfer, an associate's degree or certificate), equity gaps in students pursuing completion outcomes, and the high volume of units attempted by students pursuing a completion goal. (Please visit [www.rpgroup.org/projects/completion](http://www.rpgroup.org/projects/completion) to read articles about Bahr's analysis and to access an inquiry guide that applies this research to the current focus on increasing completion.)

While a college research office can conduct its own cluster analysis based upon local data to replicate Bahr's study, doing so is a time-intensive effort. As a more expedient alternative, this document provides a simplified rule set to sort students into the classifications identified by Bahr. This document also includes sample discussion questions on how to use these results to build a deeper understanding of student course-taking behavior and its relationship to student success.

LearningWorks was founded by the Career Ladders Project for California Community Colleges, the Research and Planning Group for California Community Colleges (RP Group), and the California Community Colleges Success Network (3CSN) to facilitate, disseminate and fund practitioner-informed recommendations for changes at the system and classroom levels, infusing these strategies with statewide and national insights.

## ***Cluster Names***

Bahr's study identified six clusters. For the purpose of this rule set, the clusters have been renamed to help bring forward the completion-related aspects of each group, which are described in "What's Completion Got to Do with It? Using Course-Taking Behavior to Understand Community College Success" available at [www.rpgroup.org/projects/completion](http://www.rpgroup.org/projects/completion). The simplified rule set also required that two of the original clusters be collapsed into one group to increase the accuracy of identification. The original and modified cluster names are:

- (1) The Completion-Likely group includes the Most Likely to Complete cluster (which Bahr calls the Transfer cluster) and Somewhat Likely to Complete cluster (which Bahr calls the Exploratory cluster)
- (2) Bahr's Terminal Vocational cluster is referred to as the Career Technical Education (CTE) group in this document
- (3) Bahr's Experimental cluster is referred to as the Completion-Unlikely group
- (4) Bahr's Drop-In cluster is referred to as the Skills-Builder group
- (5) The final group, Noncredit, is given the same name in both the original study and this document

---

<sup>1</sup>Bahr, P. R. (2010). The bird's eye view of community colleges: A behavioral typology of first time students based on cluster analytic classification. *Research in Higher Education*, 51, 724-749 and Bahr, P. R. (2011). A typology of students' use of the community college. *New Directions for Institutional Research*, S1, 33-48.

## File Preparation

Prepare a file of first time students, tracked for two years, with at least the following variables:

Variable Name	Description
number_terms_enrolled	Count of terms in which a student was enrolled (fall, spring, and summer, but not winter intersessions; quarter system colleges would include winter quarters)
course_success	Percentage of courses in which student earned a grade of C- or better
mean_unitload_regterms	Mean number of units attempted during primary terms (fall and spring semesters or fall, winter, and spring quarters) with at least one course enrollment
ua_vocnontrx	Units attempted in all non-transferable CTE courses (CB05 = C and CB09 = A,B,C)
noncred	Count of non-credit courses attempted

To gain further insight into the relationships between student classifications and other aspects of course-taking behavior, additional variables to consider include, but are not limited to:

>> Total units attempted	>> Number of units attempted in degree applicable non-transferable units
>> Total units earned	>> Level of first basic skills course taken
>> Number of units attempted in transferable math	>> Term to term retention
>> Number of units attempted in transferable English	>> Degree or certificate completion
>> Number of units attempted in transferable physical and life sciences	>> Transfer to university
>> Number of units attempted in transferable behavioral and social science	>> Ethnicity
>> Number of units attempted in other transferable units	>> Gender
>> Number of units attempted in occupational transferable units	>> Age
	>> Financial aid status
	>> Citizenship

The recommended file to use as a basis is the First File available on the California Community College Chancellor's Office (CCCCO) Data on Demand (DOD) site. The First File has several different "flavors" of first.

The recommended type of first to use shows the first post-high school enrollment at any college. The fields for that flag are as follows:

NSA stands for 'not special admit' and filters out concurrent high school enrollments. To select for students who are first time at your college, set FIRST\_TERM\_NSA\_LOC to equal '1'.

FIRST_TERM_NSA =	The first term the student met the FIRST_TERM criteria as a non-Special Admit Student (Education Status SB11 not equal to '10000')
FIRST_TERM_NSA_LOC =	L – Met the FIRST_TERM_NSA criteria at this college. O – Met the FIRST_TERM_NSA criteria at another CCC. X – Never met the criteria at a CCC.

