Student Data Warehouse POC Analysis Plan  
Su Jin Jez  
8/29/14

**Project Overview**

As closely as possible, I have replicated major components of the Institute for Higher Education Leadership & Policy’s *Student Flow Analysis: CSU Student Progress Toward Graduation* (<http://www.csus.edu/ihelp/PDFs/R_CSU_MOA_excerpt.pdf>). The analysis in the IHELP report has been widely used to understand student persistence and completion and provides a baseline understanding from which further research can be conducted.

Using the data extracted and cleaned by IRT, I planned to analyze differences between students in the incoming freshman Fall 2008 who graduated, who remain enrolled, and who have dropped out on the following milestones and indicators:

Milestones (things all students must do to complete)

* When remediation was completed (although, this may be less relevant given changes in remediation policy)
* Earn one year of college-level credits (30 units)
* Complete GE+GR coursework
* Graduate

Indicators (things that research indicates successful students are more likely to do)

* Begin remedial coursework in first term (if needed)
* Complete college-level math in the first year or two
* Complete college-level English in the first year or two
* Complete a college success course or first-year experience program
* High ratio of course completion (low rate of course dropping and failure)
* Complete 20-30 units in first year
* Earn summer credits
* Enroll fulltime
* Enroll continuously, without stop-outs
* On-time registration for courses
* Maintain satisfactory academic performance

Limitations in data availability have limited the analyses to the following milestones and indicators:

Milestones:

* First semester grade point average
* Second semester grade point average

Indicators

* Complete a college success course or first-year experience program
* Earn summer credits
* Enroll fulltime

To conduct these analyses, I used SPSS Statistics with previous research and theory driving the analysis of data. The findings from this analysis can be used to compare the results of models using SPSS Modeler, which uses a data mining procedure to analyze the data. However, it is worth noting that currently, SPSS Modeler has not been directly connected to Sacramento State’s Data Warehouse, so the ability to data mine is limited.

**Methods**

To compare students based on their graduated, enrolled, or dropped out status, I conducted one-way analyses of variance tests. This simple test compares the means of each group on the variable of interest, for example percent of students who completed a freshman seminar class.

I used the data extracted by IRT, many of which were proxies for the variable typically used by IHELP in the miles and indicators analyses. See Appendix A for approximations and details on variables used. The variables were extracted from IRT’s Data Warehouse and many, if not all, required multiple steps before being in a form usable for analysis. Details on the creation of the variables should be directed to IRT. I used the variables made available.

**Findings**

Students who graduated, who were still enrolled, and who dropped out (what I will often refer to as “final status” for ease of prose) frequently differed from each other on the milestones and indicators analyzed. See Appendix B for full ANOVA results.

*Enrolling in summer courses.* There was a statistically significant relationship at the p< 0.05 level between enrolling in summer classes after the freshman year and final status [(F = (2, 2978) = 5.514, p < 0.004]. Students who were more likely to enroll in summer classes were more likely to graduate. Those who were least likely to take summer courses were more likely to drop out of college. The difference between graduates (M = 0.03, SD = 0.143) and drop outs (M = 0.01, SD 0.090) was statistically significant. However, it is worth emphasizing that the percent of students who enrolled in summer courses was small. Only 2% of students enrolled in summer courses, ranging from 3% of those who graduated to 1% of those who dropped out.

*Completion of a freshman seminar course.* Completion of a freshman seminar also has a statistically significant relationship at the p<0.05 level with a student’s final status [(F = (2, 2978) = 20.430, p < 0.001]. Students who graduated (M = 0.50, SD, 0.500) had higher means on this variable than than students who were still enrolled (M = 0.45, SD = 0.499). Students who were still enrolled had higher means on this variable than students who dropped out (M = 0.36, SD = 0.481). The means and standard deviations do not represent percentages, as they were coded 0-2. I am not clear what the values for this variable represent, however they do predict a student’s final status.

*Enrolling fulltime in first year.* Enrolling fulltime both semesters in the first year was statistically significantly related to a student’s final status [F = (2, 2978) = 158.539, p < 0.001]. Students who graduated (M = 0.6400, SD = 0.48029) were statistically significantly more likely to enroll fulltime both semesters in their first year than students who were still enrolled (M = 0.4639, SD = 0.49894). Students who were still enrolled were more likely to do so than students who dropped out (M = 0.2568, SD = 0.43707).

