

A Different Way to Look at Student Groups and Their Success

Executive Summary

This research supports College of Marin's (COM) efforts to assure equitable educational progress among all student groups and to improve student completion overall. Previously, an [analysis of disproportionate impact](#) required by the California Community Colleges Chancellor's Office was conducted for the Student Equity Plan. Both the required student equity metrics and the Student Success Scorecard performance metrics disaggregate the data by student demographic groups. While doing so does highlight progress and success differentials between some demographic groups, these descriptive statistics do not address the reasons for the disparities. Further, they can leave the impression that the identified groups are uniform in their needs and progress. Not only is this not the case, such an approach does not lend itself to identifying at-risk students without over-generalizing.

For these reasons, this study uses cluster analysis to make finer distinctions between student groups taking into account demographics but also college preparation, enrollment and course taking patterns, educational goals and short-term academic progress milestones, some of which are stronger influences on students' success than their gender, race, age or other ascribed characteristics. Each group's risk of failing to complete is quantified as well. Such findings offer a more nuanced means of identifying and responding to students' differential needs.

The data for this study consists of first-time degree and/or transfer-seeking students that the California Community College Chancellor's Office (CO) used in calculating its Student Success Scorecard metrics. It includes 2,435 students who first enrolled at COM from academic year 2004-05 through 2008-09. These students were tracked across 6 years to determine their completion. Therefore, 2008-09 is the most recent cohort. However, the analyses focus on characteristics and behavior during these students' first semester and year in relation to completion within 6 years. With this approach, findings can be applied to support new entering student cohorts.

This research identified eight distinct student clusters, only two of whom were prepared for college. As would be expected, they completed at high rates, 71% and 79%. Though the majority of students enter COM unprepared, their outcomes vary tremendously. These clusters' completion rates ranged from 19% to 74%. Demographic characteristics were among the most important variables only in the two most homogenous groups—the two prepared groups, which consisted of predominately White and not economically disadvantaged students—and the group with the most racial/ethnic diversity. For all groups, college preparation, course-taking patterns, course success, number of units attempted, and persistence were stronger classifying characteristics. Each cluster exhibits different combinations of these primarily behavioral characteristics. For example:

- One group appears to have had a modest need for developmental English and then moved on to complete and transfer at high rates.
- One group consistently enrolled part-time, made good academic progress, but appeared to be persisting without a clearly defined goal but completed at only 45%.

- Another group enrolled nearly full-time. Their course taking and educational goal suggested they were attempting to fulfill degree requirements, but struggled academically and left after their first semester. 28% completed.
- Another group appeared to be testing the water, enrolling in a small number of units, no math or English and undecided educational goal, struggled academically, stopped out, then returned for one more semester to try again. 39% completed.
- 2 clusters with similar academic progress struggles completed at different rates, 19% vs. 45%; the more racially/ethnically diverse and economically disadvantaged cluster had a lower drop in first to second semester GPA and completed at the higher rate.
- Some groups took no English or math, some took mostly developmental or only college-level, and some took both levels concurrently, especially English. The mix of course-taking patterns among the clusters is complex and does not clearly correspond to high or low completion, warranting further investigation.
- In some clusters, more than 25% of students were undecided on their educational goal, raising the issue of advising and educational plans.

The findings of this study show that many students complete college despite being unprepared at the beginning. Without the data, we cannot determine what types of supports any of these students may have received from COM. However, the observed behavioral patterns of these student clusters suggest COM should review and discuss prerequisites, placement testing, advising, educational plans, support programs, English and math course taking, and, ideally, an academic alert system that could identify at-risk students early enough to intervene with support designed to address their particular challenges and improve their chances of success.

A Different Way to Look at Student Groups and Their Success

The College of Marin (COM) is developing plans, implementing new programs and services, and modifying policies and practices to assure equitable educational progress among all student groups and to improve student completion overall. Previously, an [analysis of disproportionate impact](#) required by the California Community Colleges Chancellor's Office was conducted for the Student Equity Plan. Both the required student equity metrics and the Student Success Scorecard performance metrics disaggregate the data by student demographic groups. While doing so does highlight progress and success differentials between some demographic groups, these descriptive statistics do not address the reasons for the disparities. Further, they can leave the impression that the identified groups are uniform in their needs and progress. Not only is this not the case, such an approach does not lend itself to identifying at-risk students without over-generalizing.

For these reasons, this study uses cluster analysis to extend those findings beyond student demographics to take into account college preparation, enrollment and course taking patterns, educational goals and short-term academic progress milestones, some of which are stronger influences on students' success than their gender, race, age or other ascribed characteristics. Cluster analysis permits making finer distinctions between groups of students, including differences in their degree of risk. Such findings offer COM administrators, faculty and staff a more nuanced means of identifying and responding to students' differential needs. Further, these results are intended to prompt discussion and suggest a broader range of strategies for improving the success of all COM students.

Data Used in Analyses

The data used in this study were those of first-time degree and/or transfer-seeking students that the California Community College Chancellor's Office (CO) used in calculating its Student Success Scorecard metrics, e.g., completion, attainment of 30 units, and persistence for the first three consecutive terms. All California community colleges are expected to improve their performance on these metrics. As such, in-depth exploration of these data to unearth enrollment patterns and characteristics of students who complete and fail to complete will aid in that goal. In addition, since these metrics are based on tracking students across six years, this research will help to create short-term metrics that can be used as milestones to gauge our progress along the way.

Students included in this dataset (N=2,435) are first-time students in the CA CC system who began at COM and completed at least 6 units and attempted any math or English in their first three years of enrollment. As noted, these students are tracked over a 6-year period as the basis for the Scorecard metrics. Students are tracked whether they remain at COM or complete at another institution. The most recent completion data available at the time of this study was for the cohort entering COM in the 2008-09 academic year. Cohort years included in this study are 2004-05 through 2008-09.

Based on previous analyses and the student success literature which shows that first term and first year are crucial to long-term success, and the need to use research findings to identify at-risk

students and develop early interventions and support programs, we focused our analyses on the first term and first year of enrollment. We divided the students into 2 separate groups: those who completed two consecutive semesters (N=1,672); and those who stopped out or dropped out of COM after their first semester (N=763). Students who stopped out/dropped out comprise 31% of the dataset.

The set of variables used in this study are commonly shown in the literature to influence student success, including completion/graduation. In addition, the choice of factors that could be examined was constrained by the availability and accuracy of COM data. Factors in the analyses include students' course success and GPA, math and English courses taken, and unit load during the first term and first year of enrollment. Patterns that exist in these data may be useful in identifying at-risk students and developing early interventions to help them progress. Other factors--student demographics, academic preparedness, and enrollment patterns over the first 3 years (6 terms) of enrollment—were also included. (See Appendix A: Methodology, Models and Descriptives for methodology detail and other factors considered, but currently unavailable.)

