DOCUMENT RESUME

| | |
|---|---|
| ED 414 972 | JC 980 044 |
| AUTHOR | Boughan, Karl |
| TITLE | Toward a Model of the Academic Progress: Summarizing Cohort 1990 Progress and Achievement. Enrollment Analysis EA97-6. |
| INSTITUTION | Prince George's Community Coll., Largo, MD. Office of Institutional Research and Analysis. |
| PUB DATE | 1997-03-00 |
| NOTE | 63p. |
| PUB TYPE | Reports - Evaluative (142) |
| EDRS PRICE | MF01/PC03 Plus Postage. |
| DESCRIPTORS | *Academic Achievement; Academic Persistence; Career Guidance; Cluster Analysis; *Cohort Analysis; *College Outcomes Assessment; Community Colleges; Data Analysis; Longitudinal Studies; Models; Path Analysis; School Holding Power; *Student Characteristics; Student Motivation; *Two Year College Students; Two Year Colleges |
| IDENTIFIERS | Prince Georges Community College MD |


ABSTRACT

To track student progress and develop a model of the total academic process at Prince George's Community College (PGCC), the Office of Institutional Research and Analysis acquired 5 years of background, behavioral, and outcome data for a cohort of first-time students entering the college. The academic process was broken down into three divisions: input (environmental variables measuring pre-college student attributes and college-external personal circumstances), through-put (process variables measuring student progress), and output (outcome variables measuring fulfillment of educational objectives). Exploratory modeling was used to create a structured academic process. The research first used regression analysis and created 10 new academic process factor scales, plus 7 student background indicators. Next, a model was created to explain student outcomes through the application of causal path analysis. The last phase of research was to produce a second model complementary to the path model but less abstract. Cluster analysis was applied to the cohort data, resulting in 10 clusters describing student attributes and behaviors. Though the path and cluster models were highly tentative, valuable insights were offered, including the importance of personal motivation, taking student career differences seriously, and creating academic support programs. Appendices include cluster data tables and methodology of social background variable construction. (YKH)

# Tracking Student Progress at PGCC

## Toward a Model of the Academic Process: Summarizing Cohort 1990 Progress and Achievement



PRINCE GEORGE'S
COMMUNITY COLLEGE



Office of Institutional Research and Analysis
Enrollment Analysis EA97-6
March 1997

PRINCE GEORGE'S COMMUNITY COLLEGE
Office of Institutional Research and Analysis

# TRACKING STUDENT PROGRESS AT P.G.C.C.: TOWARD A MODEL OF THE ACADEMIC PROCESS – SUMMARIZING COHORT 1990 PROGRESS AND ACHIEVEMENT

Enrollment Analysis EA97-6
March 1997

## *Introduction*

This is the fourth and final academic outcomes report[1] based on the Office of Institutional Research and Analysis's tracking of a cohort of first-time freshmen (N = 2,643) entering in the Fall of 1990.[2] While the earlier papers focused on specific aspects of Cohort 1990's social attributes and educational achievements, in this summary report we will attempt to move *toward* a general model of student academic progress at PGCC. The adverb in the last sentence is italicized because we must stress that the general depiction of the college's academic process arrived at in the research presented in this paper does not pretend to be definitive in any sense.

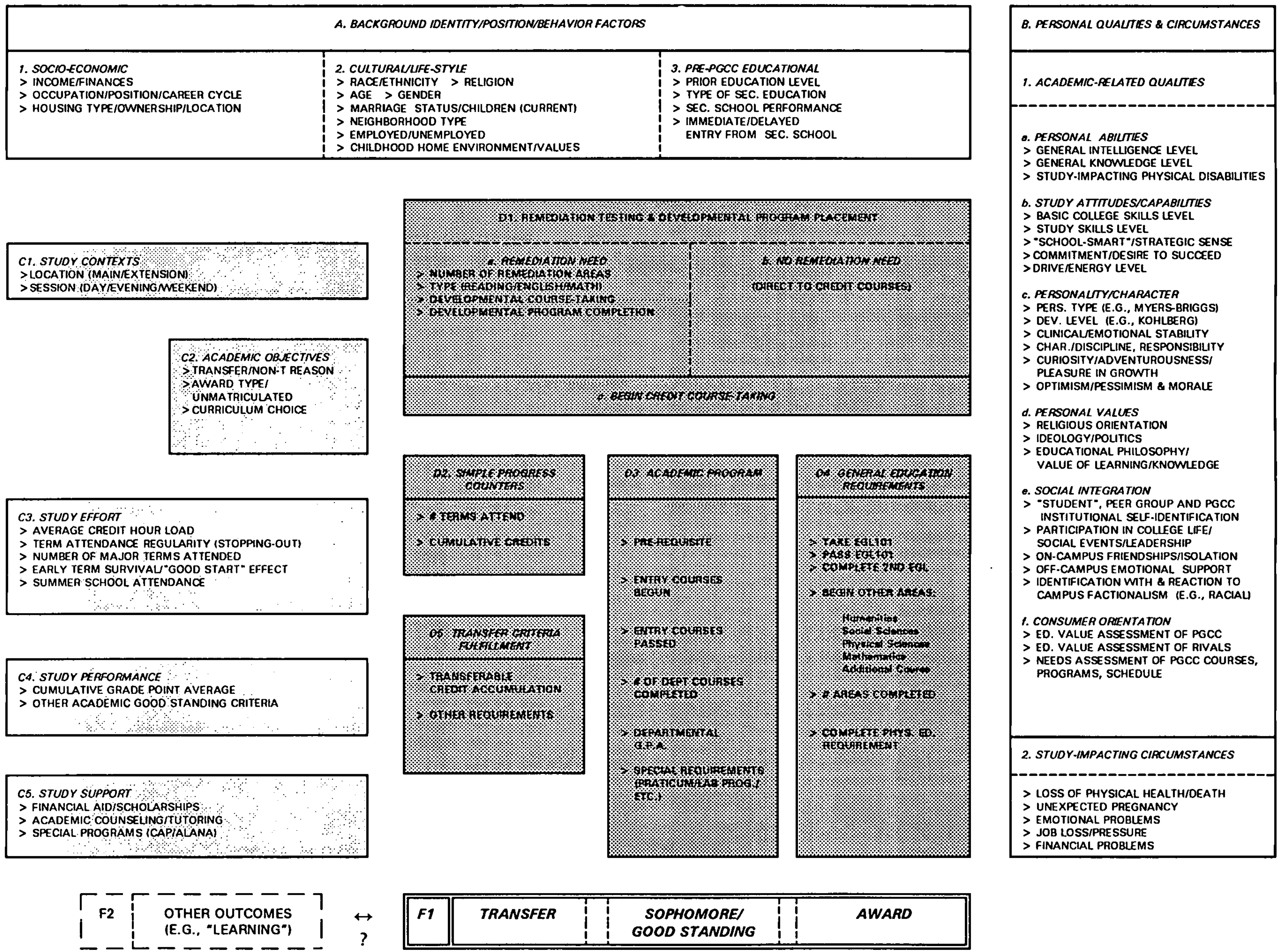
This tentativeness is partly the result of limitations in the analytic techniques we employed, which will be discussed in the body of the report. But more fundamentally, it reflects our sense of the great scope and complexity of the phenomenon being modeled and how inadequate the currently available data are in doing justice to that scope and complexity. To make this point clearer, we have inventoried to the best of our understanding all of the conceptual elements which need to be considered in groping toward any comprehensive theory of the academic process as it presently plays out in U.S. community colleges. Figure 1, below, shows these theory elements arranged by type within a structure very broadly suggesting a causal patterning leading up to final academic outcome. Causal path arrows, however, have deliberately *not* been drawn to emphasize the unfinished nature of the figure. It is not itself a portrayal of a comprehensive model of the academic process but rather a concept-gathering paradigm

---

[1]See *Tracking Student Progress at P.G.C.C.: Basic Findings of the 1990 Entering Cohort Academic Outcomes Analysis* (Enrollment Analysis EA95-7, June 1995), *Tracking Student Progress at P.G.C.C.:Fall 1990 Entering Cohort Four-Year Patterns of Attendance and Timing of Outcomes* (Enrollment Analysis EA96-1, July 1995), and *Tracking Student Progress at P.G.C.C.: Student Racial Background and Cohort 1990 Four-year Academic Outcomes* (Enrollment Analysis EA96-6, June 1996).

[2]The Cohort 1990 data set is drawn from PGCC student record databases, augmented with material supplied by the Maryland Higher Education Commission's Transfer Student System to enable us to identify cohort members who ceased community college attendance due to transfer to a Maryland four-year public post-secondary institution. Attendance, study progress and related data are all organized on a term-by-term basis so that we may assess student academic status and level of achievement at any point in the four-year process, connect patterns of attendance with outcomes, and summarize any part of the process in terms of time to outcome.

Figure 1. Toward a General Model of the PGCC Academic Process and Correlates of Academic Success




BEST COPY AVAILABLE

for guiding research out of which a highly specified, fully determined academic process model may ultimately emerge.

Block A in Figure 1 represents the most "exogamous," causally-prior factors impacting on academic outcomes: the classic socio-economic, cultural, life-style and pre-college educational "background" variables. Student attributes regarding these are not only chronologically prior but also in an important sense fixed. They are biological givens or very slowly changing social environmental statuses which neither students nor the institution can alter much or control, but which may continue, in different ways at different stages, to influence academic progress.

