Report I.2. Feasibility of Using Cross-Term Student Enrollment Microdata for Measuring Student Progress in Community College Technology Programs

Ron Anderson
Professor Emeritus, University of Minnesota
and Research Associate, Rainbow Research, Minneapolis, MN
rea@umn.edu

Contents

Acknowledgements .............................................................................................. 2
Background, Need and Goals............................................................................... 2
Data Requirements............................................................................................... 4
Data Solicitation and Access (To Be Expanded) .................................................. 5
Data Extraction (To Be Expanded) ....................................................................... 5
Data Restructuring & Management ...................................................................... 6
Problems of Non-Response and Missing Data ..................................................... 8
Additional Issues and Challenges ........................................................................ 9
1. Developmental Courses ................................................................................... 9
2. Transfer credits................................................................................................. 9
3. Simultaneous enrollment in multiple colleges ................................................... 9
4. The Summer Term Challenge ........................................................................ 10
5. Program Completion....................................................................................... 10
6. Data Quality .................................................................................................... 11
7. Tracking Periods ................................................................................................ 11
Indicators ............................................................................................................ 12
1. Outcome Indicators ....................................................................................... 12
2. Process Indicators ......................................................................................... 13
3. Background Indicators ................................................................................... 14
Preliminary Analysis of the Data ........................................................................ 15
1. Demographic Differences ............................................................................... 18
2. Momentum and Deceleration Forces .............................................................. 22
3. Skips or Gaps in Enrollment ........................................................................... 24
4. Technology Programs .................................................................................... 26
Conclusions ........................................................................................................ 28
References ......................................................................................................... 29
APPENDIX A. Codebook for Completion Data................................................... 31
APPENDIX B. DATA ACQUISITION AND SOLICITATION DOCUMENTATION 35
Acknowledgements

This project was funded by the National Science Foundation ATE Program for Targeted Research. The grant went to Colorado University’s DECA Project, Liesel Ritchie, Principal Investigator (PI), with a subcontract to Rainbow Research for Project I, Strategies for Improving Recruitment, Retention and Placement. I wish to especially thank Wayne Welch, Co-PI on the project, Liesel Ritchie, Director of the DECA project, and Martin Spielauer, Senior Research, Statistics Canada, for his construction of the microsimulation model, MicroCC. This project would not have been possible without the extremely generous time of William LeBlanc, Director of Institutional Research and Planning at Community College of Rhode Island, and his associate, Jeanne Savastano, and of Corby Coperthwaite, Director of Planning, Research and Assessment for Connecticut Community-Technical Colleges, and of Benjamin Klimczak, Senior Research Associate of Connecticut Community Colleges. In addition Vanessa Morest, Dean of Institutional Effectiveness at Norwalk Community College, has been invaluable in supporting this project with ideas and encouragement.

Background, Need and Goals

Recent trends in the workforce supply and demand point to the need for the community colleges to improve their ability to monitor and anticipate changes in supply and demand for different types of community college education, especially technician programs. Part of such monitoring is tracking students as they progress or flow through the student pipeline. Improved modeling and projection tools are needed to refine and improve the student pipeline, which encompasses recruitment, development, retention, and placement.

Community colleges are known to be far more heterogeneous than the typical 4-year college or university. They typically have a broad variety of programs, students enter at many different entry points, and students bring great variability in prior work experience, educational backgrounds, and demographics. All these factors affect student progression and outcomes. Unequal success rates by race and SES (socio-economic status), with the dependence of many students on financial aid, make planning and policy-making even more challenging for community college programs.

In recent years, important advances have been made in building longitudinal databases of community college students and in analytical strategies for their analysis (cf. Hagedorn & Kresse 2008). To our knowledge, these advances have not been applied separately to technician programs. We propose to explore the feasibility of doing this type of student tracking analysis on programs of the type funded by NSF’s ATE (Advanced Technological Education) grants. In addition, our goal is to refine enrollment time series data as measures of completion.

The longitudinal analysis of students in technician programs, as well as other academic programs, will provide a basis for a much greater understanding of the forces and impediments to a wide variety of measures of progress. These measures include course completion rates, program retention, student persistence, persistence by major or program, rates and types of transfer in and out of the program, graduation rates, and job placement rates. Once these measures are defined and implemented, a wide variety of additional statistical analysis can be done exploring the role of demographic and other background factors as well as process factors such as financial aid, remedial work, and so on. The potential for analysis ranges from cross tabulations to structural
equation modeling, such as Hagedorn, Moon, Maxwell & Pickett’s (2002) model of retention. Event history, hazard regression, and logistic modeling could contribute greatly to understanding technician education pathways, which will inform decision makers in ATE projects and technician education more broadly. Our focus also is on evaluating enrollment time series microdata for these kinds of analysis.

Leinbach and Jenkins (2008) demonstrated how such analyses can be done using the longitudinal data from the Achieving the Dream (AtD) project. In addition, they discovered how certain accomplishments such as completing the first course can be powerful in predicting completion of a milestone. They called these events momentum points and they found that the points differ across subgroups. Leinbach and Jenkins define subgroups of students not only in terms of past attributes like demographics but also in terms of students’ objectives for the future, which is consistent with our interest in focusing upon several subgroups. However, the AtD data do not allow disaggregating students by program until students have completed a program.

In this type of analysis, subgroups are typically separated in terms of students’ plans to transfer to a 4+ year institution, obtain a certificate, or obtain a 2-year degree. For each of these subgroups Leinbach and Jenkins found one or several events such as completion of the first 15 credits that are highly correlated with successful completion of the intended objective. This type of momentum point analysis has served as an effective basis for several community colleges in Washington state to improve their student success rates. These improvements are based upon institutional attention to making the momentum points more efficient, for instance, helping more students to complete the momentum points.

In contrast to multivariate and causal modeling, an important analytical and policy refinement tool is microsimulation. Microsimulation models can best help in evaluating the effect of demographic or policy changes over time. They have been pioneered in welfare analysis, taxation modeling, and felony processing as well as population aging, traffic control and public health. A good way to think about them is that they are projection systems that provide for variation in input settings, thus allowing for the estimation of long term consequences of policy changes.

Another goal of this project is to determine to what extent student progress microdata would be available for analysis and construction of microsimulation models to project and explore the implications of retention and other pipeline processes within community college education. Ultimately, models could be developed not only at an institutional or program level but also at the level of a state, a collection of states, or the nation as a whole.

For such models to become operational it is necessary to obtain estimates of the percent (for each relevant demographic group) who move (flow) through the system. Differences among racial minorities and other underrepresented social groups need to be emphasized. If demographic characteristics such as gender and race inhibit large sectors of the potential workforce from participating in education and placement, this would be a major concern because of the loss of these students from the workforce. From the published literature, we know that gender and race subgroups do pass pipeline milestones with differing rates of success. The design of the microsimulation model will incorporate the movement of each gender and race-based demographic group through the pipeline system. Thus, it will be possible to increase our
understanding of differences among these subgroups as they move through each stage of the pipeline.