Statistical Methods

In addition to descriptive statistics, this research was conducted using cluster analysis. Cluster analysis models use a set of input variables to classify students into distinct groups, or clusters, based on similar values for each variable. Students in a particular cluster are more similar to each other than they are to students in any other cluster. For example, all students are classified using persistence, college preparation status, English and math course taking in the first year, and other factors. But, the percentage that persisted, whether they were prepared or unprepared for college, and the level of English and math courses they enrolled in produces different combinations. On some variables, they are similar, on others different. These combinations are the basis for the clusters, or groups, into which students are categorized. (See Appendix A: Methodology, Models and Descriptives for methodological and model detail and descriptive statistics).

Findings

1st Cluster Analysis: Students Attending the First Two Consecutive Semesters

This analysis of students who consecutively enrolled at COM for their first two semesters yielded five distinct student clusters. Descriptive statistics generated for each cluster add to the picture of each group. Completion rates by cluster show each group's level of risk of success or failure. Basic progress and risk data for each cluster are summarized below, followed by a description of characteristics that makes each cluster distinct and a brief discussion. Also, to facilitate comparison between clusters, and highlight significant features and findings, the clusters are not necessarily shown in numerical order in the descriptions. The numeric labels are arbitrary and carry no meaning. They are just the way the model identifies the groups.

| Cluster # | Cluster Size | Risk of Dropping Out | Completion Rate | Transfer Rate | Course Success Rate (Year 1) | Mean # of Units Earned at COM (3 yrs) |
|-----------|--------------|----------------------|-----------------|---------------|------------------------------|---------------------------------------|
| 1 | 277 (14%) | High | 19.4% | 13.2% | 54% | 16.0 |
| 2 | 359 (23%) | Moderate | 45.1% | 23.4% | 77% | 35.4 |
| 3 | 303 (19%) | Low | 71.3% | 63.1% | 83% | 37.8 |
| 4 | 397 (25%) | Low | 74.1% | 59.5% | 77% | 49.7 |
| 5 | 331 (20%) | Moderate | 45.2% | 29.0% | 64% | 35.2 |

Cluster 1—Unprepared for college; no third-term persistence

This group is high risk for dropping out. All of them left after their second consecutive term. Less than 1% was prepared for college. They attempted, on average, 18.2 units in their first year but were the least likely group to succeed in their courses, failing or withdrawing late from nearly half (54%) of those courses. They are one of two clusters whose GPA declined from first to second term. Their GPA drop was the steepest (-.44). Ultimately, only 19% completed.

Most students in this cluster took English and/or math in their first year (See Table 1 below). However, while the majority (70%) enrolled in developmental English, one-third also took college-level English. Further, 41% took developmental math and one-quarter took college-level math. Many did not take these courses sequentially. Instead, they either skipped developmental courses, opting instead to start at college level, or enrolled simultaneously in both levels of the same subject area.

Table 1: College Preparation and English and Math Taking by Cluster

| | | Cluster 1 (n=227) | Cluster 2 (n=359) | Cluster 3 (n=303) | Cluster 4 (n=397) | Cluster 5 (n=331) |
|-------------------------|-------------------------------------|-------------------------|-------------------------|-------------------------|-------------------------|-------------------------|
| Academic preparation | % Prepared for College | 0.4 | 0.8 | 100.0 | 0.3 | 0.0 |
| | % Took dev. English year 1 | 69.6 | 25.1 | 0.0 | 98.9 | 89.1 |
| | % Took dev. math in year 1 | 40.5 | 18.9 | 0.0 | 6.6 | 97.9 |
| | % Took college-level English year 1 | 32.2 | 3.6 | 51.5 | 80.9 | 39.2 |
| | % Took college-level math year 1 | 24.7 | 11.1 | 50.8 | 70.2 | 10.3 |

A follow-up analysis of course taking sequences (Appendix B: English and Math Course Taking Sequences, Tables 1 and 2) showed that 21% of these students took both developmental and college-level English simultaneously in the same term. In addition, 3% took only college-level English. Far fewer students enrolled in both levels of math, 1%. However, 18% enrolled in college math only. These patterns raise questions about prerequisites and/or placement testing requirements during the 2004-2009 academic years when these students began at COM.

Gender was somewhat important in distinguishing this cluster of students from others. It included a larger percentage of male students (58%). Although race/ethnicity, economic disadvantage and educational goal were not strong identifiers of members of this cluster, those data can sometimes aid interpretation, so it is offered here for that purpose. While 52% of students in this cluster are White, 19% are Hispanic and 10% African American, the second largest proportions of these latter two groups next to Cluster 5—which has slightly larger percentages of these two groups. In addition, 10% are Asian. 49% are economically disadvantaged. As their educational goal, 50% selected AA/AS degree and 19% were undecided. This group had the second highest percentage of students who selected career development (13%) as their goal.

Cluster 5—Unprepared for college; high retention, moderate success, racially/ethnically diverse

This cluster is shown out of chronological order because it shares some similarities with Cluster 1 that are worth noting, yet its completion rate is substantially higher. This student group is at moderate risk; 45% completed. All students in this cluster were unprepared for college. The characteristic that most distinguishes this group from others, though, is the fact that almost all enrolled in both developmental English and developmental math in their first year (See Table 1 above). However, most

did not move into college level, especially math, in their first year. Their course success rate was low (60%). Their GPA declined from first to second semester (-0.18).

Even so, all of these students persisted. They attended COM for an average of 4.5 consecutive semesters. Their mean units attempted in the first year (21.1) was the second highest of all clusters. For most of the clusters, race/ethnicity was not an important distinguishing factor. However, for this group, it is, in that this cluster is the most racially/ethnically diverse group: 45% White, 25% Hispanic, 12% African American, and 5% Asian. In addition, this group has the highest proportion of economically disadvantaged students (62%).

Cluster 5 students were the second most likely to select an AA/AS degree as their career goal (51%), but the most likely of those who persisted in their first two terms to be undecided on their goal (26%). Even though this group is persisting, these students are struggling academically. Since more than one-quarter of them are undecided on a career goal, advising, and perhaps interest and aptitude testing, seem to be reasonable interventions with these students.

Cluster 1 and 5 Comparison

In some ways, Cluster 1 students seem similar to those in Cluster 5. Both have high rates of English and Math taking in their first year. GPA declined for both groups, though moreso for Cluster 1, and their course success rate was low. However, Cluster 1 students dropped out. Cluster 5 students persisted. Ultimately, they completed at strikingly different rates, 19% vs. 45%. What made their outcomes different?