Block B lists another set of "para-process" variables, those best conceptualized either as decidedly individual or personal characteristics of students varying only loosely with social background attributes if at all (e.g., extroversion/introversion), or attributes relating to the college as a social system rather than educational institution (e.g., college activities participation/non-participation). Block B variables include: personal abilities (e.g., native intelligence, physical disability), college study-related attitudes (e.g., motivation/commitment) and capabilities (e.g., study habits and skills), personality and emotional factors (e.g., Myers-Briggs Type, depressive tendencies), personal values (e.g., religious, political, educational), social relationships (e.g., Tinto's level of college community integration, level of familial support for studies), consumer orientation (satisfaction with college courses, services, tuition), and special unforeseen circumstances (e.g., loss of health, job relocation). Block B student characteristics are similar to Block A attributes in that they represent factors conceptually outside the formal educational process though highly relevant to it, and in that they tend to form prior to college entry; but they differ in that B-factors can alter, sometimes dramatically, under the impact of the college experience.

The C (light-shaded) and D (dark-shaded) variable sets enumerated in Figure 1 stand for the academic process proper. The former consist of student academic behavioral and performance indicators (location of study, schedule type, study objective, major curriculum of study, credit load tendency, grade point average, participation in financial aid and academic enhancement programs, and the like). Taken together, these factors constitute the *action* side of the academic process: students making study choices, enrolling in courses, doing assignments, taking examinations, getting grades, and thereby increasing or decreasing their chance of educational success according to the quality of their decisions and performance. It is important to notice, however, that none of these variables directly measures academic *progress*. For example, knowing that a student has a high cumulative GPA tells one little about how far along she has traveled toward her educational goal; many a high-performing student has been forced to leave college for financial or health reasons, or fails to fulfill core requirements because of difficulty with a required course or two. Good academic performance may contribute to or predict academic progress but is not the progress itself.

It is the D group of factors that make up this progress measuring or *effectual* side of the academic process. D1 represents the remediation sub-process, whether a student has placed into one or more academic skills developmental course sequences (English usage, reading comprehension and high school mathematics), the number and scores of placement tests taken, the taking and passing of developmental courses, and the number of required developmental programs completed. PGCC students required to take developmental courses and who have not yet completed their remedial programs are precluded from enrolling in many entry-level core requirement credit courses. "Graduating" from one's developmental programs is tantamount to progressing into the regular credit course stream. The remaining D factors represent progress within the regular credit stream: D2 portrays academic progress as measured by simple summary indicators of retention/persistence and undifferentiated credit accumulation. D3 lists stages in the fulfillment of a curriculum program leading to an associate degree or occupational certificate. D4 presents the graded steps toward fulfilling a degree-seeking student's general education requirements. Finally D5 includes the criteria set by a targeted four-year school met by a student attempting to transfer out of PGCC.

The complexity of the D sector of the academic process as pictured in Figure 1 implies that academic progress should be understood as a multidimensional rather than a unitary concept. Here, assessment of a student's overall level of relative achievement requires multiple scores on multiple progress indices, and there is no stipulation that student progress need be even down all paths. This brings up a second way of looking at the D sectors — as the *structural* or *formal* core of the academic process, the mechanics of course enrollment prerequisite, standards for satisfactory academic standing, credit accumulation procedures, and program satisfaction requirements.

From this standpoint, the academic process is viewed as an articulation of several interlocking sub-processes — the main ones being basic skills remediation, academic standing maintenance, satisfaction of general education requirements, and fulfillment of curriculum program requirements — each of which constitutes a distinct set of challenges for students to overcome. Given the articulated complexity of the system, many students who possess in some general sense sufficient academic talent and motivation to succeed, may find their overall progress temporarily balked or even permanently derailed by difficulty in negotiating some particular sub-process. For example, some may already have satisfied their general education requirements but may stall out in the gateway courses of their study major; some may have easily jumped the hurdles presented by all but one required general education area — the physical sciences perhaps; and some who might otherwise accumulate significant credit totals may not get the chance to enroll in regular courses because they are unable to complete their developmental requirements. We term this the *Lost in the Machinery* effect, and future research may prove that it accounts for a fair amount of enrollment attrition, quite apart from that explained by factors like general course performance and average study load considered as discrete independent variables.

The final sector, the two F blocks shown at the bottom of Figure 1, represents system output — the educational results of the academic process at PGCC. F1 (double-line border) stands for conventional overall academic success or achievement. OIRA uses a set of standard criteria to judge whether the final outcome of a student's PGCC study can be classified as a conventional success: A student classifies as a success at the time of outcome assessment if his study has resulted in either (1) formal academic achievement — earning any award (associate degree, occupational certificate or letter-of-recognition) and/or transferring to a four-year higher educational institution; and/or (2) significant credit accumulation, set as sophomore status (30 or more credit hours earned) plus academic good standing. The main independent variable of final outcome employed in this study (achiever = 1/non-achiever = 0) is based on sorting Cohort 1990 students according to the above criteria after the fifth year of the cohort's existence.[3]

F2 (dashed-line border) stands for all other ways of adjudging educational results, and is included in Figure 1 mostly to remind us that achieving conventional academic success is not equivalent to benefiting from a higher educational program in many important senses. For example, conventional achievement at best is no more than a loose and indirect indicator of *actual learning* even if we limit the test of knowledge acquisition to a student's field of study, and even less so if it is broadened to include knowledge constituting a classic liberal or new information age education. Nor does conventional academic achievement any better give us a handle on the fulfillment of important non-educational objectives for attending college: using the knowledge and credentials acquired to get a first job, change careers, start a business, improve performance in one's current employment, or achieve promotion or salary enhancement.[4]

---

[3] The second issue should be brought up at this point, having to do with a modification of the cohort base for this research. Since our purpose here was to gain an understanding of the correlates of academic achievement among regular program students, obvious special motive students were dropped from the cohorts. These were students the pattern of whose first-registration declarations and initial enrollment behavior strongly suggested non-program reasons for attendance. Early analysis identified 257 students fitting this description (around 10 percent of the original 2,643 cohort members), leaving an analysis base of 2,386 students.

[4] Paradoxically in fact, for one not insignificant set of students it is actually *failure* to achieve in the conventional sense which may mark success: "hidden" special motive attenders who start regular award programs and state degree-seeking study objectives to the Registrar but have no real intention to complete their programs and drop out immediately upon satisfying their true short-term job-related goals. Later in this report we will present evidence for the existence and extent of hidden special motive students in Cohort 1990.

This brief review of outcomes measurement immediately suggests an important reason why most outcomes research, including ours, relies on formal achievement dependent variables — simple data availability. The PGCC student records system and the Maryland Higher Education Commission's student tracking system readily provide individual enrollee data on credit accumulation, academic standing, award earning and transfer to four year colleges and universities, whereas measuring informal, more difficult to specify concepts like student "learning" and extra-academic success would require special time- and resource-intensive data gathering efforts, probably involving elaborate survey research.[5] This leads to a larger point: incomplete identification of all of the vital model elements aside, it would be premature to posit any general model of the academic process if only because many of the likely elements are so difficult or impossible to operationalize by means of standard data. And if anything, this is even more true on the independent variable side of the equation than it is on the dependent variable side. Consider the following *partial* list of serious gaps in our Cohort 1990 database: In the social background area (A), although we have the usual bare-bones attribute indicators (student age, race/ethnicity, gender, marriage status), we have no data on current employment status, occupational category data, personal income,[6] or financial circumstances. Nor do we have measures of current family circumstances like whether students are living with, and are being supported by, parents nor concerning the number and age of children students themselves maybe responsible for. Nor is there any information on family of upbringing (e.g., parental education and values), nor on secondary school

[5] Even reliable standard sources on formal academic achievement sometimes turn out to be lacking in some methodological respect or other. In PGCC's case, the classification accuracy of the transfer component leaves something to be desired. The Maryland Higher Education Commission, sole source for the data allowing us to flag PGCC student transfers, limits its tracking to community college transferrals to state system four-year colleges and universities. The absence of data for transfer to independent and out-of-state schools means that the probable full extent of Cohort 1990's transfer behavior goes under measured in our research. An outcomes assessment of Cohort 1984, which had the unusual feature of including exiter *survey* data, suggests that MHEC's tracking approach misses perhaps a quarter of all valid four-year transfers actually achieved by PGCC first-time students. Therefore, by some small but not unimportant degree our report will be misrepresenting the true achievement-explanatory strength of the our independent variables. Inevitably the real correlations will be undercut by the inadvertent missorting of some number of genuine transfer achievers into the non-achievement category of the dependent variable.

[6] Fortunately, we have been able to create a number of surrogate indicators of income and social class by means of *PGTRAK*[90], a marketing system developed for the college's targeted student recruitment programs developed from U.S. Census tract data . All socio-economic attribute variables used in the this research are based on student home tract. Estimating individual attributes from a group central tendency is, of course, prone to result in considerable predictive error. The one time we had data to test this (of Fall 1996 Pell Grant applicants which included both *PGTRAK*[90] income estimates and Financial Aid Office personal income category flags), we found that the tract-based estimate was fairly predictive of known personal income, but only at the grossest data cut — 70 percent agreement in estimate-to-actual income classification dichotomizing both indicators at below/above $40,000 household income.

subjects taken nor overall high school performance. The data gap is even wider when it comes to student personal attitude and capability (Area B). The college's student records system has never been charged to collect data of this kind on any regular, systematic basis, and thus we have virtually no direct knowledge of student native intelligence, motivational levels and quality of study habits, personality and value system types, general physical and emotional health or specific learning disabilities, or degree of social integration into campus life.