*First semester grade point average (GPA).* A student’s GPA in the first semester was statistically significantly related to their final status [F = (2, 2978) = 309.856, p < 0.001]. Students who graduated had higher GPAs in their first semester (M = 3.12004, SD = 0.577245) than students who were still enrolled (M = 2.87435, SD = 0.698843). Students who were still enrolled had higher first semester GPAs than students who dropped out (M = 2.20990, SD = 1.101735).

*Second semester grade point average (GPA).* A student’s GPA in the second semester was statistically significantly related to their final status [F = (2, 2808) = 430.908, p < 0.001]. Students who graduated had higher GPAs in their second semester (M = 3.09815, SD = 0.479276) than students who were still enrolled (M = 2.82727, SD = 0.520785). Students who were still enrolled had higher second semester GPAs than students who dropped out (M = 2.17370, SD = 1.954293).

**Discussion and conclusion**

Note that these comparisons of means to do not demonstrate causality. They simply demonstrate differences between those who graduated, those who were still enrolled, and those who had dropped out. In fact, it is likely (if not probable) that students’ pre-existing trajectories led to the manifestation of these differences. However, it is also likely that these differences perpetuated and further ingrained their different paths. As such, it is worth thinking about how these differences supported students who would end up graduating and harmed those who would not.

For example, a student’s first semester GPA is a reflection of a student’s college preparedness, which, in turn, impacts the students likelihood of completing college. The causal arrow is heavily pointed this way -- however, there is likely an important causal relationship going the opposite way. To a student, grades are a clear signal about his or her ability to succeed. Low grades indicate a lack of readiness, whereas high grades indicate a good fit. Students who receive low grades may take this as a sign that they do not belong in college (or at Sacramento State specifically) or they may take it as a sign that they need to increase their effort to be more successful in the following semester. Sacramento State could intervene at this critical point to help students reshape their mindset, along with providing increased academic support, to increase student success.

Each of the milestones and indicators analyzed above, while not demonstrating causal relationships, should be further investigated to understand the logic that underlies their relationship to student outcomes and how Sacramento State use these understandings to improve outcomes.

Appendix A. Details on variables

|  |  |  |
| --- | --- | --- |
| **Milestone, indicator from IHELP report** | **Data Warehouse variable name** | **Notes on Data Warehouse variable (important specifics, difference from milestone, indicator from IHELP report, etc.)** |
| When remediation was completed (although, this may be less relevant given changes in remediation policy) | Not available/extracted | Students have to do up to 2 semesters of remediation. Students must start in the first semester. |
| Earn one year of college-level credits (30 units) | Not extracted |  |
| Complete GE+GR coursework | Not available/extracted |  |
| Graduate | Grad Term – but used “final status” variable | If blank, that means they didn’t graduate. |
| Complete college-level math in the first year or two | Available but unclear what variable this is |  |
| Complete college-level English in the first year or two | Available but unclear what variable this is |  |
| Complete a first-year experience program: learning community (LCOM) or freshman seminar (any 21 course) | Complete freshman seminar | EOP also has section code 80 or 90. Unclear variable value coding for this. |
| High ratio of course completion (low rate of course dropping and failure) | Not extracted – need units completed |  |
| Complete 20-30 units in first year | Not extracted |  |
| Earn summer credits | Enrolled in summer courses after first summer | Slightly different than earning summer credits. Also, only currently have for first summer. Finally, unclear if this is enrollment at CSUS or anywhere. |
| Enroll fulltime | Yes, by semester |  |
| Enroll continuously, without stop-outs | I believe this is available but unclear what variable this is | Want something that shows if they enroll fall and spring continuously until graduation or dropping out. If they do=1; if they do not enroll one semester = 0 |
| On-time registration for courses | Should be available for extraction |  |
| First term GPA | Available |  |
| Second term GPA | Available |  |

Appendix B. ANOVA results

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Completed a freshman seminar -- significant difference for all but not clear what 0-2 values are…** | | | | | | | | | |
|  |  |  |  |  |  |  |  |  |  |
|  | **Descriptives** | | | | | | | | |
|  | FreshmanSeminarCompletion | | |  |  |  |  |  |  |
|  |  | N | Mean | Std. Deviation | Std. Error | 95% Confidence Interval for Mean | | Minimum | Maximum |
|  | Lower Bound | Upper Bound |
|  | Enrolled | 1054 | .45 | .499 | .015 | .42 | .48 | 0 | 2 |
|  | Graduated | 825 | .50 | .500 | .017 | .47 | .54 | 0 | 1 |
|  | Not Enrolled | 1102 | .36 | .481 | .014 | .33 | .39 | 0 | 1 |
|  | Total | 2981 | .43 | .496 | .009 | .41 | .45 | 0 | 2 |
|  |  |  |  |  |  |  |  |  |  |