Cluster analysis is not designed to predict completion, but some patterns observed between these groups do raise questions that can inform discussion and subsequent research. First, the higher percentage of Cluster 5 students enrolled in English and math during their first semester might be an explanation. However, another unprepared group of students (Cluster 2, to be further discussed below) completed at an equivalent rate (45%) yet far fewer enrolled in English and math during their first year.

Second, the way these two unprepared groups enrolled in their English and math courses is worth considering. Many of the drop-out cluster students took developmental and college level English and math courses simultaneously. Some skipped developmental altogether and enrolled at college level. However, in the follow-up analysis of course taking sequences (Appendix B English and Math Course Taking Sequences Tables 1 and 2), these patterns existed for both Cluster 1 and Cluster 5. While 21% of Cluster 1 students took both developmental and college-level English in the same term, 18% of the higher completing Cluster 5 group did as well. In addition, the Cluster 5 students were somewhat more likely to take college English only (8%) than was the Cluster 1 group (3%).

This same review of math course taking showed 1% of students in both clusters taking developmental and college math concurrently. Among the lower completing Cluster 1 group, 18% took college math only. No Cluster 5 students took college math only. This differential in college math taking is notable for further analysis, but insufficient in itself to draw any conclusions about its effect beyond supporting the general question of prerequisites and placement testing.

For English course-taking, given that this sequencing issue existed in both groups, and at similar levels, it is not likely an explanation for their differential completion rates. It could have lowered both groups' rates though, since passing one's courses is essential to continue in college.

Another notable observation is that racial/ethnic background is a distinguishing factor of Cluster 5 students, but far less so for Cluster 1. Cluster 5, with its substantially higher completion rate, is more diverse in this respect than Cluster 1. In fact, it includes a slightly higher portion of African American and Hispanic students and a lower percentage of White and Asian students (See Appendix A Methodologies, Models and Descriptives Table 1). Much of the student success literature and [COM's own equity metrics](#) lead us to expect this more racially diverse group's completion rate to be lower. But that is not the case

for Cluster 5. Further, the highest proportion of economically disadvantaged students among the 5 clusters is in Cluster 5. This characteristic too, is often associated with failure to complete college. However, COM's equity metrics show that economically disadvantaged students are equivalently or more likely to complete.

These findings raise questions about factors influencing retention that were less effective or absent for Cluster 1 than for Cluster 5. Are Cluster 5 students involved in COM student support programs? Is there a difference in financial aid between the two groups? Might there be a learning community effect for Cluster 5 given the high percentage enrolled in both developmental English and math? Data to test these questions were not available to include in this analysis. Given the high rate of persistence among Cluster 5 students, it may be that this group would be particularly responsive to changes in institutional practices that might improve their likelihood of completing. For instance, Cluster 5 students were about as likely as Cluster 1 students to choose an AA/AS degree as their first educational goal, but more likely than the other four clusters to select undecided (26%). Might the initial selection of an "undecided" goal trigger an appointment with an advisor or interests or aptitude testing that could help these students decide on and plan an appropriate path?

Cluster 2—Part-time, high retention, slow to fulfill degree requirements

This group is at moderate risk; 45% completed. Not enrolling in English and math in the first year, attempting few units, high persistence and lack of college preparation characterize this group. Only one-quarter or less took math or English. They were the least likely group to do so. On average, these students attempted 15.2 units in their first year, less than any of the other clusters. However, they attended COM consistently longer, 4.7 consecutive terms on average. In addition, they passed 77% of their courses and earned a 3.01 average GPA—the second highest among the five clusters.

Though gender, race/ethnicity, economic disadvantage and educational goal were the least important factors distinguishing this cluster from others, as previously stated, the characteristics may be useful for interpretation. This group has the highest proportion of females (55%) among the 5 clusters, the second highest proportion of Hispanic students (23%) and the second highest proportion of economically disadvantaged (57%) students.

While 45% of these students completed, it is unclear whether all students in this group intended to complete. They were one of the least likely clusters to indicate an AA/AS degree (42%) as their educational goal and few stated transfer or certificate as their goal. They were the most likely cluster to select career development (17%). They were the second most likely, among the groups who enrolled in the first two consecutive terms, to indicate undecided (21%).

Persistence and a clear goal usually go together. Given these students' strong persistence, yet moderate completion rate, and what from an institutional standpoint looks like uncertainty for many about what their goal is when they begin college, is there anything COM can do that would help more of them complete or progress more quickly?

Advising, and an educational plan if they do not have such a plan in place, may help. But, on average, these students are more than 3 years older (25.3) than students in the other groups. 57% are economically disadvantaged. These characteristics, along with enrolling in few units per semester, suggest they were likely employed. If so, this may preclude their ability to take a higher unit load. On the other hand, it may be that they are unaware of financial aid possibilities that could allow them to take additional units. Establishing an advisor alert that triggers when students have enrolled in a small number of units in their second consecutive semester without having enrolled in math or English, especially if their educational goal is undecided, may be a way to understand these students' circumstances and assure they are aware of their options.

Cluster 3-High achievement, college-prepared, transfer-oriented

These students are low risk. They are the only college-prepared group. None needed developmental math or English. They were the most likely cluster to transfer (63%) and along with Cluster 4 were the most likely to complete (71%). This group earned the highest first-year course success rate (83%) and GPA (3.17). Race and economic status were more important distinguishers of this cluster than other clusters. White students comprised 72% of this group (the highest proportion in the 5 clusters) and 29% were economically disadvantaged (the smallest proportion of all the clusters).

Most of these students persisted into their third term (83%). However, they attended fewer consecutive and total terms (3.8 and 4.1, respectively) than all clusters except Cluster 1, most of whom dropped out of college after two terms. Cluster 3's enrollment pattern would be consistent with their high transfer rate and, unlike other clusters, no need to complete developmental work before transfer, which allows faster completion.

This group's English and math taking behavior contributes to the questions this study's findings raise about the role of English and math in the first year. About half of Cluster 3 students did not enroll in these courses during their first year. Even so, they completed and transferred at high rates. Cluster 5 and Cluster 2 completed college at the same moderate rate despite opposite English and math taking behavior. These different patterns raise questions about who should take these courses and when. In fact, Cluster 4, described below, displays yet another pattern of English and math taking.

Cluster 4-High achievement, highest unit load, some developmental coursework needed

This group is low risk. This is the one group of students who, on average, was enrolled full-time during the first year. Most were not college-prepared. Almost all (99%) took developmental English. But, they appear only to have needed one developmental English course to be at college-level. Most (81%) took college English as well. More than half (54%) enrolled in the two courses concurrently. Apparently developmental math was not needed. Few enrolled in such a course (7%). Most took college-level math (70%). Cluster 4 students succeeded in nearly 80% of their courses. They attended COM an average of 4.6 consecutive terms and earned more units than any other cluster during the first 3 years (49.7). Along with Cluster 3, they had the highest completion (74%) and transfer (60%) rates.