But as serious as these data gaps are, we feel that it is important *not* to postpone earnest attempts to model the academic process to that happy day when student records reflect all of the forces which may bear significantly on the progress of college study. First of all, realistically that day may never come, and it would be wrong not to press the data at hand, as inadequate in coverage of the phenomenon to be explained as they may be, for whatever useful they may be able to tell us. It is, after all, possible that standard variables might take us quite far in rendering an account of the correlates of academic achievement.

Secondly and more importantly, early modeling efforts will point the way to the more refined, predictive models of the future. Such work should enable us to sort out which standard independent variable domains seem to be truly important and which appear to make only marginal contributions to academic progress variance, to begin to trace the overall causal network of forces leading up to final study outcome, to gather hints as to which variable domains not covered by standard student records data seem most likely worth the arduous effort of special data development, and to make progress solving important technical issues of methodology.

In the remainder of this report we will report the results of an effort to produce a provisional model of the progress toward academic achievement of first-time credit students entering PGCC in the fall of 1990. With a few exceptions, all of the data employed in this research was of the standard student records type.

***Research Plan and Methodological Considerations***

In this study, we used three different multivariate modeling techniques. First, for a preliminary exploration of data set we chose *linear regression analysis*. Here the objective was to gain an early sense of how adequately the set of social background and academic process variables available to us covered the phenomenon of academic achievement at PGCC, to determine the form of causal model ultimately required (independent effects or causal network), and to identify and deal with any technical problems within the data.

Regression is ideal for tackling such preliminary issues. The main product of regression is an equation predicting the case values of $Y$ (the *dependent* variable) by adding together the known case values of a series of indicators (the *independent* variables) after each has been multiplied by a *regression weight* representing that variable's unique cross-case impact on $Y$. The overall predictive success of this equation is indicated by the $R^2$ correlation statistic, which efficiently summarizes the *collective* impact of the independent variables upon the dependent variable in terms of the proportion of the total behavior of $Y$ (called *variance*) jointly accounted for (*explained*) by all predictors. In effect, $R^2$ measures the adequacy of a data set as the basis of a statistical model.

The unique impact of each independent variable upon $Y$ is given by a correlation statistic associated with its regression weight. This is $r_p$, the semi-partial correlation, which when squared yields an estimate of how much of the total variance of $Y$ is traceable to the effect of that variable and only that variable. A comparison of the sum of the squared semi-partial correlations with $R^2$ answers the question of which model type is best to use with the data at hand. If the squared semi-partial sum approaches the value of $R^2$, then the best choice is an independent effects model. Since the regression equation is, in fact, simply an independent effects model expressed in mathematical notation, modeling can stop at this point. But if the sum is small in comparison to $R^2$, then a causal network approach is more appropriate and the modeling effort must shift to a new methodological base.

Finally, regression provides an opportunity for exploring various technical problems of model conceptualization. One perennial concern of statistical modelers is making sure that there exists a true conceptual differentiation between the predictor and predicted variables. For example, study load is clearly a variable conceptually separate from achiever classification, and respectively assigning them predictor and predicted roles in a model of academic achievement is non-problematic. But the conceptual independence of a measure like credit hour accumulation is not so clear, for gathering sufficient college credits is a component part of our definition of academic achievement. Since it is only one component and since there are ways that students with relatively low credits earned records can be accorded achiever classification (i.e., transfer to a four-year school), one might argue that credit accumulation is sufficiently independent to be included among the predictors; or, one might argue that its conceptual overlap with the predicted variable is cause for immediate dismissal.

We decided to take an empirical rather than an *a priori* stand and let a comparison of the $R^2$s of regression equations including and excluding the conceptually ambiguous independent variables settle the matter: if the inclusion $R^2$ is very high, especially in comparison to the exclusion $R^2$, then the ambiguous variables should be dropped because they overlap with the dependent variable enough to be considered statistically equivalent to it. But if the two $R^2$s are near in value, then the overlap is not severe and the ambiguous should remain.

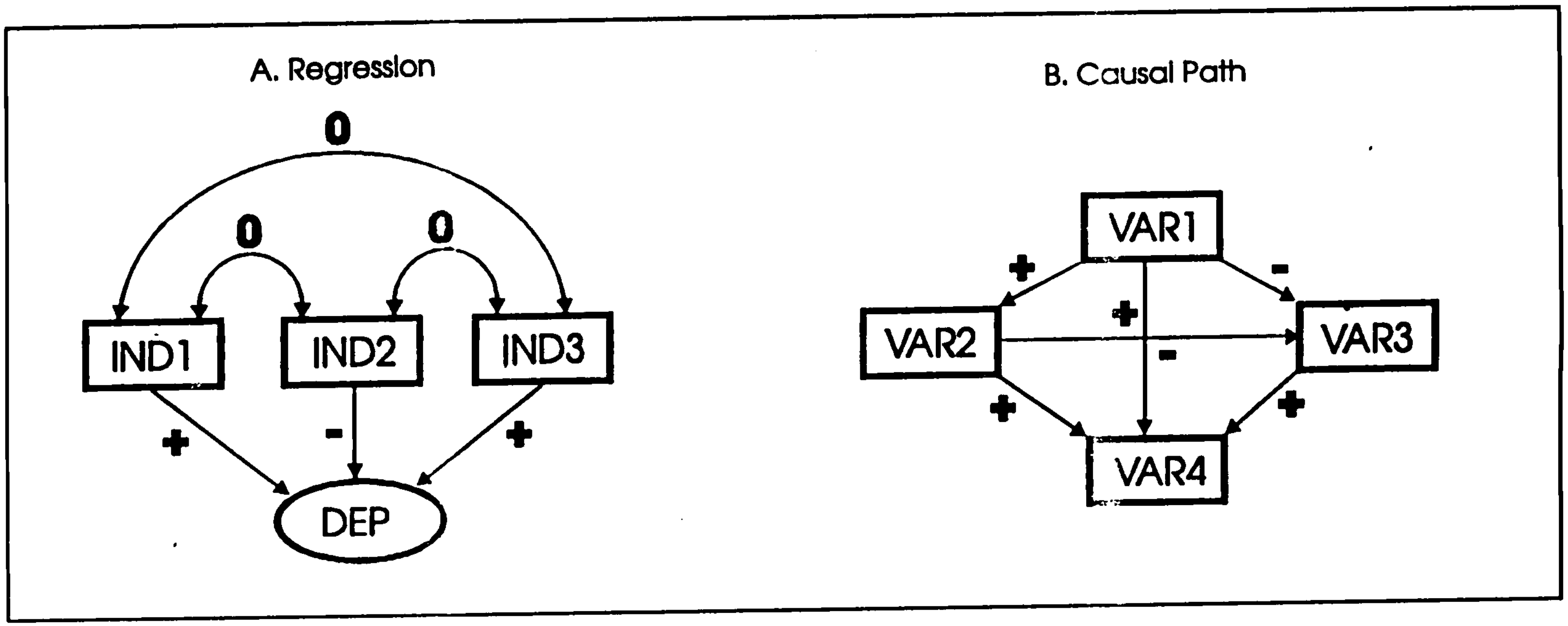
The other important conceptual concern of statistical modelers is the avoidance of high levels of what is known as *collinearity* among predictor variables. Collinearity is the tendency of conceptually related *independent variables* to intercorrelate highly, suggesting the existence of a common underlying dimension. High collinearity can cause all sorts of confounding analytical problems since each of the related variables measures the common dimension slightly differently and at a slightly different level of correlation. When collinearity levels exceed a certain point, it becomes prudent to consolidate the raw data by *data reduction* procedures which produce a smaller set of new scalar variables representing the underlying dimensions of the data. This not only solves the analytic problem, but creates a smaller revised set of predictors that more clearly conceptualize and efficiently model causality. Regression analysis provides a battery of statistics to assess level of collinearity and to identify highly intercorrelating groups of variables.

To summarize, our plan for the regression phase called for three research steps: (1) Run a test regression using all 90 variables in the raw data set to assess total achievement variance explained, hunt for possible independent-dependent variable overlap, and to gauge the level of collinearity among predictors; (2) Drop any independent variables discovered to be seriously compromised by overlap with the dependent variable, and carry out a program of data reduction on the remainder, if advisable given collinearity levels; (3) Run further regressions, using the reduced scalar data set if created, to informally explore causal patterns and to assess the best technique to be employed in the modeling phase of the research.

The objective of the second phase of the research was to build on the work of the exploratory phase by creating a full, if provisional, causal model. If phase one regression analysis found that our data could be modeled quite adequately along independent effects lines, then this would involve only some additional elaboration of the final phase one regression equation. But should this prove not to be the case, as we fully expected it would not be, then we would shift our methodology to a modeling technique known a *causal path analysis*. This expectation was based on our understanding of the real-world complexity of the academic process, which we felt was far more likely to be properly captured in a causal path model than a regression model. Figure 2 (below) graphically contrasts these two modeling approaches.

Model A represents the regression approach which assumes a phenomenon which can be grasped by means of simple, linear, additive mathematics. It has the following structure: a number of independent variables (boxes), a dependent variable (the oval), and unidirectional causal flow (arrows) connecting each independent with the dependent variable. (The mathematical plus and minus signs here, indicating the direction of correlation, are arbitrary, and for illustrative purposes only.) Also drawn are double-headed arched arrows linking all possible pairs of independent variables.