In summary, here are the principal research questions:

1. What are the events that serve as major obstacles or momentum forces for students in community college education, especially engineering, science, information technology, and other technology-oriented technician programs? Do these differ by the type of program or by the students’ objectives (e.g., certificate, degree, transfer)?
2. What is the role of gender, race/ethnicity, and age in the student pipeline, especially in the role of momentum points or obstacles? What are the implications of using different measures of retention (e.g., course completion rates, program retention, student persistence, and types of transfer in and out of a program) upon comparisons among these demographic groups?
3. Is it feasible to access, organize, and apply cross-term, time-series student enrollment data (also called student unit record (SUR) data) for measuring retention and completion and to explore answers to the above questions?
4. Is it feasible to obtain the requisite data to build a useful microsimulation model of the technician education pipeline, or of community college students in general? Determination of feasibility would be based upon tests of validity on baseline populations and on the capacity for the model to explore useful “what-if” questions that eventually might include evaluation of the effect of policy change at different levels.

We would emphasize that the purpose of this document is not to present a model of retention for technology programs in community colleges. Rather, it is to evaluate the quality of the data that can be used in estimating study progress and success, with an eye toward construction of a simulation model for strategic planning.

Data Requirements

The principal resources required for community college program completion modeling are data containing samples of cases structured across terms, which is often referred to as student unit record (SUR) data. The data desired will have fixed background characteristics as well as attributes that change across time. The most manageable time unit for this analysis and simulation is a school term, of which colleges typically have two per year plus summer. As recommended by Leinbach and Jenkins (2008), the ideal number of years for analysis is five years. It would be possible to work with as few as two or three years of data, but then many of the students will not have had sufficient time to complete their program. Using more than five years of data may not be appropriate for current policy issues.

Appendix A gives a complete list of requested data elements for the monitoring and modeling proposed. The first group of data elements in that list contains fixed characteristics for each student and the second are states that change term by term. This core subset of data elements does not require student grades. It does, however, require course credit and completion information, which implies passing rather than failing specific numbers of courses or credits attempted.
The fixed background factors desired include race/ethnicity, gender, and highest level of education of any one parent, living with parents, marital status, dependent children, financial aid or loan received, concurrent employment, financial support from parents, highest level of education completed, previous degrees or certificates, and identifiers of the first term and year of the tracking period.

The variables desired that potentially change from term to term include the targeted (desired) or enrolled program (e.g., pharmacy technician). They also include the total credits attempted and earned, type of completion status (e.g., graduation), type of transfer, or other types of departure. These characteristics of each student would necessarily have to be collected after the end of the term and assigned to the student-term later. Based upon modeling objectives, we included financial aid and marital status. However, we were aware that there might be considerable missing data for these data elements.

Data Solicitation and Access

Over a period of about six months, contacts were made with representatives of community colleges, research projects, and state departments of education. Contacts were made with community colleges and agencies known to be active in either NSF ATE activity or in community college data projects. Phone calls and e-mailed letters, such as the letter in Appendix B were used in this solicitation.

In January 2010, a representative of Boston Area Advanced Technological Education Connections (BATEC) assisted in making contact with the New England Board of Higher Education (NEBHE). Fenna Hanes of BATEC arranged for the President of NEBHE to send a letter to each of the Chancellors or Commissioners of the community college system in each of the New England states. They were asked to invite the institutional research director of the community college system or of individual community college systems to participate in the project by providing data for analysis.

Two state community college systems, RI and CT, agreed to participate. Data files were assembled over a period of several weeks. The preparation of data involved numerous phone calls and e-mails to clarify priorities and issues.

Data Extraction

Data extraction required extensive time, effort and skill on the part of the IR staff to create a file of student attributes across terms. The challenge stems from the different types of data systems needed for institutional queries and reporting on the one hand and research analysis on the other. CCRI (Community College of Rhode Island) is the only community college in the state but has several campuses. They provided flat or rectangular files from their relational database system for each of four cohorts beginning with 2006-06 through 2008-09. The data arrived as four Excel files following the codebook with a single line or record for each student who was admitted in the first term of the cohort. In the analysis reported here, only a single cohort beginning with the fall term of the 2005-2006 academic years was used. Data were provided for 4.5 years.
ending in fall of 2009. The other cohorts had 3.5 or fewer years of data, so they set aside for later analysis. The cohort used in this report had 2,502 students.

The CTCC (Connecticut Community College) system is much larger and has 12 separate community colleges. The file received from them had 689,938 term level records for 276,469 students and included 10 cohorts, the first one beginning in the 1999-2000 year and the most recent one beginning fall 2006. The first cohort had 20 terms for a total of 10 years. The last one had only 2 terms ending with the spring 2009 term.

Data Restructuring & Management

The CCRI data file did not need to be restructured, however, the CTCC had to be transformed into a student by student file rather than a student-term by student-term file. The size of the CTCC data file, over 40MB in an Excel file, made it prohibitive for them to restructure, so that process was done in Minnesota with the Stata statistical system. Once the newly structured file was created, it was converted to an SPSS file for analysis.

The CTCC file contained 28 variables and 689,938 records. The table on the next page shows the list of variables in the file. The key fact is that the 12 variables from var#17 thru 28 are a repeating group. The first 16 variables on each record represent the student. The repeating group represents each term in which the student was enrolled. (The first 16 variables are replicated on each of the student's term records except that variables 1 and 2 (year and term) change so as to coincide exactly with variables 17 and 18 (yearX and termX).

The complicating aspect of this file is that there are 10 cohorts, beginning with academic year 1999-2000 and ending with year 2008-2009. There were 3 terms for each year, so there were 30 terms for the earliest cohort, 27 for the next, and so on. We had to derive the starting years and terms and add them as variables to the file. We could have created 10 separate cohort files, but chose for convenience to create a single file with all 10 cohorts. The complexity was reduced by limiting the data to two terms (Fall and Spring) per year. As the summer term yielded only negligible additional credits for students enrolled in the Fall and Spring terms, dropping the summer terms did not yield a significant loss in data.

The final file structure had one and only one record per student. Furthermore, it was structured so that the first repeating group on each record was the first term for each student no matter which cohort they fell into. Alternatively, we could have allocated a fixed set of fields for each term and year. The structure we created made it possible to create time series graphs for all cohorts simultaneously, which of course made the process of analysis quicker.

Here is the list of the first few variables to illustrate the structure of the new file. The first 16 data elements or variables are the initial or starting states of each attribute. The next 12 variables (from 17 to 28) are the first repeating group for the first term for every student. Note that the number “1” is appended to the variable name of each of these 12 variables. The next subset of 12 variables is for the second term and a “2” is appended on each variable name. This pattern continues for 20 terms, however by the
20th term there only a handful of students because those would be the students that had been enrolled over a 10 year period.