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |  |  |  |
|  | **Post Hoc Tests** | | |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  | **Multiple Comparisons** | | | | | | |  |  |
|  | Dependent Variable: | FreshmanSeminarCompletion | | |  |  |  |  |  |
|  | Bonferroni |  |  |  |  |  |  |  |  |
|  | (I) FinalProgramStatusNEW\_n | | Mean Difference (I-J) | Std. Error | Sig. | 95% Confidence Interval | |  |  |
|  | Lower Bound | Upper Bound |  |  |
|  | Enrolled | Graduated | -.056\* | .023 | .043 | -.11 | .00 |  |  |
|  | Not Enrolled | .086\* | .021 | .000 | .03 | .14 |  |  |
|  | Graduated | Enrolled | .056\* | .023 | .043 | .00 | .11 |  |  |
|  | Not Enrolled | .142\* | .023 | .000 | .09 | .20 |  |  |
|  | Not Enrolled | Enrolled | -.086\* | .021 | .000 | -.14 | -.03 |  |  |
|  | Graduated | -.142\* | .023 | .000 | -.20 | -.09 |  |  |
|  | \*. The mean difference is significant at the 0.05 level. | | | | | | |  |  |

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |  |  |  |
|  | **ANOVA** | | | | | |  |  |  |
|  | FreshmanSeminarCompletion | | |  |  |  |  |  |  |
|  |  | Sum of Squares | df | Mean Square | F | Sig. |  |  |  |
|  | Between Groups | 9.921 | 2 | 4.960 | 20.430 | .000 |  |  |  |
|  | Within Groups | 723.025 | 2978 | .243 |  |  |  |  |  |
|  | Total | 732.945 | 2980 |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
| **Enrolled in summer in first summer - only sig diff between grads and drop outs (but few of either group took summer classes in first summer)** | | | | | | | | | |
|  |  |  |  |  |  |  |  |  |  |
|  | **Descriptives** | | | | | | | | |
|  | EnrolledFirstSummer2105 | | |  |  |  |  |  |  |
|  |  | N | Mean | Std. Deviation | Std. Error | 95% Confidence Interval for Mean | | Minimum | Maximum |
|  | Lower Bound | Upper Bound |
|  | Enrolled | 1054 | .02 | .143 | .004 | .01 | .03 | 0 | 1 |
|  | Graduated | 825 | .03 | .165 | .006 | .02 | .04 | 0 | 1 |
|  | Not Enrolled | 1102 | .01 | .090 | .003 | .00 | .01 | 0 | 1 |
|  | Total | 2981 | .02 | .133 | .002 | .01 | .02 | 0 | 1 |