Figure 2. Two Types of Analytical Models



These are assigned zeroes, to indicate absence of correlation. Like spokes connected to a wheel hub, a regression model, then, posits a series of entirely *uncorrelated* predictors, each of which independently impacts (hence, *independent* variable) on a dependent "target" variable. Put another way, the effect of each independent is a *direct effect*, owing nothing to the operation of the other independent variables. From the standpoint of modeling the academic process, the assumption that causal variables be uncorrelated is, of course, totally unrealistic. In the real world GPA can be expected to correlate with Academic Standing which in turn is known to influence Term Credit Hour Load.

The academic process comes much closer to being realistically depicted by the B Model, to which the mathematics of causal path analysis conforms. In the B instance, model shape varies greatly. Variable connections are not forced to comply with a rigid spoked wheel image but link up according to an a priori theory suited to the phenomenon under study; the causal flow (arrow heads) are determined by the conceptual structure of the theory and by common sense and chronicity (e.g., race may effect developmental placement but not the reverse). Model B variables, therefore, tend to interact in a variety of causal patterns and there is no intrinsic distinction between independent and dependent variables. The researcher, for example, might choose as the primary research focus VAR4, which in the figure parallels Model A's DEP and is also the only variable in its model exclusively on the causal receiving end. (This would correspond to achiever classification in an academic process model.) But he is also at liberty to shift attention to other interesting variables to understand their roles in the overall matrix of causality leading to VAR4.[7]

Model B resembles a net rather than a spoked wheel, each cross-knot a variable which may be looked at as the product of the effects of variables preceding it, or as the proximate cause of the next variable down the line, or as the remote cause, along with other variables following it, of some other variable far down the line. To illustrate using the language of variable effects, VAR2 may impact on VAR4 in Model B three distinct ways: *indirectly* through intervening VAR3, *indirectly* as part of preceding VAR1's impact on VAR4, and *directly* (what is left over when the one-on-one correlation between VAR2 and VAR4 is controlled for the effects of VAR1 and VAR3). Using causal path analysis, we would be able without difficulty to substitute for VAR1, VAR2 and VAR3 the three academic process variables from our earlier example, and go on to disentangle all of the direct and indirect effects of GPA, Academic Standing and Term Credit Hour Load among each other and upon achievement (VAR4).

---

[7]Two of these actually take the "dependent" position with respect to some other variables in the total system: VAR4 is "dependent' with respect to VAR1, VAR2 and VAR3; VAR3 vis-a'-vis VAR1 and VAR2; and VAR2 *vis-a'-vis* VAR1. VAR1 is the only true "independent" variable (*exogenous* in the language of path analysis) in Model B.

Our plan for phase two was first to specify a B-like model of study success according to our best understanding of the likely causal flow among social/educational background variables and academic process variables, and then to test it using Cohort 1990 data as revised in phase one and AMOS, a path analysis software package designed to draw causal path diagrams and calculate the empirical strength of inter-variable links. The strength of any linkage is traditionally measured in path analysis by a measure known as the *path coefficient*, which is analogous to the direct effect correlation between two variables once the effects of all prior variables leading to them in the model are controlled for.[8] (In Model B, for example, the path coefficient of the VAR2 > VAR4 link would indicate the strength of impact of VAR2 upon VAR4, controlling for the effects of VAR1 upon VAR2, VAR1 upon VAR4, VAR2 upon VAR3, and VAR3 upon VAR4.) Lastly in this phase, we would examine the pattern of the path coefficients in our model to identify the main pathways to achievement and the most strategically placed variables along those pathways.

The objective of the concluding phase of the research was a little different. Up to that point the research centered on developing a general causal model of the PGCC academic process leading up to study success or non-success. The variables and pathways of the model would indicate just how achievement probabilities altered for the *hypothetical typical student* located at any particular point in the causal matrix. As important and valid as modeling the process in this theoretical way was, it provided a view of limited practical utility. The "typical student" was after all just a methodological device, more myth than real. Real students vary greatly and follow any number of different paths to the final outcome of their study careers. At least as a supplement, academic policy makers needed a more concrete take on how the academic process worked, one centering on the pattern of variation in study career paths . Therefore, we determined to round off our research by re-working the same data which generated the causal path model of phase two using a radically different analytic technique — *cluster analysis*. In cluster analysis, a population is broken down into a number of sub-populations called *clusters*, according to how similar or dissimilar members are to each other across a series of attributes. The set of clusters is formed by the systematic comparison of the attribute profile of each case with all other cases. Based on this, the procedure sorts cases according to the patterns of attribute similarity/dissimilarity into a set of clusters, each exhibiting a maximum internal homogeneity, while the whole set of clusters exhibits a maximum heterogeneity.

---

[8] Technically, a path coefficient is the direct effect standardized regression weight of the causal variable upon the caused variable.

The plan in phase three was to re-analyze the academic process attributes of Cohort 1990 members by cluster analysis, thereby placing students into *study career* clusters, each of which represented a different stable pattern of remediation, study load, term attendance, course performance, *etc*. In effect, then, these study career clusters would represent the set of actual pathways through the academic process most commonly trod by our students, and could be analyzed by member social/educational background and assessed in terms of academic achievement likelihood.

### *Findings from Regression Analysis*

As a set-up for our substantive regression analysis, we began phase one of our research with a "omnibus" test achievement regression, entering all of the original 90 independent variables available to us, the results of which are easily summarizable:

- When all 90 independent variables were force-entered, the regression equation produced had an incredibly high associated $R^2$ of .85 (nearly nine-tenths of all achiever placement variance explained). This strongly suggested the existence of an independent variable/dependent variable overlap problem. We hypothesized that independent variables conceptually close to the dependent achiever variable (e.g., cumulative earned credits hours, number of required general education areas completed) were responsible for $R^2$ inflation, that we were inadvertently guilty, in effect, of partly correlating the dependent variable with itself. Sure enough, when the 18 progress variables were dropped in a second run entering only the remaining background and performance variables, $R^2$ also dropped dramatically — to .63. The decision was made, therefore, to drop these variables permanently from the analysis.

- The finding that an achievement regression involving only truly independent variables yielded an $R^2$ of .63 gave us an estimate of the maximum power of the data available to us to explain achievement variance: around three-fifths. How one evaluates the adequacy of the data for supporting academic process modeling depends on one's point of view, however. From the comparable research perspective, a "more than half-full" result looks quite good, for published works on student academic achievement rarely report variance explanations of over 30 percent. Our comprehensive approach doubled the technical explanatory power of multivariate analysis in this area. However, from the standpoint of achieving a complete theoretical understanding of the phenomenon under study, our research result had an "almost half-empty" appearance. This did not surprise us, given that so many of the components of total academic process (as inventoried in Figure 1) were left uncovered or undercovered by the data available,[9] but it did caution us against making untenable claims for the definitiveness of any product of our modeling efforts.

- The battery of diagnostic statistics provided in the output for the 72-variable regression pointed to the existence of severe collinearity problems in the data. For example, conditional index tests found *that every single academic process variable* seriously intercorrelated with at least two others. Clearly, a comprehensive program of data reduction was called for before we could proceed to the next step.

A *data reduction* procedure takes highly intercorrelating sets of individual variables and re-works them into scalar equivalents measuring the single dimensions represented by each set. When the need for down-scaling is general, as it was here, the technique normally employed is called *factor analysis*. The technique involves, first, the systematic discernment of the patterns found in a matrix representing the correlation of each variable with every other variable in a data set. This leads to the generation of that small set of linear variable combinations which can stand in for all variables with the maximum preservation of the original intercorrelation patterns. These linear variable combinations, called *factors*, are in fact scales, and although each factor includes *all* variables in its construction, each is always strongly dominated by a unique group of highly inter-correlating (multicollinear) indicators which point to its proper interpretation.

The researcher fixes the meaning of each factor by examining the correlations of all of the original variables with the new scales, identifying the group of variables most heavily contributing to its construction (all but a few will correlate only trivially), and then searching for the common theme uniting the variables in the defining set. Finally, the procedure assigns each individual case (here, cohort student) a position, termed a *factor score*, on each scale generated by the procedure, expressed in standard deviation units. The end result is the replacement of an unmanageably large set of individual independent variables, the impacts of which are next to impossible

---

[9] The independent variables entered into the regression represented direct Sector B psychological, attitudinal and aptitudinal influences not at all, and Sector A socio-cultural, economic and pre-college educational influences only weakly (e.g., no direct socio-economic measures, no data on current employment or familial circumstances and pressures, no data on parental education or on the environment and values of the childhood home, no secondary school performance data). Available Sector A data included a small set of simple standard background variables drawn from the student records database (age, race/ethnicity, married/single, pre-PGCC level of educational attainment and identity of secondary school) and a few additional data index constructions (e.g., an inferential socio-economic index construction based on Census tract data, and a delayed entry indicator created by comparing high school graduation year with the cohort inaugural year). Only Sector C (measuring persistence, effort and performance) and D1 (measuring remediation need and development program participation and progress) were comprehensively represented in the data.