<table>
<thead>
<tr>
<th>VAR NAME</th>
<th>variable label</th>
<th>Seq #</th>
</tr>
</thead>
<tbody>
<tr>
<td>YEAR</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>TERM</td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>STATUS</td>
<td>starting status</td>
<td>3</td>
</tr>
<tr>
<td>ID</td>
<td>student ID</td>
<td>4</td>
</tr>
<tr>
<td>COLLEGE</td>
<td></td>
<td>5</td>
</tr>
<tr>
<td>GENDER</td>
<td></td>
<td>6</td>
</tr>
<tr>
<td>RACE</td>
<td></td>
<td>7</td>
</tr>
<tr>
<td>AGE</td>
<td></td>
<td>8</td>
</tr>
<tr>
<td>LIV_PAR</td>
<td>live with parents</td>
<td>9</td>
</tr>
<tr>
<td>MARITAL</td>
<td>marital status</td>
<td>10</td>
</tr>
<tr>
<td>CHILDREN</td>
<td></td>
<td>11</td>
</tr>
<tr>
<td>ED_PAR</td>
<td>educ of parents</td>
<td>12</td>
</tr>
<tr>
<td>ED_PRIOR</td>
<td>prior educ of student</td>
<td>13</td>
</tr>
<tr>
<td>AID</td>
<td>financial aid</td>
<td>14</td>
</tr>
<tr>
<td>AID_PAR</td>
<td>fin from parents</td>
<td>15</td>
</tr>
<tr>
<td>JOB</td>
<td>current job</td>
<td>16</td>
</tr>
<tr>
<td>YEAR1</td>
<td>academic year</td>
<td>17</td>
</tr>
<tr>
<td>TERM1</td>
<td>term</td>
<td>18</td>
</tr>
<tr>
<td>PROGRAM1</td>
<td>major field of study</td>
<td>19</td>
</tr>
<tr>
<td>CE1</td>
<td>courses enrolled</td>
<td>20</td>
</tr>
<tr>
<td>CC1</td>
<td>courses completed</td>
<td>21</td>
</tr>
<tr>
<td>DCE1</td>
<td>dev courses enrolled</td>
<td>22</td>
</tr>
<tr>
<td>DCC1</td>
<td>dev courses completed</td>
<td>23</td>
</tr>
<tr>
<td>CRE1</td>
<td>credits enrolled</td>
<td>24</td>
</tr>
<tr>
<td>CRC1</td>
<td>credits completed</td>
<td>25</td>
</tr>
<tr>
<td>DCRE1</td>
<td>dev credits enrolled</td>
<td>26</td>
</tr>
<tr>
<td>DCRC1</td>
<td>dev credits completed</td>
<td>27</td>
</tr>
<tr>
<td>OUTCOME1</td>
<td>outcome status</td>
<td>28</td>
</tr>
<tr>
<td>YEAR2</td>
<td></td>
<td>29</td>
</tr>
<tr>
<td>TERM2</td>
<td></td>
<td>30</td>
</tr>
<tr>
<td>PROGRAM2</td>
<td></td>
<td>31</td>
</tr>
<tr>
<td>CE2</td>
<td></td>
<td>32</td>
</tr>
<tr>
<td>CC2</td>
<td></td>
<td>33</td>
</tr>
<tr>
<td>DCE2</td>
<td></td>
<td>34</td>
</tr>
<tr>
<td>Etc</td>
<td>Etc</td>
<td>Etc</td>
</tr>
</tbody>
</table>
The above table is the top section of the first file restructured. This file had a total of only 304 variables. However, we discovered that many of the so-called fixed data elements changed over time. For example, college changed because students often switched colleges within the system. (Some even were enrolled in 3 or 4 separate colleges at the same time, which will be addressed in a later section.) Even demographic variables like marital status changed over time and of course age progresses. And students change majors or program during their tenure.

This resulted in a file with over 500 variables including some new indicators at each term for such factors as cumulative credits earned and the presence of failed or terminated courses. The final working file contains 927 variables for 126,482 students. The reason that this file contains many fewer students than the original is that the most recent 4 cohorts were dropped from the analysis (which amounted to 34% of the students) because they did not have a full 5 years of enrollment history. In addition, all the students that were returning students at the time of the first term were dropped from the analysis (23% were identified as returning students.) Eliminating these two groups of students resulted in the total number of students dropping from 276,469 to 126,482. All of these dropped cases are available for later analysis, of course.

Once the restructuring was accomplished with the CTCC data, the recoding of variables from both systems’ files began. This recoding will be detailed in later sections, but here it should be noted that the code restructuring encompassed program type as well as enrollment variables. In addition, missing values had to be handled consistently across variables and data sets.

**Problems of Non-Response and Missing Data**

The codebook specified that missing values (including non-responses and unavailable data) should be coded as “9” for single-column fields and “99” for double-column fields. All of these values were either defined as internally missing within the working files or defined as zeroes. For certain variables such as cumulative counts of credits, missing values are equivalent to zeroes because zero is an easier quantity to work with in calculations.

One of the more complicating factors was identifying when a student had truly dropped out because some community college students drop out for one or several terms at a time. In both data files, a missing values code for all of the variables for a given term signified that the student had not registered. This meant that it was either a permanent or a temporary non-enrollment.

Such cases had to be distinguished from students who attempted courses by enrolling in them but never completed them because of a failed grade. (Both college systems defined pass or completion of a course as receipt of a grade of D or higher.) In such instances, the course enrollment would be a course count but the course completion would be a missing value. Such complexities make missing values one of the most challenging aspects of data management and analysis.

While data on a number of background or demographic variables (e.g., children, marital status, etc.) had been requested, this information was not generally available. CCRI used Accuplacer, a database of those taking the recommended computerized placement test, to obtain partial data for the following variables: children, concurrent job,
parents’ education level, and student’s education level. Because as many as 75% of the students were not in the Accuplacer database, these data elements were not used in the current analysis. CTCC was not able to provide valid response for many of their students either. Both systems were able to supply information about financial aid for a few students, and some analysis was done on whether or not the student received financial aid, but the patterns were not stable and the majority of students were missing on this variable.

Additional Issues and Challenges

1. Developmental Courses

Developmental courses, formerly called remedial courses, are mandated courses based upon admissions test scores but they are courses for which the credits do not count toward graduation, certificate, or transfer credit. In the CCRI system, the developmental credits are included in the attempted but not in the earned credits. The developmental courses are also included in the count of courses attempted but not in the earned. In the CTCC system, developmental courses are only included in three of the 12 community colleges, and they are included only for attempted and not for completed course counts or credits.

Because developmental courses and credits are included with other courses and credits in the enrollment or “attempted” counts, the only way to calculate the total credit and course enrollments is to subtract developmental units from the total enrollment (attempted) counts. This was a major reason for including the developmental courses and credits in the data codebooks. Another reason was to be able to analyze the impact of developmental course taking and failures upon performance and completions otherwise.

2. Transfer credits

In CCRI, transfer credits are included in both the attempted and earned credit counts, but not until the student declares a major. So they may not be included for a student for their first few terms. Transferred credits do not appear as transferred courses because they don’t know how many courses were taken to receive a given number of transferred credits. The CTCC system course and credit counts do not include any transfer credits. This difference between the two systems may account for small differences in course and credit completions between the two systems.

3. Simultaneous enrollment in multiple colleges

In the CT data we found the following: There were 7,061 student-terms with 2 college enrollments, 227 student-terms with 3 college enrollments, 18 student-terms with 4, 1 with 5, and 1 with 6. Most of these have the same course/outcome information, but that is not the case for all of them; some colleges apparently kept their records
separately. To collapse cases to one record per student, we tried making "college" the last college the student was enrolled in before the multiple-enrollment occurred, but many students started out in multiple colleges. We opted for assigning the college in which the student had enrolled in the most often, and picked one at random in the case of ties.