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  | |  | |  | |  | |  | |  | |  | |
|  | **ANOVA** | | | | | | | | | |  | |  | |  | |
|  | EnrolledFirstSummer2105 | | | |  | |  | |  | |  | |  | |  | |
|  |  | Sum of Squares | df | | Mean Square | | F | | Sig. | |  | |  | |  | |
|  | Between Groups | .196 | 2 | | .098 | | 5.517 | | .004 | |  | |  | |  | |
|  | Within Groups | 52.826 | 2978 | | .018 | |  | |  | |  | |  | |  | |
|  | Total | 53.022 | 2980 | |  | |  | |  | |  | |  | |  | |
|  |  |  |  | |  | |  | |  | |  | |  | |  | |
|  | **Post Hoc Tests** | | | | |  | |  | |  | |  | |  | |  | |
|  |  |  | |  | |  | |  | |  | |  | |  | |  | |
|  | **Multiple Comparisons** | | | | | | | | | | | | |  | |  | |
|  | Dependent Variable: | EnrolledFirstSummer2105 | | | | | |  | |  | |  | |  | |  | |
|  | Bonferroni |  | |  | |  | |  | |  | |  | |  | |  | |
|  | (I) FinalProgramStatusNEW\_n | | | Mean Difference (I-J) | | Std. Error | | Sig. | | 95% Confidence Interval | | | |  | |  | |
|  | Lower Bound | | Upper Bound | |  | |  | |
|  | Enrolled | Graduated | | -.007 | | .006 | | .774 | | -.02 | | .01 | |  | |  | |
|  | Not Enrolled | | .013 | | .006 | | .081 | | .00 | | .03 | |  | |  | |
|  | Graduated | Enrolled | | .007 | | .006 | | .774 | | -.01 | | .02 | |  | |  | |
|  | Not Enrolled | | .020\* | | .006 | | .004 | | .01 | | .03 | |  | |  | |
|  | Not Enrolled | Enrolled | | -.013 | | .006 | | .081 | | -.03 | | .00 | |  | |  | |
|  | Graduated | | -.020\* | | .006 | | .004 | | -.03 | | -.01 | |  | |  | |
|  | \*. The mean difference is significant at the 0.05 level. | | | | | | | | | | | | |  | |  | |
|  |  |  | |  | |  | |  | |  | |  | |  | |  | |
| **Enrolled first time for both semesters in first year -- all significant** | | | | | | | | | | | |  | |  | |  | |
|  |  |  | |  | |  | |  | |  | |  | |  | |  | |
|  | **Descriptives** | | | | | | | | | | | | | | | | |
|  | FirstYearEachTermCompletedw12units\_n2 | | | | | | |  | |  | |  | |  | |  | |
|  |  | N | | Mean | | Std. Deviation | | Std. Error | | 95% Confidence Interval for Mean | | | | Minimum | | Maximum | |
|  | Lower Bound | | Upper Bound | |
|  | Enrolled | 1054 | | .4639 | | .49894 | | .01537 | | .4338 | | .4941 | | 0.00 | | 1.00 | |
|  | Graduated | 825 | | .6400 | | .48029 | | .01672 | | .6072 | | .6728 | | 0.00 | | 1.00 | |
|  | Not Enrolled | 1102 | | .2568 | | .43707 | | .01317 | | .2310 | | .2826 | | 0.00 | | 1.00 | |
|  | Total | 2981 | | .4361 | | .49598 | | .00908 | | .4183 | | .4539 | | 0.00 | | 1.00 | |
|  |  |  | |  | |  | |  | |  | |  | |  | |  | |
|  | **ANOVA** | | | | | | | | | | |  | |  | |  | |
|  | FirstYearEachTermCompletedw12units\_n2 | | | | | | |  | |  | |  | |  | |  | |
|  |  | Sum of Squares | | df | | Mean Square | | F | | Sig. | |  | |  | |  | |
|  | Between Groups | 70.542 | | 2 | | 35.271 | | 158.539 | | .000 | |  | |  | |  | |
|  | Within Groups | 662.534 | | 2978 | | .222 | |  | |  | |  | |  | |  | |
|  | Total | 733.076 | | 2980 | |  | |  | |  | |  | |  | |  | |
|  |  |  | |  | |  | |  | |  | |  | |  | |  | |

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |  |  |  |
|  | **Post Hoc Tests** | | |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  | **Multiple Comparisons** | | | | | | |  |  |
|  | Dependent Variable: | FirstYearEachTermCompletedw12units\_n2 | | | |  |  |  |  |
|  | Bonferroni |  |  |  |  |  |  |  |  |
|  | (I) FinalProgramStatusNEW\_n | | Mean Difference (I-J) | Std. Error | Sig. | 95% Confidence Interval | |  |  |
|  | Lower Bound | Upper Bound |  |  |
|  | Enrolled | Graduated | -.17605\* | .02193 | .000 | -.2286 | -.1235 |  |  |
|  | Not Enrolled | .20714\* | .02032 | .000 | .1585 | .2558 |  |  |
|  | Graduated | Enrolled | .17605\* | .02193 | .000 | .1235 | .2286 |  |  |
|  | Not Enrolled | .38319\* | .02172 | .000 | .3312 | .4352 |  |  |
|  | Not Enrolled | Enrolled | -.20714\* | .02032 | .000 | -.2558 | -.1585 |  |  |
|  | Graduated | -.38319\* | .02172 | .000 | -.4352 | -.3312 |  |  |
|  | \*. The mean difference is significant at the 0.05 level. | | | | | | |  |  |

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |  |  |  |
| **First term GPA -- significant for all** | | | |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  | **Descriptives** | | | | | | | | |
|  | @1stTermGPA | |  |  |  |  |  |  |  |
|  |  | N | Mean | Std. Deviation | Std. Error | 95% Confidence Interval for Mean | | Minimum | Maximum |
|  | Lower Bound | Upper Bound |
|  | Enrolled | 1054 | 2.87435 | .698843 | .021526 | 2.83211 | 2.91659 | 0.000 | 4.000 |
|  | Graduated | 825 | 3.12004 | .577245 | .020097 | 3.08059 | 3.15949 | 0.000 | 4.000 |
|  | Not Enrolled | 1102 | 2.20990 | 1.101735 | .033188 | 2.14478 | 2.27502 | 0.000 | 4.000 |
|  | Total | 2981 | 2.69672 | .928213 | .017001 | 2.66338 | 2.73005 | 0.000 | 4.000 |