| Table 1. Extracted Factor Scales and Their Interpretation | | |
|---|---|---|
| Factor Label | Interpretive Name | Representative Defining Variable Loadings |
| TRSSEEK | *Degree-Seeking and Future Study Intention* | Specific Career Curriculum Choice (-.87)<br>Any Transfer Curriculum Choice (.73)<br>Arts & Sciences Curriculum Choice (.46) |
| PERSMOTV | *Special Personal Motives for Attendance* | Stated Enrollment Reason: Job/Other Personal (.90)<br>Stated Enrollment Reason: Transfer (-.89)<br>Any Non-Occupational Curriculum Choice (-.24) |
| PREPARED | *College Preparedness/ Remediation Progress* | Dev. Course-Taker/Programs Completed (.84)<br>Developmental Math Required (-.65)<br>Number of Developmental Programs Required (-.63)<br>Developmental Math Placement Test Score (.51)<br>Mean Placement Test Score (.47) |
| LAUNCH | *Early Term Survival and Progress* | "Good Start" - Enrolled Fall 1/Spring 1/Fall 2 (.69)<br>Any First Year Good Academic Standing (.66)<br>10+ Credit Hours Earned (.57)<br>Enrollment beyond First Fall (.53)<br>Any First Year Credit Hours Earned (.47)<br>First Year Cumulative Grade Point Average (.46) |
| EFFORT | *Study Load Carried* | Mean First Fall/Spring Course Hour Load (.94)*<br>Mean Major Term Course Hour Load 9+ (.92)*<br>Mean Major Term Course Hour Load (.91)*<br>First Fall Course Hour Load 15+ (.86)* |
| PERFORM | *Course Performance and Academic Status* | First Year Cumulative Grade Point Average (.83)<br>Last Cumulative Grade Point Average (.82)<br>Earned/Attempted Credit Hour Ratio (.79)**<br>Always in Academic Good Standing (.63)<br>Proportion of Major Terms in Good Standing (.62)<br>Any Probationary Standing Period (-.43) |
| PERSIST | *Attendance Persistence/ Continuity* | Attendance Duration from Fall 1 to Last Term (.86)<br>Number of Major Terms Enrolled (.75)<br>Major Term Enrollment beyond First Year (.74)<br>Major Term Enrollment beyond First Fall (.63)<br>10+ Credit Hours Earned (.55)<br>Sequential Major Terms/No "Stopping Out" (.53) |

| Table 1. (Continued) | | |
|---|---|---|
| Factor Label | Interpretive Name | Representative Defining Variable Loadings |
| PROBLEMS | *Patterns of Remediation Difficulties and Stalled Academic Progress* | Number of Course-Taking Developmental Areas (.77)<br>First Year Developmental Course-Taking (.73)<br>Number of Developmental Courses Repeated (.71)<br>Any Restricted Academic Status Period (.66)<br>Always in Good Academic Standing (-.61)<br>Mean Placement Test Score (.-.57)<br>Any Credit Course Attempts (-.43)<br>Passed at least 1 Attempted Credit Course (-.39)<br>Developmental Math Required/Incomplete (.33) |
| ATTITUDE | *Implied Study Motivation & Success Commitment* | Attended Both Day/Evening - Any Major Term (.71)<br>Attended Both Main/Extension Center Classes (.69)<br>Attendance Duration from Fall 1 to Last Term (.64)<br>Any Summer Term Enrollment (.63)<br>Any Change in Study Major (.61)<br>Sequential Attendance/No "Stopping Out (.49)<br>"Good Start" - Enrolled Fall 1/Spring 1/Fall 2 (.43) |
| SUPPORT | *Institutional Financial & Academic Support* | Any Pell Grants Received (.88)<br>Any First Year Pell Grants Received (.82)<br>Any Minority Retention Program Participation (.35)<br>Any Job Planning/Study Technique Courses (.21) |
| HISES | *Inferred Student Socio-Economic Background Level* | Lifestyle Cluster Social Class Trichotomy (.82)***<br>Home Census Tract % Upper White Collar (.76)<br>Home Census Tract Median Household Income (.74)<br>Home Census Tract % College Graduate Adults (.74)<br>Home Census Tract % Female-Head Families (-.73)<br>Home Census Tract % on Public Assistance (-.70) |

* Includes developmental course hours    ** Regular credit hours only
*** Based on *PG-TRAK90*, a P.G. County marketing and targeting system

to assess due to high collinearity among types of variables, with a small, manageable set of independent variable scales without any multicollinearity-within-type problems by definition.[10]

[10] The factor analysis was performed using *oblique* rotation rather than the more standard *varimax* rotation. The latter produces factor scales of heightened power which are mathematically prevented from intercorrelating, a quality preferred when the phenomenon being modeled can adequately be represented by a set of direct effect predictors. Oblique rotation, on the other hand, heightens scalar power without sacrificing the empirical tendencies of factors to intercorrelate. This form of rotation was preferable in our case because the phenomenon of academic achievement under study was too complex to be captured in a simple direct effects model.

Table 1, above, presents the array of factor scales generated by two factor analyses: the multi-scalar rendering of the original 56 Sector C and D1 independent variables, plus a supplementary factoring of a range of social class indicators into a single socio-economic scale. The table shows which scales emerged from our factor analyses, gives the standard label used to identify each in all further report tables and discussion (e.g., TRSSEEK), indicates which of the original variables dominated and defined each factor, along with their scale correlations (called *loadings* in factor analysis), and provides a brief description of the dimension underlying each factor, a summary interpretation of the theme common to the top loading original variables.

As can be seen, the oblique factor analysis reduced the original 56 Sector C and D1 independent variables to just ten factor scales. Arranged in rough causal distance from academic final outcome, these were TRSSEEK (Degree-Seeking and Future Study Intention), PERSMOTV (Special Personal Motives for Attendance), PREPARED (College Preparedness and Remediation Progress), LAUNCH (Early Term Survival and Progress), EFFORT (Study Load Carried), PERFORM (Course Performance and Academic Status), PERSIST (Attendance Persistence and Continuity), PROBLEMS (Patterns of Remediation Difficulties and Stalled Academic Progress), ATTITUDE (Implied Study Motivation and Success Commitment), and SUPPORT (Participation in College Financial Aid and Special Academic Support Programs). Half of the items on this list were expected and fell neatly into the academic model component boxes of Figure 1 — TRSSEEK, PERSMOTV, EFFORT, PERFORM and PERSIST; but the remainder were a surprise, and their appearance in the factor analysis constituted the first real findings of our research.

First, apparently two separate dimensions underlay the original set of variables relating to college preparedness and participation in remediation programs. Rather than a single REMED factor, our analysis automatically generated two scales: one measuring students in terms of level of study readiness and developmental *placement* status (PREPARED), and a second (PROBLEMS) combining variables tracking absence of student developmental *progress* with several drawn from the credit course performance side of the academic process. These latter indicators flagged various negative credit study conditions (e.g., restricted and probationary academic standing, failure to enroll in credit courses, and failure to convert credit enrollments into earned credit hours), and when taken together with developmental non-progress pointed to the existence of certain special patterns of academic frustration and stagnation. The emergence of the PROBLEMS factor implied that when students experience difficulties across a range of academic processes, including and especially the strategically placed developmental process, the negative effects of these individual difficulties tend to coalesce into a self-reinforcing progress-inhibiting syndrome with its own independent effects.

Second, the factor analysis discerned an important timing aspect in the measurement of academic performance. Along side EFFORT, PERFORM, PERSIST and PROBLEMS, which assessed a student's progress-relevant behavior over his or her whole study career, a separate LAUNCH factor also emerged, based on variables exclusively concerned with gauging what happened to students during the earliest terms. Special to the LAUNCH factor were its top loading variable, the "Good Start" indicator (attendance in all three initial major terms), and launch period survival (enrollment beyond the first year). But the other key variables were simply Launch Period versions of the same sort of indicators out of which the EFFORT, PERFORM, PERSIST and PROBLEMS scales were constructed (for example: First Year Cumulative GPA versus All Term Cumulative GPA). This suggests that the common intuition that there is something special and formative about a student's first several college terms is substantiated: The Launch Period seems to function as a shakedown time which separates the absolutely unready and unmotivated from those more willing and able to handle the challenge of college-level study.

Third, the factor analysis unexpectedly brought forth two scales relating to two subsidiary dimensions of the academic process. A general measure of student receipt of special institutional support (SUPPORT) was created by the procedure out of variables measuring student participation in financial aid programs (Pell Grants), in academic counseling and support programs (ALANA, a minority mentoring program and the TRIO-funded Student Support Services), and in study skills training and career planning courses. Prior to the factor analysis we were uncertain whether the few variables we possessed to track the presence and use of institutional support efforts would factor into a single independent scale, or whether they would be absorbed in the construction of the more theoretically central factors. Factor analysis proved that this less noticed side of the academic process has a statistical life of its own.

Also, the procedure created a truly unexpected factor (ATTITUDE) by combining the effects of a seemingly miscellaneous set of Sector C variables. Contributors to the scale included flags of enrollment in both day and evening classes, in both main campus and extension centers classes, and in summer term classes, plus measures of attendance consistency and duration, and a flag for having made any change in study major. All of these were indicators of student academic choices requiring extra effort regarding class attendance, or in one instance (change in study major) extra thoughtfulness regarding study objectives. The key word was "extra," suggesting that such choices *indirectly* reflected a *highly motivated attitude* toward college. The unexpected emergence of this arguably *psychological* factor, the sole representative of Academic Process Sector B among our factor scales, was very gratifying. Lacking direct psychological measures, we had initially feared that we might have to leave this entire critical area unexplored in our first modeling venture.