More information on the DOD site can be found at: <http://www.rpgroup.org/resources/california-community-college-chancellors-office-data-mart-2>. Information on the First File and a comparison to other data sources can be found at <http://www.rpgroup.org/resources/examining-two-resources-cohort-transfer-tracking>

## for Applying Student Classifications for Two-Year Cohorts

After the student file is prepared, apply the rules in this section to classify your students. **Note that the rules depend on precedence. That is, the order in which rules are applied will influence the findings.**

These rules were created originally from semester-based colleges. A suggested conversion of the script for quarter-based colleges is as follows below. **Note that this conversion from semester-based rules to quarter-based rules has not been field tested.**

1. Semester units were multiplied by 1.5 to convert to quarter units
2. Number of terms were translated as follows:
  - (a) 1 quarter = 1 semester
  - (b) 2 or 3 quarters = 2 semesters
  - (c) 4 or 5 quarters = 3 semesters
  - (d) 6 quarters = 4 semesters
  - (e) 7 or 8 quarters = 5 semesters
  - (f) 9 quarters = 6 semesters

### *Caveat*

In testing, this rule set correctly identified students' classifications nearly 80% of the time. One should keep in mind, however, that the rule set provided here is intended for classifying students in the aggregate (as a group) and should not be used to interpret the course-taking and enrollment behavior of individual students.

### *Feedback*

The RP Group would like to hear about your experience implementing these rules and disseminating the findings, as well as any insights gained by your college community. Please send your comments and suggestions to:

Terrence Willett, Senior Researcher, RP Group  
twillett@rpgroup.org  
(831) 461-4586

# semester-based college rule set in SPSS

if (number\_terms\_enrolled = 1 & course\_success >= 0.65) classification = 1.

if (number\_terms\_enrolled = 1 & course\_success < 0.65) classification = 3.

if (number\_terms\_enrolled = 2 & course\_success >= 0.65 & mean\_unitload\_regterms < 8) classification = 1.

if (number\_terms\_enrolled = 2 & course\_success < 0.65) classification = 3.

if (number\_terms\_enrolled = 2 & course\_success >= 0.65 & mean\_unitload\_regterms >= 8) classification = 4.

if (number\_terms\_enrolled = 3 & course\_success >= 0.65 & mean\_unitload\_regterms < 6) classification = 1.

if (number\_terms\_enrolled = 3 & course\_success < 0.65) classification = 3.

if (number\_terms\_enrolled = 3 & course\_success >= 0.65 & mean\_unitload\_regterms >= 6 & ua\_vocnontrx < 3) classification = 4.

if (number\_terms\_enrolled = 3 & course\_success >= 0.65 & mean\_unitload\_regterms >= 6 & ua\_vocnontrx >= 3) classification = 5.

if (number\_terms\_enrolled = 4 & ua\_vocnontrx < 3) classification = 4.

if (number\_terms\_enrolled = 4 & ua\_vocnontrx >= 3) classification = 5.

if (number\_terms\_enrolled = 4 & noncred >= 4 & mean\_unitload\_regterms < 6) classification = 2.

if (number\_terms\_enrolled = 5 & ua\_vocnontrx < 3) classification = 4.

if (number\_terms\_enrolled = 5 & ua\_vocnontrx >= 3) classification = 5.

if (number\_terms\_enrolled = 5 & noncred >= 5 & mean\_unitload\_regterms < 6) classification = 2.

if (number\_terms\_enrolled = 6 & ua\_vocnontrx < 3) classification = 4.

if (number\_terms\_enrolled = 6 & ua\_vocnontrx >= 3) classification = 5.

if (number\_terms\_enrolled = 6 & noncred >= 6 & mean\_unitload\_regterms < 6) classification = 2.

value labels classification 1 'skills-builder' 2 'noncredit' 3 'completion-unlikely' 4 'completion-likely' 5 'CTE'.

EXECUTE.