Why would these students need developmental English and be able to successfully take that course simultaneously with college English? This pattern, combined with their high transfer and completion rates, suggests these students needed little assistance to be college ready. Since, on average, they were the youngest of the consecutively enrolled group (19.2 years old), perhaps they were still close enough to having taken English in high school to recall much of what they learned, once reminded. Additional information that may assist in understanding this cluster include: They were the most likely to select earning an AA/AS (58%) as their educational goal. The majority (55%) was male. More than half of these students were White (56%), 14% Asian, 12% Hispanic and 2.5% African American.

Although students in this cluster complete at the highest rate of all the clusters, might this be a group whose time to completion could be shortened (one of COM's strategic planning goals) by intervening with students who simultaneously test into the highest level of developmental English and college-level math? Perhaps through test preparation and re-test, completing a module rather than a full semester course, or through some form of accelerated English, possibly including ESL if the proportion of students that are Asian (14%) and Hispanic (12%) are from families in which a language other than English is spoken at home.

The course taking patterns of this group also contribute to the question about the need for and effect of English and math enrollment in the first year. While slightly more than half of the college-

prepared Cluster 3 enrolled in these courses in their first year, an additional 20-30% more of this group did, yet they completed and transferred at approximately equivalent rates.

2nd Cluster Analysis – Students stopping out or dropping out after the first semester

The 5 clusters of students described above included all students in the Student Success Scorecard dataset who enrolled at COM in their first two consecutive semesters. The following clusters include only students who did not enroll in a second consecutive term. These students comprised 31% of the Scorecard dataset used for this study. Though enrolled at COM for an average of only 2 terms, ultimately 50% completed in the 6-year tracking period, some at COM but most either at another community college or a 4-year college or university. This completion rate compares to 54% completion for the students who enrolled in the first two consecutive semesters.

The cluster analysis for this set of students yielded 3 unique student groups. Basic data for each cluster is summarized below, followed by a description of characteristics that makes each cluster distinct and a brief discussion. (See Appendix A: Methodology, Models and Descriptives for model detail, description and descriptive statistics for each cluster.)

| Cluster # | Cluster Size | Risk of Dropping Out | Completion Rate | Transfer Rate | Course Success Rate (Term 1) | Mean # of Units Attempted in Term 1 |
|-----------|--------------|----------------------|-----------------|---------------|------------------------------|-------------------------------------|
| 6 | 197 (29%) | High | 27.9% | 23.4% | 49% | 10.6 |
| 7 | 268 (39%) | High | 39.1% | 30.1% | 55% | 5.5 |
| 8 | 214 (32%) | Low | 79.0% | 75.7% | 76% | 7.1 |

Cluster 6-Low completion, attempting degree requirements

Students in this group are high risk for dropping out, even though half indicated their goal was an AA/AS degree. Only 28% completed. They enrolled for 10.6 units, on average, but succeeded in only 49% of their courses. These students were unprepared for college. Most enrolled in developmental English (97%) in their first semester; 39% simultaneously took college-level English. For math, 29% enrolled in developmental. However, almost all were unprepared yet 20% enrolled in college-level math.

Many students in this cluster appear to have been attempting to fulfill completion/transfer requirements by taking English and math in their first term and enrolling nearly full-time, with half having declared an AA/AS degree as their educational goal. However, they struggled academically. Though none enrolled in their second consecutive semester, 42% returned to COM for at least one additional semester. On average, this group enrolled in two non-consecutive terms.

Clearly, this group was trying to succeed. As with the high risk Cluster 1 students who consecutively enrolled in two terms yet completed at only 19%, the issue of prerequisites and placement testing requirements at the time, as well as advising, arises. Approximately 28% of all students who left after their first term were undecided about their educational goal. Among Cluster 6, 26% were undecided, again, suggesting the need for advising.

This was the most racially/ethnically diverse group of the 3 clusters who stopped out after their first semester (50% were White, 21% Hispanic, 13% African American and 6% Asian) and the most economically disadvantaged (60%). In these respects, they are most similar to Clusters 1 and 5 above, and their completion rate falls between the rates of those two clusters.

Cluster 7-Low completion, few units, unprepared, no English or Math

Students in this group are high risk for dropping out; 39% completed. Few were prepared for college. In contrast to Cluster 6, almost none of these students took English in their first term. Further, they enrolled in about half as many units (5.5). In addition, a smaller percentage of these students enrolled in math. Only 14% enrolled in developmental math, 2% in college-level. Like Cluster 6, they struggled academically, succeeding in only half of their courses (55%), despite the lower unit load.

This group took fewer than 2 courses, on average, in their first semester and had the highest proportion of students with an undecided educational goal (30%). A far lower percentage of these students than Cluster 6 students selected an AA/AS degree as their goal (27% vs. 50%). Almost half were economically disadvantaged (49%). These characteristics, along with their lack of preparation for college and lack of English and math taking, suggest the possibility that this group of students may have been testing the water, trying to determine whether and which college was right for them, or alternatively, the issue could have been whether they were able to get the advising and/or classes they needed at COM. Over half (58%) returned to COM for at least one semester after stopping out, enrolling for an average of 2.3 terms in three years. This suggests some degree of motivation to succeed, but also perhaps hesitancy that advising and intrusive support programs may be able to counter.

Cluster 8-High success, college-prepared, transfer-oriented

Students in this cluster are low risk. They are distinguished primarily by the fact that most are prepared for college, took no developmental level courses, and were far less likely than the other clusters to be economically disadvantaged (15% vs. 60% and 49%). In addition, this group was the least likely to indicate AA/AS degree as their educational goal (15% vs. 50% and 27%). They were more likely to select basic skills (28%) or undecided (26%). Another 14% chose educational development. All their math and English courses were college-level, though a relatively small percentage enrolled in those courses (22% English, 31% math). They succeeded in the majority of their courses and earned a first-term GPA of 2.93. This group is more homogenous than the other two clusters of students who left after their first semester; 78% are White.

On average, this group enrolled at COM for 7 units in their first term and attended 1.7 semesters. Only 38% of this group returned to COM after their second semester. However, 79% completed, almost all through transfer. While their reasons for leaving COM cannot be determined by the data in this study, the variety of educational goals they selected suggests they may have enrolled only to take a particular course of interest or needed for transfer, or perhaps to test the fit while considering their college options, or, given that this was the youngest first-term-only cluster (mean age 18.4), they may have been concurrently enrolled high school students or recently graduated from high school taking summer courses before enrolling at another college or university. The anecdotal belief that many students attend COM only to fulfill math or English requirements, may have been the case for some of these students, but the majority did not take those courses.