TABLE 2. Regression of Student Background Variables And Academic Factor Scores upon Study Achievement (N = 2,386)

| $r^2$ Rank | Independent Variables* | Zero-Order $r^2$ | Semi-Partial $r_p^2$ |
|---|---|---|---|
| ALL VARIABLE REGRESSION | | $R^2$ = .469 | $\sum r_p^2$ = .111 |
| 1 | ATTITUDE | .239 | .035 |
| 2 | PERSIST | .207 | .003 |
| 3 | LAUNCH | .198 | .018 |
| 4 | PERFORM | .159 | .018 |
| 5 | SUPPORT | .118 | .001 |
| 6 | EFFORT | .102 | .009 |
| 7 | PREPARED | .101 | .015 |
| 8 | PROBLEMS** | -.098 | -.006 |
| 9 | *White*+ | .058 | .001 |
| 10 | *HS Quality* | .055 | .002 |
| 11 | HISES | .033 | .000 |
| 12 | PERSMOTV** | -.027 | -.001 |
| 13 | *Young* | .027 | .001 |
| 14 | *Immed Entry* | .027 | .000 |
| 15 | TRSSEEK | .014 | .000 |
| 16 | *Single* | .007 | .000 |
| 17 | *Female* | .002 | .000 |

*Factor scale independent terms, described in the text, are indicated by CAPITALIZED variable names; non-factor-based variables are as follows: *White*+ (White/Asian/International Students = 1, African American/Hispanic Students = 0); *Female* (Female Student = 1, Male Student = 0); *Young* (Under 21 = 1, 21 and Over = 0); *Single* (Never Married/Separated/Divorced = 1, Now Living with Spouse = 0); *Immed. entry* (From High School to PGCC within 1 Year = 1, More than a Year = 0), *H.S. Qual* (P.G. Private = 4, High Rated P.G. Public = 3, Other P.G. Public = 2, Other H.S./G.E.D./No Diploma = 1).

** Mathematically, all squared values are positive; the minus signs shown preceding the correlation values for PROBLEMS and PERSMOTV are meant merely to flag the fact that the original unsquared correlation was negative.

Finally, the table shows the derivation of the only social/educational background factor we created through factor analysis to be used in our regression work — HISES, measuring home neighborhood socio-economic ranking. Because of their intrinsic interest, we decided to enter the remaining six background variables into our regressions in their original form: Five were dichotomous attribute variables representing race/ethnicity (*White+*), gender (*Female*), age (*Young*), marriage status (*Single*) and duration between high school graduation and college entry (*Immediate Entry*). The last was a three-point scale of quality of high school education based on the school ratings of locally knowledgeable college staff (*HS Qual*). The labels of the dichotomous variables were selected to point to the background quality being measured, and as customary, the presence of a quality was set to 1 and its absence to 0. For example, for the *Female* indicator, those in the criterion category Female were flagged with a 1, those Not Female (i.e., males) with a 0.[11]

The data reduction program completed, we were finally ready for our first substantive regression analysis. The basic results of the forced entry regression of the 11 new factor scales and 6 remaining original variables against achiever classification can be seen in Table 2 above. The table displays the following statistics: (1) the whole-equation $R^2$, allowing us once again to assess the adequacy of the data set to cover the phenomenon of academic achievement; (2) the whole-equation sum of the squares of all variable semi-partial correlations ($3\ r_p^2$), a measure of how much of the total variance is explained by individual variable *direct effects* and therefore an indirect test of the appropriateness of the regression model; (3) the square of the Pearson zero-order correlation for each independent variable ($r^2$), which gives the proportion of the dependent variance it explains by both its direct and indirect effects; and (4) each independent variable's semi-partial (direct effect only) correlation with achiever classification ($r_p^2$).

The following points can be made based on the regression results found in Table 2:

- Our data reduction program was successful. Although there is always some loss in explanatory power when one scales any set of related variables, the scales emerging from the factor analysis retained most of the original independent variable joint correlation with achievement. The regression equation $R^2$ fell by only 16 percent of the achievement variance explained (.63 to .47, 72-variable and 17-variable $R^2$s respectively).

---

[11] The variables defining HISES are given in the factor descriptions in Table 1. Along with the numerically highly predominant white student contingent, the criterion category of *White+* also included the small Asian American and international student minorities in the cohort, which behaved very similar to the white students with respect to academic achievement. See methodological Appendix B for a detailed discussion of the construction of, and methodological consideration relating to, the five student attribute dichotomous variables. Appendix B also provides a full description of the derivation of the High School Quality four-point scale.

- Given the closeness of original and revised $R^2$ correlations, our earlier remarks assessing the level of achievement variance we were able to explain with the predictors available remain valid — basically, while we were able to obtain better than average results compared with other studies of the correlates of study success, we still were unable to account statistically for around half the phenomenon in question because of important data gaps.[12]

- The $\sum r_p^2$ result strongly implied that regression was the wrong technique for modeling academic achievement. Less than a quarter (.11) of the joint explanatory power of the independent variables (.47) proved traceable to their individual *direct effects* upon achiever classification. This confirmed from a mathematical point of view what we already believed to be the case from a theoretical point of view: That the phenomenon of academic achievement is best conceptualized as a complex web of background and process variables *working together* to produce study success probabilities. With the possible exceptions of course performance and term persistence, the two variables most directly bearing on study success, there is no logical reason why any of the predictors *should* exhibit discernable direct effects.

- Most of the independent variables exhibited negligible individual levels of unique impact on achievement likelihood. But the direct effects of five predictors, though of low absolute magnitude, were strong enough to be noticeable: ATTITUDE ($r_p^2 = .035$), LAUNCH (.018), PERFORM (.018), PREPARED (.015) and EFFORT (.009).[13]

---

[12] If we shift the predictiveness standard from the theoretical to the practical, the performance of the regression equation looks a good deal better. Practical predictiveness is simply the ability to properly prophesy case placement on the dependent variable scale. Here we are talking about the overall rate of success to guessing student achiever /non-achiever classifications using the equation-based estimate of achiever probability for each case. We calculated the proper placement rate, often called the *coefficient of concordance*, for the 17 variable equation, using an achiever probability of >.5 as the criterion for predicting an achiever classification. The coefficient's value turned out to be .84 — that is, in 84 percent of the cases, estimated student achievement classification was in accord with known student achievement classification. The reason for the large gap between the theoretical and practical predictiveness of the equation is that there were relatively few assured estimates of case achiever classification (say, probability below .3 and above .7) and a preponderance of unsure estimates in the middle range (above .5 probability). In other words, the attributes of the typical case tended to work at cross-purposes with respect to achievement likelihood or provided only scanty clues (low theoretical predictiveness), but when processed through the equation their combined slight weight toward one side or the other was sufficient to "tip the scale" in the correct direction (high practical predictiveness).

[13] The noticeable direct effect of PERFORM was non-problematic because it fitted with the theory, but those of the other variables were somewhat counter-theoretical. The most likely explanation for these minor discrepant readings is that they represented the *indirect effects* of some of the influences on study success gone unmeasured due to the limitations of our data.

- The table shows independent variables in rank order according to $r^2$ level, which simultaneously measures, in explained variance terms, both the direct *and* indirect impact of a single variable upon a dependent. When the second variable represents the outcome of a complex causal network and the first variable is one of the predictors of that network, then $r^2$ can be loosely interpreted as reflecting the *overall* importance of the role played by the predictor in that network's effect upon the dependent variable. Eight predictors registered $r^2$s of around the .10 proportion of achievement variance explained or better: ATTITUDE (.24), PERSIST (.21), LAUNCH (.20), PERFORM (.16), SUPPORT (.12), EFFORT (.10), PREPARED (.10) and PROBLEMS (–.10).

- The overall causal network importance of the eight predictors just mentioned was furthermore confirmed by a supplementary *stepwise* regression analysis. When a regression is performed using the stepwise inclusion/exclusion procedure, the only predictors entered from a larger set of independent variables are those which a probability algorithm determined were essential to maximizing the predictive power of the regression equation. For regression entry the stepwise procedure selected the above eight, plus *White*+ and *HS Quality*. This yielded an equation with an $R^2$ almost identical to that of the 17-variable equation; in other words, the network could be reduced to just these ten predictors with almost no loss to its power to determine achievement probabilities.[14]

- That the factor scales measuring course performance, attendance persistence, term study load, and college preparedness/developmental placement should turn up in a list of those most central in a network of academic achievement causality is perhaps not surprising. These are, after all, the variables representing the core elements of the academic process, and two of them — PERSIST and PERFORM — directly capture the essence of the two simple but crucial activities which, if engaged in often enough, mechanically guarantee success at college: enrolling in courses and passing courses.

- The entries on the list most central to the network of achievement predictors which deserve real comment were ATTITUDE, LAUNCH and PROBLEMS. None of these factors measure commonly accepted aspects of the academic process, yet they all apparently made important contributions to setting the probabilities of college success for Cohort 1990 students. In fact, level of student success motivation (as implied by engaging in various "extra effort" study behaviors) was at the top of the list ($r^2$=.24), and survival of and quality

[14] The stepwise regression output also flagged five predictors from among the ten entered as contributing the most to the step-by-step calculation of the regression equation — in order of equation entrance, these were ATTITUDE, LAUNCH, PREPARED, EFFORT and PERFORM.

academic performance and attendance during the "launch period" (first three major terms) came in third (.20). The tendency of some students to fall into a debilitating syndrome of academic difficulties was last on this list (-.10), but the fact that it made the list at all is notable because the phenomenon measured by the PROBLEMS factor has until now gone unreported as a separate factor in conditioning success probabilities.