CT Institutional Research extracted and sent us the data and reported that the course counts across multi-terms for a student should not be overlapping, that is, more than one college should not count the same credit twice, which would be equivalent to double-counting. Apparently, students enrolling in multiple institutions typically have credits transferred to their home institution, but that these courses should not show up on both of the college records for a single term in the data file. Our procedure was to combine these data unless the total credits completed was over 18, which is equivalent to over 6 typical courses in any one term.

4. The Summer Term Challenge

Summer term data were included in the RI file but not the CT file. CT reported that they did not include the summers because they don't centrally keep track of the number of courses attempted during the summer. They keep track of the number of credits only. Furthermore, we found that they have a cumulative credits indicator that includes not only credits earned in the summer term but also credits that are successfully transferred in from another institution either within or outside of CT. However, that is not an indicator provided for our data file.

While the data for summer terms were provided by CCRI, in the majority of our analysis, they have been dropped. We found that only a very small number of regular students enrolled in summer school. The effect of dropping summer terms completely on enrollment and completion trends was negligible. Therefore, they were dropped in the interest of more readily interpretable trend lines in the analysis.

5. Program Completion

For the data extraction, a program completion variable labeled "outcome" was requested for each student-term. That attribute had graduation or certification, transfer to 4-year, dropout or no status change. The resulting distribution for outcome from RI for 13 terms had only 1% graduation, very few dropouts, and 17% with transfers. Transfers were estimated by RI by checking the National Student Clearinghouse for discontinued students to see if they were enrolled in a different college. That process did not distinguish between the types of colleges, so transfers in the RI data included transfers to other community colleges as well as transfers to colleges offering 4-year degrees. This transfer information was not available in the data received from CT. In fact, the graduation and certificate completion data were so minimal (e.g., 1 percent of students) for CT so as to be not usable.

For both databases, we added estimates of “transfer readiness,” as a measure of completion. Transfer readiness was defined as completion of X number of credits within a specified number of years. Usually the transfer readiness criterion used was 48 credits, which is equivalent to 16 courses or a full-time load of 4 courses for each of 4
terms. Additional criteria for completion of credits were used, namely 30 credits and 15 credits. Using 48 credits as an additional criterion for completion with the RI data added 23% to graduation and transfers, giving us a total of 41% completing a program within 4.5 years. This would seem to have some validity because the national community college completion rate within five years that is often used in news articles is 40%.

Using 48 credits as the criteria for completion with the CT data gave us a total of only about 15% completion for that system. As this seems to underestimate completion, in the analysis we tended to use 30 credits more often than 48 credits as the operational criterion for student completion. It has become apparent that the majority of community colleges do not track completions precisely, which is why there is consideration interest in refining “measures of success” for community colleges.

6. Data Quality

In this type of secondary analysis project where one is removed from the original data generation, it is very difficult to assess data quality. Consistency checks across indicators can be and were made. No data errors of this type were found in the CCRI data but a few were found in the CTCC data, which was many, many times larger and hence more subject to error. The identified problems occurred when the completed course counts were subtracted from the attempted course counts, after removing all cases with missing values for either of these variables. In a handful of cases, the results were minus numbers, which is logically inconsistent. Another problematic area was that in less than 100 out of 276,469 students, their age was listed as less than zero. These erroneous values were assigned as missing values so that they would not affect calculations. Given the enormous size of the files, these problems did not seem worthy of additional attention at this time.

7. Tracking Periods

As noted already, in the CTCC system data the records of student cohorts after 2004-2005 were set aside and not included in the analysis. This reduced the total size of the population of students and it also prohibited analysis of more recent student cohorts. When the objective is to compare recent student performance with earlier years, then they will be added back in. However, generalizations about completion would be limited to fewer numbers of terms. For the CTCC data our intention was to have at least 5 years on which to base completion, so the last 4 cohorts were dropped. As can be seen in Figure 2.1, there appears to be a slight increase in completion rates for the cohorts in the mid-2000s compared to those from the early years of the decade of the 2000s. To determine if this trend is continuing, it would be possible to examine that question by adding the later years back into the analysis file.

Apart from the unavailability of data, there is another argument for using shorter enrollment periods for the completion measures. For some of the indicators of cumulative progress, the trend lines flattened out and no longer rose after about 5 or 6 years. Because of this tendency the trend line graphs using in the analysis section sometimes continue for 10 terms and other times for 12 terms. The data were coded for
analysis of 14 terms, but in the interest of retaining nearly similar proportionality in the size of the graphs, most of the analysis was based upon 12 terms (6 years) of analysis.

Indicators

The indicators used in the analysis are defined into three categories: outcome, process, and background. They are listed and discussed in that order. The operational definitions of each are cryptic. Most created indicators are defined as the value “1” if the characteristic is present. If not, the cases are coded as “0.” In most instances, but not all, a missing value category has been defined if data are missing.

1. Outcome Indicators

Most of the outcome indicators were measured at the term level, which allowed for the construction of indicators across the careers of the students at the colleges. This type of indicator was named with a capital “X” at the end of the name. The remainder of the 31 outcome indicators listed below, do not have an “X” in the name and represent a single measure related to completion or retention. Most of the analysis done so far concentrates mostly on those indicators than span multiple terms because they capture the process or evolution of student performance over time.

Acc16 is term when 16 courses first completed.
Acc48 is term when 48 credits first completed.
ActiveX is 1 if any course(s) completed in term X.
AnydceX is 1 if student enrolled in any developmental course in termX.
Ate is any technology major at initial term.
AtesexX and ateraceX are recoded versions of techX with other defined as sysmis.
AteX is 1 if any tech major by termX otherwise 0 (collapsed version of techX).
Complete is 1 if at least 15 total credits completed.
CompletedX is 1 if 30+ credits completed.
CompleteX is 1 if 15+ credits completed by termX.
CreditsX is count of credits completed each X with missing set to zero.
CumcoursesX is cumulative sum of totcoursesX completed at termX.
DevfailcrX is 1 if student failed any developmental course in a given term X, otherwise 0.
Drop is 1 if last < 13 or finish13 = 0.
DropoutX is 1 if student did not enroll in any later terms without first earning 48 credits.
EnrX is 1 if enrolled in any course during term X.
Eveready is 1 if 48+ total credits by last term in series.
FailX is 1 if the subtraction of earned courses from attempted courses is > zero during termX.
FinishX is 1 if student graduated or transferred or completed 48 credits.
Last is the term after which no more enrollments occurred within 13 terms for RI & 14 for CT.

OutcomeX is a 1 if reached graduation, transfer, or transfer readiness (16+ courses) in termX.
PcompletedX is percent of attempted courses that were completed in term X.
Persist2000, 2001, 2002 is 1 if enrolled in spring as well as fall term of each of 3 years (CT only).
Persist2006, 2007, 2008 is 1 if enrolled in spring as well as fall term of each of 3 years (RI only).
Persist ance is the sum of years in which persist ance (i.e., continued from fall to spring) occurred.
Persist ance2 is persist ance recoded so as to collapse values above 1 to 1.
ProgramX: see background indicators below.
SaveoutX is unrecoded (uncollapsed) version of outcomeX.
SuccessX is cumulative of outcome, that is, any completion outcome including 16+ courses.
TechX is count by type of technology program intended or enrolled for each termX.
TotcoursesX is total number of courses completed in termX.
TotcreditsX is total number of cumulative credits completed by termX.
TreadyX is 1 is 48 credits completed by term X.