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |  |  |  |
|  | **ANOVA** | | | | | |  |  |  |
|  | @1stTermGPA | |  |  |  |  |  |  |  |
|  |  | Sum of Squares | df | Mean Square | F | Sig. |  |  |  |
|  | Between Groups | 442.257 | 2 | 221.128 | 309.856 | .000 |  |  |  |
|  | Within Groups | 2125.247 | 2978 | .714 |  |  |  |  |  |
|  | Total | 2567.504 | 2980 |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |  |  |  |
|  | **Post Hoc Tests** | | |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  | **Multiple Comparisons** | | | | | | |  |  |
|  | Dependent Variable: | @1stTermGPA | |  |  |  |  |  |  |
|  | Bonferroni |  |  |  |  |  |  |  |  |
|  | (I) FinalProgramStatusNEW\_n | | Mean Difference (I-J) | Std. Error | Sig. | 95% Confidence Interval | |  |  |
|  | Lower Bound | Upper Bound |  |  |
|  | Enrolled | Graduated | -.245689 | .039270 | .000 | -.33975 | -.15163 |  |  |
|  | Not Enrolled | .664444 | .036396 | .000 | .57726 | .75163 |  |  |
|  | Graduated | Enrolled | .245689 | .039270 | .000 | .15163 | .33975 |  |  |
|  | Not Enrolled | .910134 | .038892 | .000 | .81697 | 1.00329 |  |  |
|  | Not Enrolled | Enrolled | -.664444 | .036396 | .000 | -.75163 | -.57726 |  |  |
|  | Graduated | -.910134 | .038892 | .000 | -1.00329 | -.81697 |  |  |
|  | . The mean difference is significant at the 0.05 level. | | | | | | |  |  |

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |  |  |  |
| **2nd term CSUS GPA -- significant for all groups** | | | | |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  | **Descriptives** | | | | | | | | |
|  | @2ndTermCSUSGPA | |  |  |  |  |  |  |  |
|  |  | N | Mean | Std. Deviation | Std. Error | 95% Confidence Interval for Mean | | Minimum | Maximum |
|  | Lower Bound | Upper Bound |
|  | Enrolled | 1045 | 2.82727 | .520785 | .016110 | 2.79566 | 2.85888 | 0.000 | 4.000 |
|  | Graduated | 823 | 3.09815 | .479276 | .016707 | 3.06536 | 3.13094 | 1.550 | 4.000 |
|  | Not Enrolled | 943 | 2.17370 | .954293 | .031076 | 2.11271 | 2.23469 | 0.000 | 4.000 |
|  | Total | 2811 | 2.68733 | .786444 | .014833 | 2.65824 | 2.71641 | 0.000 | 4.000 |
|  |  |  |  |  |  |  |  |  |  |
|  | **ANOVA** | | | | | |  |  |  |
|  | @2ndTermCSUSGPA | |  |  |  |  |  |  |  |
|  |  | Sum of Squares | df | Mean Square | F | Sig. |  |  |  |
|  | Between Groups | 408.142 | 2 | 204.071 | 430.908 | .000 |  |  |  |
|  | Within Groups | 1329.824 | 2808 | .474 |  |  |  |  |  |
|  | Total | 1737.967 | 2810 |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |  |  |  |
|  | **Post Hoc Tests** | | |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  | **Multiple Comparisons** | | | | | | |  |  |
|  | Dependent Variable: | @2ndTermCSUSGPA | |  |  |  |  |  |  |
|  | Bonferroni |  |  |  |  |  |  |  |  |
|  | (I) FinalProgramStatusNEW\_n | | Mean Difference (I-J) | Std. Error | Sig. | 95% Confidence Interval | |  |  |
|  | Lower Bound | Upper Bound |  |  |
|  | Enrolled | Graduated | -.270877\* | .032072 | .000 | -.34770 | -.19405 |  |  |
|  | Not Enrolled | .653572\* | .030910 | .000 | .57953 | .72761 |  |  |
|  | Graduated | Enrolled | .270877\* | .032072 | .000 | .19405 | .34770 |  |  |
|  | Not Enrolled | .924448\* | .032828 | .000 | .84581 | 1.00308 |  |  |
|  | Not Enrolled | Enrolled | -.653572\* | .030910 | .000 | -.72761 | -.57953 |  |  |
|  | Graduated | -.924448\* | .032828 | .000 | -1.00308 | -.84581 |  |  |
|  | \*. The mean difference is significant at the 0.05 level. | | | | | | |  |  |