Conclusion

This study's findings offer a nuanced way to identify students who are at-risk of failing to complete their education. Demographic characteristics that are often used in identifying and intervening with students are not as salient for most of the cluster groups in this study as is their preparation for college, persistence, English and math course-taking patterns, academic progress, number of units enrolled and educational goal. Among the eight clusters identified, two were college-prepared. They

completed at 71% and 79%. Completion rates of the six clusters that were unprepared for college varied widely, from 19% to 74%.

COM students, especially those who were unprepared, were using various strategies in college with different levels of success. The primarily behavioral characteristics used in this analysis, while not “predicting” completion, suggest needs and issues that may influence these groups’ outcomes. For example, one group appears hesitant, enrolling for few units, struggling academically then stopping out after their first term to wait awhile and try again later. Another group begins with nearly a full course load, taking required courses intending to earn a degree, but struggles academically and drops out after the first semester. Some re-engage later, but most never complete. A third group consistently enrolls in multiple semesters, but only part-time. They succeed in their coursework, but seemingly without a clear end goal for their education in sight. Yet another group seems to only need a little developmental help with English, then they achieve at high rates. Clearly, these students experience college very differently. The ability to distinguish the multiple pathways our students take through college, and the different needs accompanying those pathways, can help us identify such patterns and intervene to improve students’ chances of success.

Findings from this research also raised questions about institutional practices such as pre-requisites and placement testing, advising, and English and math requirements. The issue of prerequisites has likely been resolved in the interval since the most recent cohort in this study entered COM. A follow-up analysis will show whether prerequisite requirements are in place and enforced. If they are, not only will it be evidence of institutional improvement, it may be feasible to conduct a study of their effect on student success. Another issue worth discussing and investigating is the variety of English and math taking patterns evident among these groups: developmental only, college-level only, concurrently or sequentially enrolled, English only, math only, or both--sometimes at different levels, yet with no clear relationship to high, moderate or low risk of completing college. Statistical modeling may help clarify.

Finally, these findings can be used to identify at-risk students through technological means, such as alert systems. The foundation of such a tool is a robust, reliable data system from which the necessary research can be conducted and identifying and notification triggers built. COM is beginning forays into an alert system via COMCare and the Student Success Collaborative.

Appendix A: Methodology, Models and Descriptives

These analyses were conducted using data for the cohort of 2,435 students who entered COM as first-time, degree and/or transfer-seeking college students during academic years 2004-05 through 2008-09. These are the cohorts used by the California Community College Chancellor's Office (CO) to calculate the student success outcomes for the Student Success Scorecard. The CO defines degree and transfer/seeking based on a minimum of 6 units earned and any English or math attempted in the first three years. Completion is tracked over a 6-year period for all students in this dataset, including whether they transferred to another institution. The 2008-09 cohort was the most recent group for which 6-year completion data are available.

Student-level completion and demographic data for each cohort were downloaded from the CO's Data on Demand repository. For each student, this data file includes the 6-year outcomes for completion, transfer to a 4-year institution, AA/AS degree earned, college-prepared status, and basic demographic data.

In order to determine term-by-term GPA, enrollment, courses taken, and other factors, the course-level data for each student was extracted from the CO's MIS system data. For each student, various metrics were then calculated cumulatively and for each of the first 6 terms from their first term of enrollment. These metrics include: GPA, course success, developmental and college-level English and math courses taken, units attempted, and units earned. Stop-outs in enrollment and number of terms enrolled were also calculated. These data were then merged with the completion data to create the final dataset.

Descriptive analyses of completion rates were conducted for groups of students who differed in background characteristics, academic preparedness, enrollment patterns, and academic progress at COM. Based on a review of the literature and previous studies on completion/graduation, as well as the availability of accurate data to use for this study, the following factors were included:

- 1) Student demographics:
 - a. Age at first term
 - b. Race/ethnicity
 - c. Economically disadvantaged (CO definition is students eligible for CalWorks or WIA, those receiving BOGW or Pell grant, those with VTEA economically disadvantaged status, and those whose SSN matches with Dept. of Social Services)
 - d. Gender
- 2) Academic preparation:
 - a. Level of English and math courses attempted for each term (English 150 and higher and math 103 and higher were considered college-level)
 - b. Prepared status (CO definition of prepared is lowest attempted English and math courses attempted are college-level; unprepared is if either math or English is below college-level)
- 3) Academic progress—cumulative and for each term of enrollment:
 - a. GPA
 - b. Course success (calculated as the percentage of courses attempted resulting in a grade of A, B, C, CR, or P)
- 4) Enrollment patterns:
 - a. Number of units attempted by term
 - b. Number of units earned in first 3 years from first term

- c. Persistence (enrolling in the first 3 consecutive terms)
 - d. Stop-outs (skipping one or more terms of enrollment followed by re-enrollment)
 - e. Total number of terms enrolled in the first 3 years from first term
- 5) Educational goal
- a. Stated educational goal at the first term of enrollment (AA/AS degree, basic skills, career development, vocational certificate, educational development/enrichment, transfer to a 4-year institution, and undecided)

NOTE: Various other data such as financial aid information, Math and English placement scores, student support programs, counseling, educational plans, and some student demographic data such as veteran's status and first-generation college student, may have an impact on student success and completion. The use of such data was explored for this study, but these data were not available for the students and cohort years in this study either because they were not being collected at the time or because the data collected were incomplete or unreliable. Increasing institutional efforts to collect and maintain these data systematically will improve COM's capacity for research that supports planning and student success.

Analysis Groups

Initial descriptive findings showed that 31% of students in our dataset did not enroll in the second consecutive semester, though approximately 50% of them ultimately completed. Students who did enroll in a second consecutive term completed at 54%. Despite the relatively small difference in completion rates, the student success literature shows that first term and first year are crucial to long-term success, and it is possible that important differences exist between students who enroll consecutively and those who stop-out and return. For these reasons, the analyses focused on the first term and first year of enrollment and the dataset was divided into 2 separate student groups: Group 1, those who completed two consecutive semesters (N=1,672); and Group 2, those who stopped out or dropped out of COM after their first semester (N=763). For Group 1, analyses focused on academic progress during the first year of enrollment. Group 2, who did not enroll in the second consecutive semester, was analyzed separately using their first term data, rather than first year.

Cluster Analysis

Cluster analysis was used to identify students based on their shared background characteristics, enrollment patterns, and academic progress. These findings help to quantify different groups' risk of dropping out, determine specific risk factors inherent to each cluster group, and can be used in developing appropriate interventions.

A 2-step clustering procedure in SPSS was selected because it allows inclusion of categorical variables as predictors. The procedure includes an initial step in which each case is scanned and assigned to a pre-cluster, then a traditional hierarchical clustering algorithm is used to cluster the pre-clusters.¹ Cluster analysis is exploratory and does not distinguish between independent and dependent variables. It groups cases (in this study, students) on the basis of their homogeneity according to input variables are specified. Post-hoc significance testing was used to determine significant differences between cluster groups on evaluation or criterion variables (in this study, completion).