2. Process Indicators

Process indicators constitute significant events that affect the evolution of performance and completion over time. A momentum point is such an indicator. While we created two momentum point indicators (mpt1 and mpt2) only mpt1 was used because it included more students and because it had been shown to be effective in other research. The remainder of the process indicators captured skips or gaps in course registration over time. The skipX indicator captured whether a student skipped any given term; the remaining skip indicators were simply the sum of skipped terms by any given student.

Mpt1 is 1 if cumcourses1 is 4+, representing a year 1 equivalent momentum point.
Mpt2 is 1 if cumcourses2 is 8+, representing a year 2 equivalent momentum point.
Skip sum is the total of all terms skipped before last enrolled term.
Skip sum3 is skip sum with higher values of skip sum collapsed into 3.
Skip sum5 is skip sum with higher values of skip sum collapsed into 5.
SkipX is 1 if a skip occurs during term X, which can occur only if X is smaller than last.
3. Background Indicators

Below is a list of the demographic and other background variables included in the data files. The analysis to date has concentrated upon gender, race, and age, and to some extent fulltime versus part-time enrollment and “living with parents.” The remaining variables were largely ignored because of the large amount of missing data already mentioned.

Age is the actual age in years of the student at the first term.
Age19 and Age19ct are 1 if age is 19+ otherwise 0.
Agecat2 is age in 3 categories: to 20; 20-25; and 26+.
Aid is 1 if financial aid received .
Ate is 1 if technology program major declared in first term.
Ed_par parents educ: high-sch college
Ed_prior is 1 if prior college.
Children 1 if children (only 50 but 2047 missing).
College is the college enrolled in at term1.
Gender is 1 for male and 2 for female.
Fulltime is defined as those taking 4+ courses the first term; all others are part time.
Fulltime2 is 1 if totcourses1 is 4+ & totcourses2 is 4+, which is equivalent to 8+ courses in year1.
Full is same as fulltime except it is based on 12+ credits completed.
Job is 1 if known job.
Liv_par is 1 if living with parents, 2 if not (40% missing).
Marital is 1 if married (only 147).
Program is the un-recoded CIP code for student major or program.
Race categories are white black Latino other and 99 for missing.
Racecat4 collapses missing values with other while racecat2 collapses all except white into other.

Some of these so-called background indicators can also be considered outcome indicators. For example, ATE represents a technology program selection for term1. It is also part of the AteX measure listed in the outcome indicator section above.
Preliminary Analysis of the Data

As noted earlier, CCRI (Community College of Rhode Island) is the only community college in the state of Rhode Island and they provided several rectangular files. However, because of the limited time span of some cohorts, only the cohort of students entering in the fall term of the 2005-2006 academic years was used. This cohort had 9 non-summer terms for an entire period of 4.5 years ending in fall of 2009. This cohort had 2,502 students, and these are the CCRI students shown in the charts in this section.

The CTCC (Connecticut Community College) system is much larger and has 12 separate community colleges. Their file received had 689,938 term level records for 276,469 students and included 10 cohorts, however, only the first six cohorts were used in this analysis. The six cohorts were those students entering CTCC for the first time in the fall terms of 1999, 2000, 2001, 2002, 2003, 2004, and 2005. These cohorts add up to 126,482 students. The graphs for CTCC in this section include all of these students.

The purpose of this section is to describe the time-series student data from the two community college systems to illustrate the challenges and character of the data. These data reveal the effects of demographic and process factors upon student progress.

A variety of outcome measures are used to illustrate the implications of selecting specific indicators of student success. A second purpose of displaying a variety of indicators over time is to identify the indicators that should be built into any microsimulation model that follows this report.

Figure 1 Cumulative Percent Completion* by Gender, CCRI 2005 Cohort

*Completion includes any of the following: graduation with a degree or certificate, transfer to another college, or completion of 16 courses (the equivalent of completing full time coursework for two years).

While Figure 1 reveals a gender gap, its main purpose here is to illustrate the gradual slope of students as they reach completion. It shows that 40% of CCRI students
in the fall 2005 cohort, of which there were 2,502 students, gradually reached completion status by the end of 9 terms (4.5 years). The completion slope for the first two years is slightly steeper for the first two years, no doubt a consequence of full time students finishing up and leaving the system during that time. The shape of the slope suggests that a significant number of students may complete their programs after the 9 terms. The gender difference will be discussed later.

In the following trend charts (Figure 2 and 3) a different outcome measure as well as population are depicted. The data are the 126,482 students in six cohorts of the CTCC system. The outcome indicator is the cumulative average credits earned by the students in each cohort. After 10 terms (5 years), only 20 to 25% credits have been completed on average. This average is relatively low because of the large number of dropouts and the large number of part time students in typical community colleges.

Note that the lines represent cohorts and these cohorts include students from 12 colleges. While the cohort lines can not be easily distinguished in black and white, the more recent cohorts reveal an increase in completion rate compared to the earlier cohorts. This indicates that student progress, in terms of average credits earned, was improving in CTCC over time.
In Figure 3, the lines represent the 12 colleges. The outcome measure, average credits earned, is the same as Figure 2, but this chart graphs 14 terms (7 years) instead of 10. It reveals that the growth of credits earned does not increase substantially after the 12 term (6 years).

The next graph (Figure 4) shows the percent registered or enrolled in any course for each term. The chart represents the CTCC students showing the typical decline in registration (enrX) for one or more courses over their life history with the college. While over 126,000 students were enrolled in courses during their first term (by definition) by the end of their 14th term, only a couple of hundred students in those cohorts were still enrolled.

Keep in mind that this pattern is for courses attempted, not courses earned or completed. If the chart were to depict student course completions per term, it would of course reveal lower percentage estimates for the entire time.
1. Demographic Differences

Gender differences in completion are depicted in the next chart (Figure 5).

However, in contrast to Figure 1 where males were shown to be more likely to complete than females, this trend line depicts dropout rates. Males were more likely to
dropout than females after the second term, which means that dropouts do not account for the higher rate of males than females completing their programs.

We did not find gender differences in course failures within the CTCC system, so it would appear that external factors, rather than process variables, account for the gender differences in completion. Some interesting gender differences have been uncovered with respect to technology programs; however, they will be discussed in a later section on technology programs.

Turning now to differences by race, the following chart (Figure 6) shows typical patterns of completion with whites and “others” completing their programs more rapidly than African Americans and Latinos in CCRI. This pattern is replicated in the CTCC system.
If we shift (see Figure 8) to the average courses completed term by term, we see only a small decline for those remaining registered. The chart below shows students in CCRI, and the average number of courses completed only drops from about two to 1.5 courses per term. The younger age group had a higher rate of enrollment across time. This pattern of age differences tends to be replicated across other indicators of completion.