# semester-based college rule set in STATA

```
#delimit ;

generate classification = .;

label define classification 1 "skills-builder" 2 "noncredit" 3 "completion-unlikely" 4 "completion-likely" 5 "CTE";

label values classification classification;

replace classification = 1 if (number_terms_enrolled == 1) & (course_success >= 0.65);

replace classification = 3 if (number_terms_enrolled == 1) & (course_success < 0.65);

replace classification = 1 if (number_terms_enrolled == 2) & (course_success >= 0.65) & (mean_unitload_regterms < 8);

replace classification = 3 if (number_terms_enrolled == 2) & (course_success < 0.65);

replace classification = 4 if (number_terms_enrolled == 2) & (course_success >= 0.65) & (mean_unitload_regterms >= 8);

replace classification = 1 if (number_terms_enrolled == 3) & (course_success >= 0.65) & (mean_unitload_regterms < 6);

replace classification = 3 if (number_terms_enrolled == 3) & (course_success < 0.65);

replace classification = 4 if (number_terms_enrolled == 3) & (course_success >= 0.65) & (mean_unitload_regterms >= 6) & (ua_vocnontrx < 3);

replace classification = 5 if (number_terms_enrolled == 3) & (course_success >= 0.65) & (mean_unitload_regterms >= 6) & (ua_vocnontrx >= 3);

replace classification = 4 if (number_terms_enrolled == 4) & (ua_vocnontrx < 3);

replace classification = 5 if (number_terms_enrolled == 4) & (ua_vocnontrx >= 3);

replace classification = 2 if (number_terms_enrolled == 4) & (noncred >= 4) & (mean_unitload_regterms < 6);

replace classification = 4 if (number_terms_enrolled == 5) & (ua_vocnontrx < 3);

replace classification = 5 if (number_terms_enrolled == 5) & (ua_vocnontrx >= 3);

replace classification = 2 if (number_terms_enrolled == 5) & (noncred >= 5) & (mean_unitload_regterms < 6);

replace classification = 4 if (number_terms_enrolled == 6) & (ua_vocnontrx < 3);

replace classification = 5 if (number_terms_enrolled == 6) & (ua_vocnontrx >= 3);

replace classification = 2 if (number_terms_enrolled == 6) & (noncred >= 6) & (mean_unitload_regterms < 6);
```

# quarter-based college rule set in SPSS

if(number\_terms\_enrolled = 1 & course\_success >= 0.65) classification = 1.

if(number\_terms\_enrolled = 1 & course\_success < 0.65) classification = 3.

if((number\_terms\_enrolled = 2 | number\_terms\_enrolled = 3) & course\_success >= 0.65 & mean\_unitload\_regterms < 12) classification = 1.

if((number\_terms\_enrolled = 2 | number\_terms\_enrolled = 3) & course\_success < 0.65) classification = 3.

if((number\_terms\_enrolled = 2 | number\_terms\_enrolled = 3) & course\_success >= 0.65 & mean\_unitload\_regterms >= 12) classification = 4.

if((number\_terms\_enrolled = 4 | number\_terms\_enrolled = 5) & course\_success >= 0.65 & mean\_unitload\_regterms < 9) classification = 1.

if((number\_terms\_enrolled = 4 | number\_terms\_enrolled = 5) & course\_success < 0.65) classification = 3.

if((number\_terms\_enrolled = 4 | number\_terms\_enrolled = 5) & course\_success >= 0.65 & mean\_unitload\_regterms >= 9 & ua\_vocnontrx < 4.5) classification = 4.

if((number\_terms\_enrolled = 4 | number\_terms\_enrolled = 5) & course\_success >= 0.65 & mean\_unitload\_regterms >= 9 & ua\_vocnontrx >= 4.5) classification = 5.

if(number\_terms\_enrolled = 6 & ua\_vocnontrx < 4.5) classification = 4.

if(number\_terms\_enrolled = 6 & ua\_vocnontrx >= 4.5) classification = 5.

if(number\_terms\_enrolled = 6 & noncred >= 4 & mean\_unitload\_regterms < 9) classification = 2.

if((number\_terms\_enrolled = 7 | number\_terms\_enrolled = 8) & ua\_vocnontrx < 4.5) classification = 4.

if((number\_terms\_enrolled = 7 | number\_terms\_enrolled = 8) & ua\_vocnontrx >= 4.5) classification = 5.

if((number\_terms\_enrolled = 7 | number\_terms\_enrolled = 8) & noncred >= 5 & mean\_unitload\_regterms < 9) classification = 2.

if(number\_terms\_enrolled = 9 & ua\_vocnontrx < 4.5) classification = 4.

if(number\_terms\_enrolled = 9 & ua\_vocnontrx >= 4.5) classification = 5.

if(number\_terms\_enrolled = 9 & noncred >= 6 & mean\_unitload\_regterms < 9) classification = 2.

value labels classification 1 'skills-builder' 2 'noncredit' 3 'completion-unlikely' 4 'completion-likely' 5 'CTE'.

EXECUTE.