¹ http://www.spss.ch/upload/1122644952_The%20SPSS%20TwoStep%20Cluster%20Component.pdf

Parameters selected for analysis

Several “test” models were run using various input factors and analysis parameters available in SPSS. Some input variables were assigned too much importance by the algorithm, overwhelming or “swamping” the analysis and creating clusters based solely on that factor. Consequently, those were omitted. Including some others reduced the quality of the clusters to an unacceptable level because they did not sufficiently differentiate. We also tested the impact of using a fixed number of clusters and different goodness-of-fit measures. After careful evaluation of each test model, the following parameters were selected:

- SPSS’s default algorithm was used to automatically determine the optimal number of clusters based on goodness-of-fit. For both the first term only and two-term student group models, this process yielded few enough clusters so that they were understandable and interpretable, but sufficient to discern subtle yet important differences between groups.
- The clustering criterion (goodness-of-fit) method selected was Akaike’s Information Criterion (AIC), a somewhat less conservative method typically resulting in a greater number of clusters.
- The log-likelihood distance measure was selected because the dataset included categorical input variables.
- Continuous variables (units attempted, course success, total terms enrolled) were not assumed standardized.
- The parameters for pre-cluster differentiation were left at the default values.

Model 1- Students attending the first 2 consecutive semesters

A total of 13 input factors were entered into the final model.

Categorical inputs:

- Took developmental English in first year (yes/no)
- Took developmental math in first year (yes/no)
- Took college-level English in first year (yes/no)
- Took college-level math in first year (yes/no)
- Economically disadvantaged (yes/no)
- Race (all categories)
- Gender (male/female)
- First educational goal (AA/AS degree, basic skills, career development, vocational certificate, educational development/enrichment, transfer to 4-year institution, and undecided)
- Prepared for college (yes/no)
- Persisted into the third term (yes/no)

Continuous inputs:

- First year course success (percentage)
- Number of consecutive terms enrolled at COM from first term in first 3 years (1-6)
- Total number of units enrolled in first year

Model 1 yielded 5 distinct clusters, ranging in size from 277 (14% of the sample) to 397 (25%). The silhouette measure of cohesion and separation (overall goodness-of-fit) measure for the final model was .29, indicating a “fair” cluster quality. This indicates an acceptable level of fit. However, having additional data to input into future models may improve the quality.

Model 2- Students stopping out or dropping out after the first semester

A total of 12 input factors were entered into the final model.

Categorical inputs:

- Took developmental English in first semester (yes/no)
- Took developmental math in first semester (yes/no)
- Took college-level English in first semester (yes/no)
- Took college-level math in first semester (yes/no)
- Economically disadvantaged (yes/no)
- Race (all categories)
- Gender (male/female)
- First educational goal (AA/AS degree, basic skills, career development, vocational certificate, educational development/enrichment, transfer to 4-year institution, and undecided)
- Prepared for college (yes/no—lowest level math or English course was college-level)

Continuous inputs:

- First semester course success (percentage)
- Total number of terms enrolled at COM from first term in first 3 years (1-5, Maximum value was 5 because these students did not attend the second semester)
- Total number of units enrolled in first semester

Model 2 yielded 3 distinct clusters, ranging in size from 197 (29% of the sample) to 268 (40%). The silhouette measure of cohesion and separation (overall goodness-of-fit) measure for the final model was .21, indicating a “fair” cluster quality. Again, this quality level is acceptable but could be improved with additional data to input into the model.

Cluster validation and interpretation

The 2-step procedure provides each input variable’s relative importance for the construction of a specific cluster, based on chi-square distributions for categorical variables and t-tests for continuous variables. See **Figure 1** (for Model 1) and **Figure 2** (for Model 2) to view input variables in order of importance within each cluster. The most important variable is shown first in the column; the last is least important.

For both models 1 and 2, review of the descriptive statistics for each cluster group suggested that the clusters yielded had face validity—that is, the clusters defined by the models appeared reasonable and made sense. Completion rates among the clusters varied significantly, indicating that the cluster input variables selected for the model distinguished completion satisfactorily.

The clusters were then interpreted and validated by running post-hoc significance tests on each of the underlying variables with which we began our descriptive analyses (whether or not they were included as input variables in the cluster model). Academic progress, enrollment patterns, and demographics all yielded significant differences between cluster groups. See Appendix A Methodology, Models and Descriptives Table 1 (for Model 1) and Table 2 (for Model 2) for complete profile statistics by cluster.

How to Read the Cluster Models

In Figures 1 and 2 below, the number at the top of each column simply identifies the cluster by assigning it a number. Size indicates the number and percentage of all students in the dataset that comprise the cluster. Inputs are the variables the model uses to classify students into distinct clusters based on similar values for each variable. Further, the model determines the importance of each variable within each cluster. The most important variable that distinguishes the clusters is shown first, at the top of the column; the last is least important. For example, in Figure 1, for Cluster 1 students, persistence is the key distinguishing feature. First year course success is the second most important in defining this group of students.

The legend at the top right of the model diagram indicates the importance of the input variables in the overall model, not to a particular cluster. The shading represents the importance of each input variable in defining the overall model.

Students Attending the First Two Consecutive Semesters

Figure 1. SPSS Cluster Analysis Model 1-Students Attending First Two Consecutive Semesters