A partial explanation for these differences can be seen in Figure 17 which shows the average terms skipped. Figure 17 contrasts racial groups and shows that groups with lower completion rates have higher stop-out or term-skip rates, which of course slows down their progress.

![Figure 8. Percent of Courses Completed Each Term by Age Group: CCRI 2005 Cohort](image)

Figure 9 returns to the outcome indicator of average credits earned. This time the trendlines represent age groups. These data confirm that for both community college systems, the younger age groups have more rapid progress.
A similar advantage can be seen when contrasting those that live with their parents versus those who don't (Figure 10). Thus, the advantage clearly does not accrue to age itself but to benefits that younger age students have in terms of time and financial support.
2. Momentum and Deceleration Forces

Leinbach and Jenkins (2008) found that certain milestones such as completing the first 16 credits among Washington state community college students gave them a boost in terms of completion of an entire program. This is graphically depicted in the chart below (Figure 12) with a steep rise in program completion as measured by “outcomeX,” particularly during terms 3 and 4 (the student’s second year).

The opposite of momentum force can be called deceleration or impediment, and this can be visualized in Figure 14 where taking any developmental course slows down progress toward completion. Having to take developmental courses (remedial courses) produces deceleration or negative momentum toward program completion too.

Although not depicted in any charts, any course “failure” during one’s first term results in an impediment. The distinction feature of these impediments is that their effect persists throughout the students’ career at the college.
Attending part-time functions similar to event-based impediments. Figure 15, which contrasts part-time with full-time enrollees, shows the dramatic difference in progress of these two groups over time.
Not surprisingly developmental (remedial) course demands curtail progress. Figure 16 shows the percent of average developmental course failures across time. In CCRI about 10% of the students in the first term of the 2005 cohort failed one or more developmental courses. This declines to less than 1% by the 9th term, but there were still a few students continuing to fail developmental courses during their 4th year in the cohort.

Figure 16 also shows that there was a non-white, compared to a white, disadvantage in terms of the relative number of students failing developmental courses. The points to an area that colleges might focus on to improve completion rates.

Figure 16 Percent of Developmental Course Failures by Race: CCRI 2005 Cohort

3. Skips or Gaps in Enrollment

The next two figures (Figures 17 and 18) show how stop-outs (skipped terms) interfere with student progress. Stop-outs apply only to students that re-enroll or register before the term sequence expires. For CCRI it expired at 9 terms, whereas for CTCC it expired at 14 terms. Those who never re-enrolled or who did not re-enroll until after the 9th term of the CCRI cohort or the 14th term of the CTCC cohort were not counted in these skip patterns. The trend lines in Figure 17 show the percent of students who took a short-term skip in enrollment over the course of their tenure at the college. Despite the fact that stop-outs necessarily can only occur in the middle of the term sequence, Figure 17 reveal substantially large differences among racial groups.
The total number of terms skipped also made a difference, as one might expect, in progress toward completion. Figure 18 visually shows that if a student skips more than two terms, the chances of completion are extremely small.
The outcome indicators above focus upon the inverse of completion, specifically on course failures, skips in enrollment, and dropouts. Minority groups are over represented in developmental group enrollments and developmental course failures. This puts them at a disadvantage in completion of a program of courses. Older age groups likewise experience a disadvantage, perhaps because of external commitments.

4. Technology Programs

Using the 2010 CIP codes for programs the advanced technology education (ATE) fields were divided into 4 categories in descending priority:
1. Engineering, electronics and manufacturing
2. Computers, computer networking, and IT
3. Technology and technician fields not already coded above,
4. Science

The codes were applied for each term of each cohort, but only the first term’s program selection was used in some analysis. In CCRI, 8% of the students had chosen an ATE field while in CTCC 6% had done so. In CTCC, the choice of engineering declined slightly over the six cohort years, while the distribution of ATE choices differed substantially across the 12 colleges in the CTCC system. The following chart (Figure 19) shows that the completion rates for engineering and science tended to be higher than for other technologies, and for computer fields in particular.

In both systems, the gender differences are very large in technology programs and can be seen in Figure 20 below where three times as many males as females
continued to chose a technology field term after term. In should be noted that “program” or “major” is not constant for any given student because a students declare majors and then change their declared field. These changes on the part of individual students are incorporated in the trendlines of Figure 20.

An unexpected difference was uncovered in the two systems. In CTCC, attrition from an ATE technology program was about equal over time, whereas in CCRI not only was the attrition from their intended ATE major in term one substantial (almost half), but women were over twice as likely to choose a non-ATE program.

The substantial exit of women from advanced technology programs can be seen visually in Figure 21. Females were more than twice as likely to drop a technology major as were males by their 4th term. By their 9th term in CCRI, females were almost 3 times as likely to drop their technology program as were males. The attrition in CTCC technology programs was greater for women than men, but not as dramatic as at CCRI.
Conclusions

Because of the exploratory nature of the project and complexity of the data, it is not possible to emphasize enough that these results are preliminary and need to be replicated by others with these and other data sets. That said, it is hard not to be intrigued by the patterns uncovered.

In these preliminary analyses, we have confirmed the major role that obstacles and momentum forces, as well as demographic characteristics, play in completion success. The time-series, longitudinal, and student-tracking perspectives that the trendlines provide show that gender, race, and age play a major role in the student pipeline and that the magnitude of this role changes across time.

These term by term data lay the ground work for a microsimulation model because such a model necessarily artificially traces the student career through a program. The indicators of process and outcomes across time will help to provide baselines for variables in “what-if” microsimulation models.

In this effort, we have established that is it feasible to acquire and structure large microdata files of student progress data in order to examine enrollment and completion patterns. The difficulties and challenges encountered will help to refine any future work that is done along these lines.
References


Ewell, Peter & Jenkins, Davis (2008) “Using state student record data to increase community college student success” Pp 71-82 in Bers, Trudy Hl (Editor) Student Tracking in the Community College (New Directions for Community Colleges, No. 143), NYC: John Wiley & Sons.