- Finally, what independent variables did not make the network-central list is also of some interest. Absent were the two academic process factor scales dealing with curriculum choice and study objective (TRSSEEK, measuring student preference for transfer-related programs over career-related programs and no program choice, and PERSMOTV, measuring the degree to which a student has non-standard motives for attending college — e.g., personal enrichment). Also absent were all variables measuring student social/educational background (although our stepwise regression did include *White*$^+$ and *HS Quality* as the last two terms of its equation).

Table 3. $R^2$ for Regressions Of Study Achievement with Independent Variables Selected by Process Type (N = 2,386)

| PROCESS TYPE | (VARIABLES INCLUDED) | $R^2$ |
|---|---|---|
| *Whole Model* | (All 17 Variables) | .469 |
| General Academic | (EFFORT, PERFORM, PERSIST, PROBLEMS) | .355 |
| Launch Period | (PREPARED, LAUNCH) | .256 |
| Special Attributes | (ATTITUDE, SUPPORT) | .249 |
| Background | (HISES, *White*$^+$, *Female, Young, Single, Immed Entry, HS Qual*) | .104 |
| Study Orientation | (TRSSEEK, PERSMOTV) | .034 |

Such findings made us wonder how independent variable impacts might vary by larger predictor categories, a question we explored in a concluding supplementary series of regressions resulting in Table 3 just above. Displayed are the $R^2$ results, from highest to lowest, for five regressions, each of which was generated by a different set of independent variables. Each entry group included only predictors related by a defined academic process type — e.g., TRSSEEK and PERSMOTV are predictors of the "Study Orientation" type. Here $R^2$ indicates how much of the total achievement variance can be collectively explained by all variables of a given process type working together. Broadly speaking, $R^2$ for a predictor type is analogous to $r^2$ for an individual predictor. Table 3 findings were as follows:

- As we expected, given the individual variable $r^2$ results of Table 2, the regressions based on the core academic process variables (General Academic $R^2 = .36$), on early term performance and behavior (Launch Period .26), and on unusual considerations like success attitude and level of institutional support (Special Attribute .25), account for the most achievement variance.

- Collectively, the two Study Orientation variables TRSSEEK and PERSMOTV seemed to impact on achievement variance hardly at all ($R^2 = .03$). This finding lends little credence to the belief common among educators that students in traditional academic programs or attending for traditional academic reasons have an important edge on those studying for job-related or personal reasons.

- Even working collectively, the seven social/educational background variables managed to explain only a very modest proportion of the total achievement variance ($R^2 = .10$). Considering all of the attention devoted by the education community to the supposed influence of ethnicity, gender, poverty and the like on learning, this calls out for comment.

- We can think of two technical factors which might have held down Background regression $R^2$: (1) Our set of variables did not cover important aspects of student background, environment and personal characteristics like familial and job pressures, high school performance and native intelligence, to name just a few; (2) Some of the background indicators we used were methodologically weak (e.g., HISES indirectly derives student social class from home Census tract socio-economic data). But whatever the dampening influence one attributes to these, the fact still remains that race/ethnicity and gender variables were unaffected by such possibilities, and neither of these classic background indicators proved to be prime predictors of study success at PGCC.

Before pushing on to reporting our findings based on causal path analysis, we need to make one additional point regarding our regression findings: It is easy to be misled by our regression estimates of the relative "causal network-centrality" of achievement predictors, and one must take great care in interpreting them. A good general reason for exercising caution here is that regression is, at best, an awkward analytic tool for getting at such matters. But the particular reason is that the idea of "causal centrality" is not equivalent to the idea of "causal importance." Just because a predictor was not found to be a prime focal point around which a good deal of the total interaction among many other elements in the network was organized, does not necessarily mean that the predictor had no vital role to play. As will become quite clear as we next examine the results of a proper causal path analysis of our data, many "non-central" predictors play important *localized* roles within the causal network.

*Findings from Path Analysis*

With what we learned from the regression phase of our research, we were able to proceed to causal path analysis. In a nutshell, path analysis results in a tested structural model of a causal network. Using a set of linear predictors, the researcher constructs a model of the phenomenon of interest, based on his best theoretical understanding of how the phenomena works. He lines up his variables to represent causal sequences, drawing arrows of causality to link them (arrow heads point in the direction of the causal flow). Not all possible paths among the variables will be drawn, only those informed by the theory.

The decision on which paths and directions to be included in the model is important because path analysis is fundamentally a *theory testing* procedure. Its output provides a set of *path coefficients* (standardized partial regression weights), for measuring the predictive association of any pair of variables controlling for the effects of all other variables causally linked to the pair. In a good model, most of the coefficients should exceed $p$ .10. In addition, it includes a variety of general assessment statistics for judging the overall adequacy of a model (*goodness-of-fit*) in terms of fitting the total reality embodied in the data. Starting with the same data set, different models, arising from different theories, may lead to very different statistical outcomes and adjudged levels of adequacy.

After processing the data through the model, if a review of the path coefficients and the goodness-of-fit statistics implies a need for model improvement, the researcher can make adjustments, such as adding, subtracting or re-formulating variables, adding new path possibilities or "pruning" away weak original paths, or changing the causal direction of some of the links. He can then re-test the model as revised, and re-assess goodness-of-fit. The cycle of model-assess-revise-reassess continues until the researcher is satisfied with the model's performance.

Our initial model began with all seventeen original achievement predictors, but in testing failed on the grounds of *non-parsimony* — a form of over-complexity which, in path analysis, implies a condition akin to multi-collinearity in regression analysis: Some sets of variables were so highly intercorrelated that they were overlapping in their predictive effect. A bit of experimentation traced most of the problem to the seven background variables, which it will be recalled were not subjected to factor analysis during the regression data reduction step. A quick factor analysis at this point found two factor scales which could be substituted for six of the seven background variables with little loss of predictive power — ADVANTAGD, a scale measuring tendency to come from a socially and/or educationally advantageous background (highly correlating with *White*$^+$, HISES and *HS Qual*), and TRADSTU, a scale measuring the tendency to fit the description of the traditional beginning college student (highly correlating with *Immed Entry*, *Young* and *Single*).