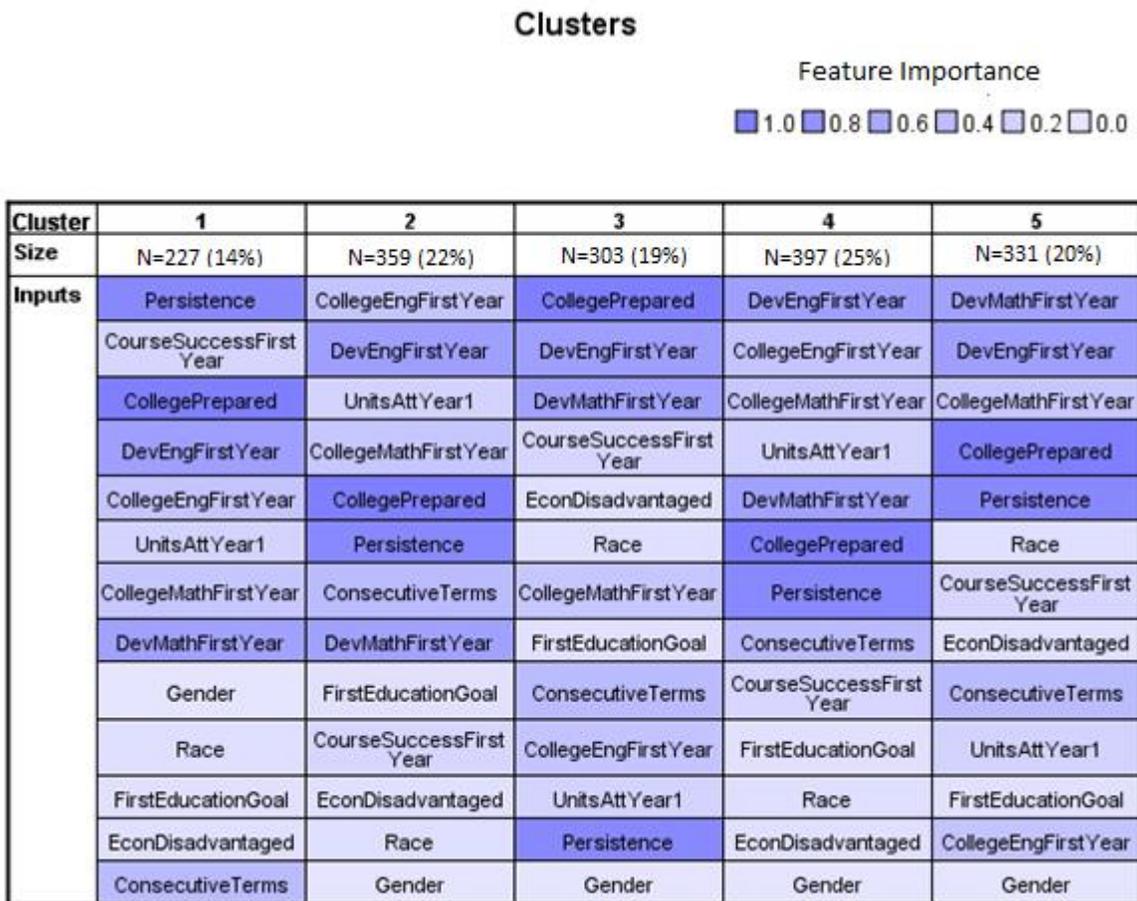


Table 1. Cluster Profile Descriptive Statistics-Students Attending First Two Consecutive Semesters

| | | Cluster 1 (n=227) | Cluster 2 (n=359) | Cluster 3 (n=303) | Cluster 4 (n=397) | Cluster 5 (n=331) | All students |
|-----------------------------------|--|-------------------------|-------------------------|-------------------------|-------------------------|-------------------------|-----------------|
| Outcomes | % Completed* | 19.4 | 45.1 | 71.3 | 74.1 | 45.2 | 53.5 |
| | % Transferred to 4-year institution | 13.2 | 23.4 | 63.1 | 59.5 | 29.0 | 39.4 |
| Academic success | Mean course success in first year | 0.54 | 0.77 | 0.83 | 0.77 | 0.64 | 0.72 |
| | Mean GPA in first year | 2.41 | 3.01 | 3.18 | 2.93 | 2.68 | 2.87 |
| | Mean change in GPA from 1st to 2nd term | -0.44 | 0.06 | -0.02 | 0.02 | -0.18 | -0.07 |
| Academic preparation | % Prepared** | 0.4 | 0.8 | 100.0 | 0.3 | 0.0 | 19.1 |
| | % Took dev. English year 1 | 69.6 | 25.1 | 0.0 | 98.9 | 89.1 | 57.9 |
| | % Took dev. math in year 1 | 40.5 | 18.9 | 0.0 | 6.6 | 97.9 | 31.5 |
| | % Took college-level English year 1 | 32.2 | 3.6 | 51.5 | 80.9 | 39.2 | 42.9 |
| | % Took college-level math year 1 | 24.7 | 11.1 | 50.8 | 70.3 | 10.3 | 34.8 |
| Retention | Mean # of units attempted year 1 | 18.2 | 15.2 | 19.0 | 23.9 | 21.1 | 19.7 |
| | Mean # of units earned in 3 years | 16.0 | 35.4 | 37.8 | 49.7 | 35.2 | 36.6 |
| | % Persistence | 0.0 | 99.7 | 82.8 | 100.0 | 100.0 | 82.7 |
| | Mean total terms attended in 3 years from first F/S term | 2.8 | 4.9 | 4.1 | 4.8 | 4.6 | 4.4 |
| | Mean consecutive terms attended from first F/S term | 2.0 | 4.7 | 3.8 | 4.6 | 4.5 | 4.1 |
| Demographics | % Economically disadvantaged*** | 49.3 | 56.6 | 29.4 | 41.3 | 61.6 | 47.7 |
| | % White | 51.5 | 48.8 | 71.9 | 56.4 | 44.7 | 54.6 |
| | % Hispanic | 18.9 | 23.1 | 5.3 | 12.3 | 24.8 | 16.9 |
| | % African-American | 10.1 | 4.5 | 1.0 | 2.5 | 11.8 | 5.6 |
| | % Asian | 9.7 | 11.4 | 9.6 | 14.4 | 5.4 | 10.3 |
| | % Male | 58.2 | 45.1 | 51.8 | 55.4 | 48.9 | 51.5 |
| | % Mean age | 20.6 | 25.3 | 21.2 | 19.2 | 21.4 | 21.6 |
| First Educational Goal | % AA/AS | 50.2 | 41.5 | 42.6 | 57.7 | 51.1 | 48.9 |
| | % Basic skills | 7.1 | 7.5 | 10.2 | 1.3 | 2.4 | 5.4 |
| | % Career development | 13.2 | 16.7 | 7.9 | 7.1 | 7.9 | 10.4 |
| | % Certificate | 0.4 | 1.4 | 0.3 | 0.3 | 0.6 | 0.6 |
| | % Educational development | 0.9 | 2.5 | 7.3 | 1.5 | 1.2 | 2.7 |
| | % Transfer | 6.6 | 3.6 | 11.2 | 8.8 | 7.0 | 7.4 |
| | % Undecided | 19.4 | 20.6 | 16.5 | 19.9 | 26.3 | 20.7 |

*CO definition: Percentage of first-time students with minimum of 6 units earned who attempted any math or English in the first 3 years and achieved any one of the following: earned AA/AS or CO-approved credit certificate; transferred to a 4-year institution; achieved "transfer-prepared" (completed 60 UC/CSU transferable units with GPA>=2.0)

**CO definition: Lowest attempted English or math course is college-level

***CO Definition includes: Students eligible for CalWorks, WIA; those receiving BOGW or Pell grant; those with VTEA economically disadvantaged status; those whose SSN matches with Dept. of Social Services

Students Stopping Out or Dropping Out After the First Semester

Figure 2. SPSS Cluster Analysis Model 2-Students Stopping Out or Dropping Out After First Semester