# quarter-based college rule set in STATA

```
#delimit;

generate classification = ;

label define classification 1 "skills-builder" 2 "noncredit" 3 "completion-unlikely" 4 "completion-likely" 5 "CTE";

label values classification classification;

replace classification = 1 if number_terms_enrolled == 1 & course_success >= 0.65;

replace classification = 3 if number_terms_enrolled == 1 & course_success < 0.65;

replace classification = 1 if (number_terms_enrolled == 2 | number_terms_enrolled == 3) & course_success >= 0.65 & mean_unitload_regterms < 12;

replace classification = 3 if (number_terms_enrolled == 2 | number_terms_enrolled == 3) & course_success < 0.65;

replace classification = 4 if (number_terms_enrolled == 2 | number_terms_enrolled == 3) & course_success >= 0.65 & mean_unitload_regterms >= 12;

replace classification = 1 if (number_terms_enrolled == 4 | number_terms_enrolled == 5) & course_success >= 0.65 & mean_unitload_regterms < 9;

replace classification = 3 if (number_terms_enrolled == 4 | number_terms_enrolled == 5) & course_success < 0.65;

replace classification = 4 if (number_terms_enrolled == 4 | number_terms_enrolled == 5) & course_success >= 0.65 & mean_unitload_regterms >= 9 & ua_vocnontrx < 4.5;

replace classification = 5 if (number_terms_enrolled == 4 | number_terms_enrolled == 5) & course_success >= 0.65 & mean_unitload_regterms >= 9 & ua_vocnontrx >= 4.5;

replace classification = 4 if number_terms_enrolled == 6 & ua_vocnontrx < 4.5;

replace classification = 5 if number_terms_enrolled == 6 & ua_vocnontrx >= 4.5;

replace classification = 2 if number_terms_enrolled == 6 & noncred >= 6 & mean_unitload_regterms < 9;

replace classification = 4 if (number_terms_enrolled == 7 | number_terms_enrolled == 8) & ua_vocnontrx < 4.5;

replace classification = 5 if (number_terms_enrolled == 7 | number_terms_enrolled == 8) & ua_vocnontrx >= 4.5;

replace classification = 2 if (number_terms_enrolled == 7 | number_terms_enrolled == 8) & noncred >= 5 & mean_unitload_regterms < 9;

replace classification = 4 if number_terms_enrolled == 9 & ua_vocnontrx < 4.5;

replace classification = 5 if number_terms_enrolled == 9 & ua_vocnontrx >= 4.5;

replace classification = 2 if number_terms_enrolled == 9 & noncred >= 6 & mean_unitload_regterms < 9;
```



## discussion questions

Bahr's student clusters can be used to highlight common student pathways through institutions, assess the comprehensiveness of success measures and identify factors that are associated with varying degrees of completion—meaning the achievement of a degree, certificate or transfer. Below are examples of ways to focus research results around each of these topics, followed by a series of potential discussion questions.

### *Understanding Student Pathways*

Draw a series of maps of the pathways that each student cluster is likely to take through your institution. Then lead a conversation with questions such as:

- (1) How do the pathways differ among the different clusters?
- (2) Who are the students in each cluster? Are there several clear subgroups within each cluster? Is it apparent what each of these groups is trying to achieve?
- (3) How do our college's policies add momentum or create barriers for students achieving their goals? Do these policies affect how long students stay at our institution? Are there different momentum points and barriers for different clusters?
- (4) What interim measures would help us understand students' progress toward completion and likely points where students drop out? What external factors (e.g., full-time versus part-time enrollment, term-to-term persistence) should we evaluate? What internal factors (e.g., motivation, hope) could we assess?

### *Investigating Non-Completion Successes*

If your college has a large number of skills-builder or non-credit students, determine what types of courses they are taking. Then lead a conversation with questions such as:

- (1) By examining the courses these clusters frequently take, is it apparent which skills they are attempting to build? What percentages of students are pursuing workforce training, life-long learning or basic skills?
- (2) Is there a way our institution could capture these

successes? How could we measure or quantify their activities?

- (3) Are there factors in our community that make non-completion pathways a priority?
- (4) Are there ways to help students achieve short-term successes, such as stackable certificates, and connect them with longer-term pathways, such as preparing for transfer to a four-year program?

### *Examining the Gap in Completion Attainment*

If your college has a large number of completion-unlikely students, gather additional information about the characteristics that differ between this group and completion-likely students, such as participation in student support programs, course-taking patterns in the first semester or basic skills placement scores. Then lead a conversation with questions such as:

- (1) What are the differences between completion-likely and completion-unlikely students? Are there clear subgroups within each cluster? What else do we need to know to better understand who is more likely to be in each group?
- (2) What should we be doing to determine if students are making progress toward their goals? Should the college intervene in some way between a student's first and second semester? Are there actions that should be taken during a student's first semester?
- (3) Do the courses taken in the first semester appear to affect whether students enter the completion-likely or completion-unlikely clusters? Does their basic skills placement score?
- (4) Are there student support interventions that appear to be making a difference in students being in the completion-likely group?