APPENDIX A. Codebook for Completion Data

Field List:
1. **Student id** can be any values uniquely identifying each student. Preferably it will be a random number. Keep in mind that you may want to add data to these records in 2011 or later years, in which case you will want to be able to use this field in merge more data into this file. It is likely that we will continue this project for several years, enhancing the model and using data from for more than five years.
2. **Institution id** is a code number or character string to uniquely identify. If you do not include one, I will do so, in order to keep your institution separate. It would be best for it to be a short random number.
3. **Gender** 1 for male; 2 for female; Use 0 (zero) or 9 for missing (not available).
4. **Race** 1 for white; 2 for black; 3 for Latin; 4 for Other; 0 or 9 for unknown
5. **Age** This can be either age at the starting quarter, year of birth, or age collapsed into age groups. We may collapse 22 and above as high and below 22 as low, so if you use age groups, it would be best if you used a group that breaks between age 22 and 23.
6. **Living with parents** 1 for yes; 2 for no; 9 or zero for not known
7. **Marital status** 1 for married; 2 for non married; 3 for divorced or separated; 0 or 9 for unknown
8. **Children** (Does student have any children?) 1 for yes; 2 for no; 9 or zero for not known
9. **Highest level of education of either parent** 1 for high school only; 2 for some prior college without degree; 3 for any college degree; 4 for postgraduate study; 9 or 0 for unknown
10. **Highest prior education of student** 1 for high school only; 2 for some prior college without degree; 3 for any college degree; 9 or 0 for unknown
11. **Financial aid or loan received** 1 for yes (any amount); 2 for no; 9 or 0 for unknown
12. **Financial support from parents** 1 for yes (any amount); 2 for no; 9 or 0 for unknown
13. **Concurrent job** 1 for yes (either part time or full time)
14. **Year of term:** 2004-2005 (or another year for first term, e.g., 2006-2007)
15. **Term:** 1 for Fall
16. **Intended or enrolled program type** (e.g., engineering, network technician), if any; Use any coding scheme. Please provide definition of categories of program type. Use 0 in unknown.
17. **Number of courses enrolled in** by student during term; (Do not consider a student “enrolled” if student withdrew early in term, without record on transcript.) Code 1 through 8; use 9 or 0 for unknown.
18. **Number of courses completed** by student during term; Code 1 through 8; use 9 or 0 for unknown. “Completed” is defined as receiving a passing grade.
19. **Number of developmental courses taken** by student during term; (Do not consider a student “enrolled” if student withdrew early in term, without record on transcript.) Code 1 through 8; use 9 or 0 for unknown.
20. **Number of developmental courses completed** by student during term; Code 1 through 8; use 9 or 0 for unknown. “Completed” is defined as receiving a passing grade.
21. **Total credits attempted**; use 2-digit code with zero or 99 for unknown (“Attempted” is same as enrollment. See #17 for definition of enrollment.
22. **Total credits earned**; use 2-digit code with zero or 99 for unknown. “Earned” means completed with passing grade.
23. **Total developmental credits attempted**; use 2-digit code with zero or 99 for unknown (“Attempted” is same as enrollment. See #17 for definition of enrollment.
24. Total developmental credits earned: use 2-digit code with zero or 99 for unknown. “Earned” means completed with passing grade.

25. Outcome status: Code 1 for degree or certificate earned; 2 for dropout; 3 for transfer to 4-year program; 4 for no status change; 0 or 9 for unknown.


27. Term: 2 for Spring.

28. Intended or enrolled program type (e.g., engineering, network technician), if any; Use any coding scheme. Please provide definition of categories of program type. Use 0 in unknown.

29. Number of courses enrolled in by student during term; (Do not consider a student “enrolled” if student withdrew early in term, without record on transcript.) Code 1 through 8; use 9 or 0 for unknown.

30. Number of courses completed by student during term; Code 1 through 8; use 9 or 0 for unknown. “Completed” is defined as receiving a passing grade.

31. Number of developmental courses taken by student during term; (Do not consider a student “enrolled” if student withdrew early in term, without record on transcript.) Code 1 through 8; use 9 or 0 for unknown.

32. Number of developmental courses completed by student during term; Code 1 through 8; use 9 or 0 for unknown.

33. Total credits attempted: use 2-digit code with zero or 99 for unknown (“Attempted” is same as enrollment. See #17 for definition of enrollment.

34. Total credits earned; use 2-digit code with zero or 99 for unknown. “Earned” means completed with passing grade.

35. Total developmental credits attempted; use 2-digit code with zero or 99 for unknown (“Attempted” is same as enrollment. See #17 for definition of enrollment.

36. Total developmental credits earned; use 2-digit code with zero or 99 for unknown.

37. Outcome status: Code 1 for degree or certificate earned; 2 for dropout; 3 for transfer to 4-year program; 4 for no status change; 0 or 9 for unknown.

38. Year of term: 2005-2006 (Note: this summer term may be part of 2006-2007.)

39. Term: 3 for Summer.

40. Intended or enrolled program type (e.g., engineering, network technician), if any; Use any coding scheme. Please provide definition of categories of program type. Use 0 in unknown.

41. Number of courses enrolled in by student during term; (Do not consider a student “enrolled” if student withdrew early in term, without record on transcript.) Code 1 through 8; use 9 or 0 for unknown.

42. Number of courses completed by student during term; Code 1 through 8; use 9 or 0 for unknown. “Completed” is defined as receiving a passing grade.

43. Number of developmental courses taken by student during term; (Do not consider a student “enrolled” if student withdrew early in term, without record on transcript.) Code 1 through 8; use 9 or 0 for unknown.

44. Number of developmental courses completed by student during term; Code 1 through 8; use 9 or 0 for unknown.

45. Total credits attempted: use 2-digit code with zero or 99 for unknown (“Attempted” is same as enrollment. See #17 for definition of enrollment.

46. Total credits earned; use 2-digit code with zero or 99 for unknown. “Earned” means completed with passing grade.

47. Total developmental credits attempted; use 2-digit code with zero or 99 for unknown (“Attempted” is same as enrollment. See #17 for definition of enrollment.
48. **Total developmental credits earned;** use 2-digit code with zero or 99 for unknown. “Earned” means completed with passing grade.

49. **Outcome status** Code 1 for degree or certificate earned; 2 for dropout; 3 for transfer to 4-year program; 4 for no status change; 0 or 9 for unknown

The last 24 fields (fields 14-37, which represent 3 terms of 12 fields each) are for the first academic year. The remaining fields, from 38 to 149, would be structured identically to the first year. The actual assignment of years and terms to field numbers is shown in the table below. Fields for the remaining terms, through spring, 2009, continue in subsets of 12 fields per term, for 193 fields as noted in note #3 above and as outlined in the table below.

<table>
<thead>
<tr>
<th>Year</th>
<th>Academic Year</th>
<th>Term</th>
<th>Fields</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2004-5</td>
<td>1 (Fall)</td>
<td>14-25</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2 (Spring)</td>
<td>26-37</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3 (Summer)</td>
<td>38-49</td>
</tr>
<tr>
<td>2</td>
<td>2005-6</td>
<td>1 (Fall)</td>
<td>50-61</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2 (Spring)</td>
<td>62-73</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3 (Summer)</td>
<td>74-85</td>
</tr>
<tr>
<td>3</td>
<td>2006-7</td>
<td>1 (Fall)</td>
<td>86-97</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2 (Spring)</td>
<td>98-109</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3 (Summer)</td>
<td>110-121</td>
</tr>
<tr>
<td>4</td>
<td>2007-8</td>
<td>1 (Fall)</td>
<td>122-133</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2 (Spring)</td>
<td>134-145</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3 (Summer)</td>
<td>146-157</td>
</tr>
<tr>
<td>5</td>
<td>2008-9</td>
<td>1 (Fall)</td>
<td>158-169</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2 (Spring)</td>
<td>170-181</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3 (Summer)</td>
<td>182-193</td>
</tr>
</tbody>
</table>

*Note: Fields for shaded rows are fully defined in the Field List as fields 14-49.*

If you have only 3 years of data available, for year one enter the academic year 2006-7 first, followed by the next two years. In that case you would have only 12 sets of 12 fields each, rather than 15 sets.