Clusters

Feature Importance



| Cluster | 6 | 7 | 8 |
|---------|--------------------------|--------------------------|--------------------------|
| Size | N=197 (29%) | N=268 (39%) | N=214 (32%) |
| Inputs | DevEngFirstFSTerm | DevEngFirstFSTerm | CollegePrepared |
| | UnitsAttFirstFSTerm | UnitsAttFirstFSTerm | DevEngFirstFSTerm |
| | CollegePrepared | CollegePrepared | EconDisadvantaged |
| | CollegeEngFirstTerm | CollegeEngFirstTerm | FirstEducationGoal |
| | FirstEducationGoal | CollegeMathFirstFSTerm | DevMathFirstFSTerm |
| | DevMathFirstFSTerm | TotTermsEnrolled3Yrs | CourseSuccessFirstFSTerm |
| | EconDisadvantaged | EconDisadvantaged | CollegeMathFirstFSTerm |
| | CourseSuccessFirstFSTerm | CourseSuccessFirstFSTerm | Race |
| | Race | FirstEducationGoal | TotTermsEnrolled3Yrs |
| | CollegeMathFirstFSTerm | Gender | CollegeEngFirstTerm |
| | Gender | Race | UnitsAttFirstFSTerm |
| | TotTermsEnrolled3Yrs | DevMathFirstFSTerm | Gender |

Table 2. Cluster Profile Descriptive Statistics-Students Stopping Out or Dropping Out After First Semester

| | | Cluster 6 (n=197) | Cluster 7 (n=268) | Cluster 8 (n=214) | All students |
|-------------------------------|---|-------------------------|-------------------------|-------------------------|-----------------|
| Outcomes | % Completed* | 27.9 | 39.1 | 79.0 | 48.5 |
| | % Transferred to 4-year institution | 23.4 | 30.1 | 75.7 | 42.7 |
| Academic success | Mean course success in first term | 0.49 | 0.55 | 0.76 | 0.60 |
| | Mean GPA in term1 | 2.46 | 2.62 | 2.93 | 2.67 |
| Academic preparation | % Prepared** | 0.5 | 7.5 | 95.8 | 33.3 |
| | % Took dev. English term 1 | 97.0 | 0.4 | 0.0 | 57.9 |
| | % Took dev. math in term 1 | 29.4 | 13.8 | 0.0 | 14.0 |
| | % Took college-level English term 1 | 39.1 | 0.0 | 21.5 | 18.1 |
| | % Took college-level math term 1 | 19.8 | 1.9 | 30.8 | 16.2 |
| Retention | Mean # of units attempted term 1 | 10.6 | 5.5 | 7.1 | 7.5 |
| | Mean total terms attended from first F/S term | 2.0 | 2.3 | 1.7 | 2.0 |
| Demographics | % Economically disadvantaged*** | 60.4 | 48.5 | 14.5 | 41.2 |
| | % White | 50.3 | 58.2 | 77.6 | 62.0 |
| | % Hispanic | 20.8 | 14.9 | 4.2 | 13.3 |
| | % African-American | 13.2 | 7.1 | 0.5 | 6.8 |
| | % Asian | 6.1 | 6.0 | 6.5 | 6.2 |
| | % Male | 55.3 | 50.4 | 52.8 | 52.6 |
| | % Mean age | 20.4 | 19.4 | 18.4 | 19.4 |
| First Educational Goal | % AA/AS | 49.8 | 26.5 | 14.5 | 29.5 |
| | % Basic skills | 3.1 | 14.9 | 28.0 | 15.6 |
| | % Career development | 12.2 | 12.3 | 5.6 | 10.2 |
| | % Certificate | 0.5 | 0.8 | 0.0 | 0.4 |
| | % Educational development | 1.5 | 5.6 | 13.6 | 7.1 |
| | % Transfer | 4.6 | 3.4 | 8.4 | 5.3 |
| | % Undecided | 25.9 | 30.2 | 25.7 | 27.5 |

*CO definition: Percentage of first-time students with minimum of 6 units earned who attempted any math or English in the first 3 years and achieved any one of the following: earned AA/AS or CO-approved credit certificate; transferred to a 4-year institution; achieved "transfer-prepared" (completed 60 UC/CSU transferable units with GPA>=2.0)

**CO definition: Lowest attempted English or math course is college-level

***CO Definition includes: Students eligible for CalWorks, WIA; those receiving BOGW or Pell grant; those with VTEA economically disadvantaged status; those whose SSN matches with Dept. of Social Services

Appendix B: English and Math Course Taking Sequence

Table 1: First Year English Course Taking Patterns—Students Attending First Two Consecutive Semesters

| | Cluster 1 | | Cluster 2 | | Cluster 3 | | Cluster 4 | | Cluster 5 | |
|--|-----------|------|-----------|------|-----------|------|-----------|------|-----------|------|
| | N | % | N | % | N | % | N | % | N | % |
| No English | 62 | 27% | 257 | 72% | 147 | 49% | 0 | 0% | 11 | 3% |
| Dev. Eng. then college Eng. | 18 | 8% | 1 | 0.3% | 0 | 0% | 104 | 26% | 46 | 14% |
| Dev. Eng. Only | 92 | 41% | 89 | 25% | 0 | 0% | 76 | 19% | 190 | 57% |
| College Eng. only | 7 | 3% | 12 | 3% | 156 | 51% | 4 | 1% | 25 | 8% |
| Dev. & college concurrently | 48 | 21% | 0 | 0% | 0 | 0% | 213 | 54% | 59 | 18% |
| Total | 227 | 100% | 359 | 100% | 303 | 100% | 397 | 100% | 331 | 100% |

Table 2: First Year Math Course Taking Patterns—Students Attending First Two Consecutive Semesters

| | Cluster 1 | | Cluster 2 | | Cluster 3 | | Cluster 4 | | Cluster 5 | |
|--|-----------|------|-----------|------|-----------|------|-----------|------|-----------|------|
| | N | % | N | % | N | % | N | % | N | % |
| No Math | 94 | 41% | 261 | 73% | 149 | 49% | 115 | 29% | 7 | 2% |
| Dev. Eng. then college Math | 13 | 6% | 9 | 3% | 0 | 0% | 21 | 5% | 30 | 9% |
| Dev. Math only | 77 | 34% | 58 | 16% | 0 | 0% | 3 | 1% | 290 | 88% |
| College Math only | 41 | 18% | 30 | 8% | 154 | 51% | 256 | 64% | 0 | 0% |
| Dev. & college concurrently | 2 | 1% | 1 | 0.3% | 0 | 0% | 2 | 1% | 4 | 1% |
| Total | 227 | 100% | 359 | 100% | 303 | 100% | 397 | 100% | 331 | 100% |