Additional Notes:

1. While this codebook suggests an implicit order, formats, and coding for these data elements, please realize that I understand that the data may not be available in these ways. I can reformat, recode, combine variables, and in other ways create a file from your data to work with on this project. Even if the data are completely missing for some of these items, we can still work with whatever data you send us.

2. Data format: the data elements can be separated by a comma, one or more spaces, a tab character, or they can be in fixed fields. If using fixed fields without spaces between the items, please supply a guide to the number of columns of each field. Comma separated value (CSV) is the preferred format.

3. The initial fields (fields 1-13 or gender to concurrent job) apply to the first (fall) term of 2004; unless you have only 3 years of data, in which case, you would probably start with fall of 2006. After the first 13 fields (data elements) the remaining data are in groups or subsets of 12 fields each, one for each term through spring, 2009. The remaining data items are structured in groups of 12 fields for each of 3 terms for each of 5 years, for 120 fields. These would be the fields from 14 to 133 in the list below, however; only the first
year of fields is listed below. All years and their fields are listed in the table at the end of this document. I would like five years as that would next year give us a 6-year period, which is often used for measures of program completion. If you can only go back to 2006, that is OK. We would like to code three terms for each academic year as follows: 1 for Fall Term; 2 for spring term; and 3 for Summer Term.

(4) For each item, codes of 0, 9, or 99 have been proposed for information not known or missing. Blanks are OK but only if you are NOT using blank to separate data fields.

(5) You may send the file as an SPSS file, an Excel file, or as a CSV (comma separated value) file. SPSS will be used for data management, analysis, and for input to the microsimulation model.

(6) Note that the CT database included an additional field for a 3 letter code: RTN for returning student after a prior term; NEW for first term at that school; HSS for high school student. This variable was called “status” and was used for more precisely defining the cohorts as those students coded “NEW” at their first term.
To Whom It May Concern:

This is to seek your participation in the microsimulation modeling project described in the synopsis appended. With funding from the National Science Foundation’s Advanced Technological Education (ATE) program, we are developing analysis strategies and a demonstration microsimulation model of student retention in community college programs, with a specific emphasis upon those programs that prepare students for jobs in advanced technology fields.

We are seeking participation from several community colleges so that the microsimulation can work with real world student data, making the comparisons and other results more useful. Your participation would only require preparation of data from students’ electronic records across a minimum of three years beginning with the 2006-07 year. (Connecticut is providing five years beginning in 2004-5.) In return, we will analyze the data with statistical techniques and a demonstration microsimulation model that are designed to achieve the following objectives: 1) tease out obstacles for specific student subgroups that inhibit their progress, thus potentially providing you a basis for improving student success rates; 2) improve projections of student enrollment and different types of departure at different points in time; and 3) provide you with a tool for asking “what if” questions, e.g., what if the number of enrollments of Hispanic males over 20 doubled, how would the overall student completion speed and success rates change?

As a token of appreciation for participation in this project, a $500 honorarium will be paid to the Institutional Research Director or the person designated as the point person.

This is just the first phase of project to build a powerful analytic and policy refinement tool for community colleges and individual programs within them. It is our intent to seek national funding for the model in order to add the following components: recruitment strategies, financial aid, economic trends, remedial (developmental) programs, life course transitions, and changes in workforce requirements, and job placements. In other words, our goal is a complete model of the student pipeline system. By participating with us now, you will have the benefit of
being able to help shape the model so that it includes solutions to strategic challenges that you face in your institution.

A decision to participation in the project is needed by December 15, because on that date we will finalized the definition of the data elements and the formats required so that the data will be comparable. We will attempt to accommodate your data constraints and make it as little work as possible for you to supply the needed data.

I want to stress that not only will the students be anonymous, but also each institution will be assigned a random identifier. Only I, and your institutional representative, will know the identity of your institution’s ID code.

I have worked with microsimulation modeling of prison and sentencing populations for many years. Working with me is Martin Spielauer, PhD, who works in microsimulation modeling full time at Statistics Canada. Some of the models he works with include education as part of life course processes. While the community college model will be a pioneering application, we have the experience to ensure its completion and usefulness.

It is our goal to have the microsimulation working by February, 2010. We would like to hold an optional informational meeting of participants at the AERA meetings in Denver, May 1-4, 2010 or at a location in New England before May to demonstrate the model and how community college IR directors and other researchers can use it.

It is my firm belief that your effort in preparing the data for use in the microsimulation model will be well rewarded in terms of strategic analysis the retention of your students. Please call me at 952-473-5910 or 612-963-6660 (travel phone) if you have any questions or send me an email at rea@umn.edu.

Sincerely yours,

Ron Anderson
Synopsis of Microsimulation Project Utilizing Student Record Data to Study Student Retention

May 25, 2010

Ron Anderson, Professor Emeritus, University of Minnesota (rea@umn.edu) (Feel free to call me at 952-473-5910 or 612-963-6660 (mobile).)

Recent, volatile trends in the supply of and demand for community college graduates, especially from advanced technology programs, argues for the need to improve analysis and forecasting of student flow through the academic system. This flow includes recruitment, remediation, retention, and placement.

We are seeking participation with community colleges that collect data of the type needed for longitudinal statistical analysis and microsimulation modeling.

The desired data are listed in the appendix below. Within the list are two groups: the first group contains fixed data elements from a cohort of students at their entry into the college; the second group consists of items needed at the end of each term for three or more years. With these data, it will be possible to map the flow of students through academic program(s), identifying obstruction points for various demographic groups, especially gender, age, and ethnicity/race.

We don't need to identify any individuals, and we don't need to identify institutions. The time frame for the initial, demonstration model is to have it running by the fall term of 2010. During this initial phase, we will do baseline testing to establish model validity. Depending upon funding, the extensions to the basic model would occur after that.

During the past six months we have obtained such data from all the community colleges in Connecticut and Rhode Island. It appears that we will be able to accommodate all Massachusetts community colleges this summer or the upcoming academic year. If you are interested in participating, please let us know as soon as possible.

The proposed participation will give institutions new tools (new data structures, statistics and microsimulation models) to utilize in their student data analysis for strategic planning. The model will allow for the exploration of “what if” analysis to aid in anticipating future scenarios. Such planning is important because of the rapidly evolving supply and demand for technicians and others graduating from community college programs.

Appendix: Desired Data Elements for Community College Pipeline Modeling

1. Data Elements for starting cohort (Fall term, 2006-07)
   Fixed unique student ID (randomly assigned)
   Fixed demographic background factors:
   Gender;
   Race/Ethnicity;
   Age;
   Family and finance:
   living with parents; (if available)
   marital status; (if available)
   dependent children; (if available)
   financial aid or loan received; (if available)
   concurrent job; (if available)
   Financial aid from parents; (if available)
   Highest level of prior educ.; (if available)

2. Data Elements Needed each term, 2006 to 2010
   (Student data needed at end of each term)
   Initial Term (e.g., Fall 2006)
   Intend or enrolled program (e.g., engineering, network technician), if any;
   Number of courses attempted;
   Number of courses completed;
   Total credits attempted;
   Total credits earned;
   Outcome status:
   Degree or Cert. earned; (if available)
   Transfer to 4-year prog.; (if available)