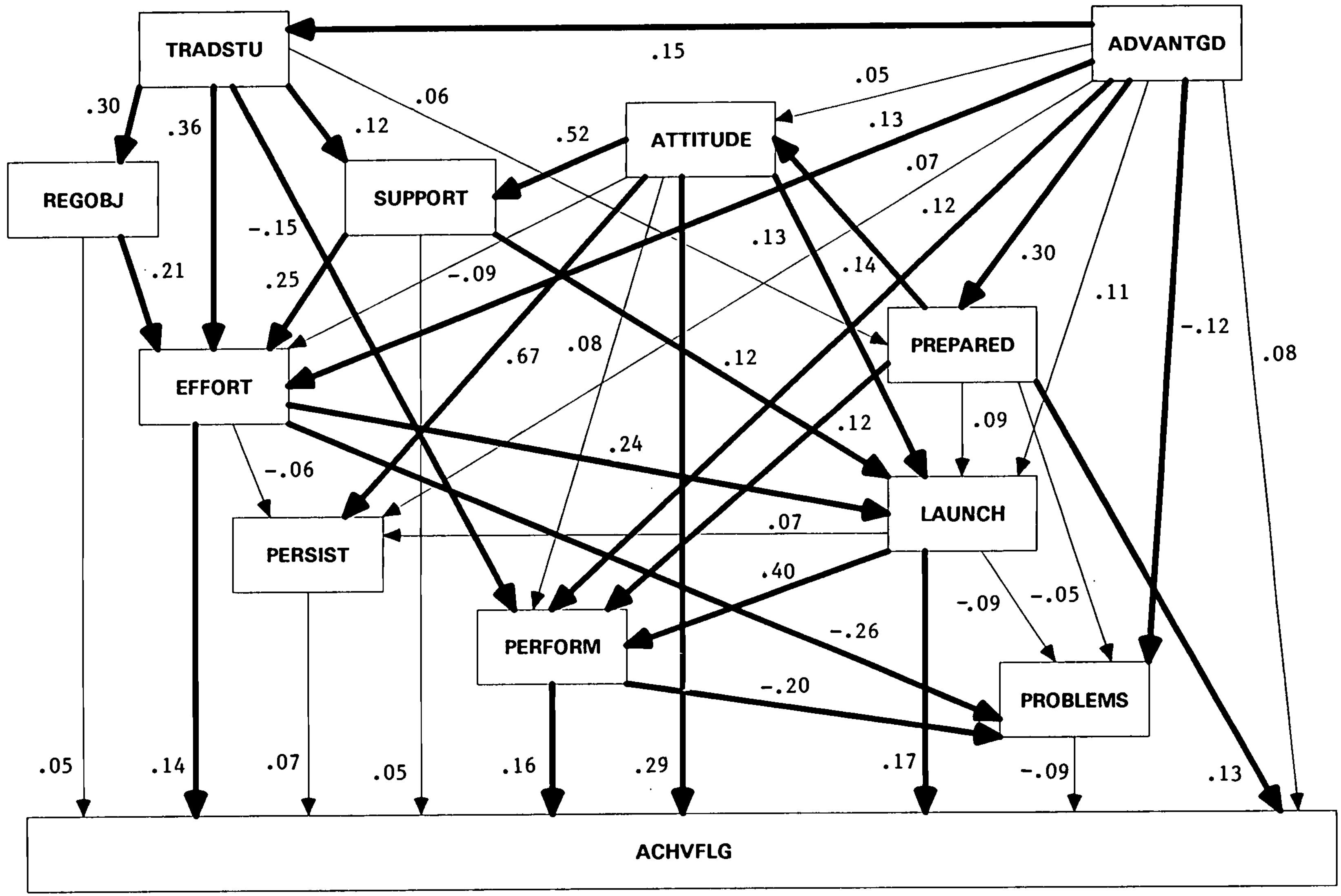



Figure 3. The Academic Achievement Causal Network

PATH COEFFICIENTS

.05 - .09

.10 and above

These two factor scales became the background variables for all subsequent path model tests; Female, which did not factor with the other background variables and showed only negligible linkages with other variables in the model was simply dropped from further analysis. Also, the two student academic choice factors (TRSSEEK and PERSMOTV) proved intercorrelated enough to be consolidated into a single factor scale — REGOBJ, standing for "regular objectives" — replacing the former in our path analyses. Once the factor scale substitutions were made, and the decision made to eliminate paths shown to be trivial according to the results of the test run (those with a $p$ of less than .05), the model as revised proved acceptable.

Figure 3, above, graphically depicts our final path analytic model. It shows the 11 predictor variables distributed in rough terms of temporal, logical and structural distance from the achievement classifier and from one another. The causal flow works downwards towards the bottom of the diagram, with many lateral links in-between. For example, the two background factors (ADVANTGD and TRADSTU), representing inherent personal qualities and pre-college circumstances, precede all other variables (top diagram position), as they do in theory and logic; next comes a broad band consisting of PREPARED, SUPPORT, REGOBJ, LAUNCH and ATTITUDE scales, measuring student entry-level preparedness, volunteer accessing of various student support programs, academic program choices, early term progress and academic success motivation level, all of which have to do with or start during the initial phase of a student's study career; lastly, and placed closest to the terminal final outcomes variable (ACHVFLG, the achievement flag), were the four general academic process variables EFFORT, PERSIST, PERFORM and PROBLEMS, respectively measuring mean study load, attendance persistence tendency, overall course performance, and academic problem symptomology.

The diagram indicates the existence of causal paths linking two variables, and the direction of path causality, by single-headed arrows. Each arrow is shown with its associated path coefficient ($p$), a probability weight measuring the impact of the first on the second variable, controlling for all causally preceding variables. Thick arrows indicate a moderate to strong link ($p$ .10) while fine arrows show marginal relationships (.05 - .09). Since path coefficients are *discrete* probability weights, *p*s for a specified sequence of paths (which we call a "trail") can be summed (discounting mathematical signs), and their total ($P_t$) can be taken as a measure of the probability weight of the entire *trail*.

The path analysis output for the model gave an $R^2 = .47$ for the variance explanatory power of the total causal system with respect to the achievement variable. This was identical to the $R^2$ reading for the earlier regression model, suggesting that a shift to a path model, though more restrictive since it limits the number of predictor-dependent variable linkages, nevertheless here involved no loss of power to explain the variance of the variable representing the key phenomenon.

Goodness-of-fit, however, is not measured in path analysis by $R^2$ because, technically speaking, in path analysis there is no single dependent variable. In any case, the diagnostic statistics supplied in our model's path analysis output indicated a level of data fit quite reasonable for an exploratory effort.[15] Below are the main findings of our model:

- Path coefficient patterns suggest that there are three large structures in the PGCC causal network of academic progress, formed by the chaining of mostly moderate-to-high variable links. Two of these are trails: An "Effort Trail" seems to run towards ACHVFLG involving TRADSTU > (REGOBJ * SUPPORT) > EFFORT > PERSIST, connecting with the achievement measure through REGOBJ, EFFORT and PERSIST. Also, a "Performance Trail" seems to travel thusly — ADVANTGD > PREPARED > LAUNCH > PERFORM > PROBLEMS — touching ACHVFLG through the latter three variables.

- The other large structure is a sort of traffic hub or rotary revolving around ATTITUDE. Moderate-to-strong paths run from it to ACHVFLG and to virtually all nodes along the Effort and Performance trails.

- Summed path coefficients for the two trails were almost equal: Effort Trail ($P_t = 1.56$), Performance Trail ($P_t = 1.58$). For the "Attitude Rotary" (all direct paths from ATTITUDE to other variables), the paths of which overlapped those of the two trails to a certain extent, $P_t$ estimate was even higher (1.88).

- At the head of both trails are background factors. TRADSTU starts off the Effort Trail by strongly linking to all three of the nearest trail nodes: REGOBJ ($p = .30$), SUPPORT (.12) and EFFORT (.36). Additionally, a moderately wide path ($p = -.15$) connects it with PERFORM. It would seem that whether a student can be described as traditional or as an adult learner has a good deal to do with the programs he or she will sign up for, whether the student will seek and receive financial aid and special tutoring, what kind of course performance can be expected, and especially how big a study load he or she will undertake on average.

---

[15] Goodness-of-fit measurement in path analysis is controversial. The technique is relatively new and statisticians have yet to reach a consensus on which of a host of diagnostic measures should be in general preferred. One of the stricter measures produced a figure placing our model just outside the bounds of technical acceptability (CMIN/DF = 8.319, 5 or less indicating acceptability), but a more generous measure implied an excellent level of model performance (GFI = .988, 1 meaning perfect fit). One could summarize this by say that path diagnostic measures put our model somewhere along a range from barely "OK" to well-performing, but importantly, all of them agree that our path model vastly outperformed our earlier regression model (e.g., perfect CMIN = 0, path CMIN = 174.7, regression CMIN = 7368.7).

- ADVANTGD starts off the Performance Trail, linking strongly with PREPARED ($p = .30$) and moderately with PERFORM (.12). In addition, it "shortcuts" to many other variables, most notably to EFFORT (.13), LAUNCH (.11) and PROBLEMS (-.12). Apparently, whether or not a student comes out of a socially or educationally advantageous background has a fair impact on his or her chances to begin study with good pre-college preparation or needing remediation, to pass or fail courses, to carry a full- or part-time study load, do well or poorly during the early terms, and to avoid or get trapped in various academic pitfalls.

- The $P_t$ for TRADSTU's three direct paths was .78 and for all of ADVANTGD's direct paths 1.13. This confirms what we suggested earlier concerning the low $R^2$ we got when regressing background variables exclusively upon achiever classification: Although background variables might not impact directly upon academic achievement, they may prove locally important within the total causal network.

- It is interesting to observe that the nodes along the Effort Trail prior to EFFORT and PERSIST were variables concerned mainly with *type of student* (traditional or Adult Learner) and *academic process* matters (program choice and institutional support), while the nodes along the Performance Trail prior to PERFORM and PROBLEMS were variables mostly having to do with *pre-college attributes* (race, age and high school type) and *initial study stage* considerations (developmental program participation and early term experiences).

- Most noteworthy regarding the latter was the major impact of LAUNCH upon PERFORM ($p = .40$): It would seem that such things as the immediate taking of credit courses, an experience of success in those course and especially the assiduous attendance during the first three major terms (the "Good Start" effect) are critical bases for latter successful course work.

- But perhaps the most notable finding of all was one already hinted at in our earlier identification of an Attitude Rotary as one of the three large structures of the model: ATTITUDE, the factor scale gauging student success drive, proved to be the most strategically important of all the model's variables. Not only was its cumulative presence felt more strongly than that of any other single variable ($P_t = 1.88$), but several of its individual impacts were highly significant in their own right: It directly linked with the course performance measure ( $p = .08$), with the study load measure (-.09), with the early progress measure (.13), and outstandingly with participation in financial aid and academic support programs (.52) and with the attendance persistence ( .67). Furthermore, among all variables ATTITUDE showed the greatest direct effect upon ACHVFLG, the achievement classifier ($p =$ .29).

Other noteworthy findings were:

- No non-trivial links could be modeled among EFFORT, PERSIST and PERFORM; the strongest was a not very consequential p=-.06 connection between EFFORT and PERSIST. Within the global network of achievement causality (controlling for all causally prior variables), study load, attendance persistence and course performance seemed to operate more or less independently of each other.

- When all controls are in effect, possessing traditional student attributes seems to be somewhat *negatively* related to course performance (TRADSTU > PERFORM, p=-.15). Perhaps this confirms a common belief among educators — that when all is said and done, Adult Learners seem to work more earnestly and succeed more often in the classroom than do younger students.

- SUPPORT, measuring participation in various student financial aid and academic support programs, linked with ACHVFLG at the $p=.05$ level and with EFFORT at the $p=.25$ level. While the former fact may be taken as somewhat confirming OIRA program assessment showing the achievement-enhancing properties of such participation,[16] the EFFORT link may be spurious at least in part, because some student support services require full-time status of participants and all are very difficult for part-time students to take advantage of due to scheduling of activities.

- The model portrays one clear, very strong link between a background measure and an important process variable: ADVANTGD > PREPARED ($p=.30$). In combination, race, social class and quality high school experience impact on college preparedness in a fashion no amount of statistical controlling can explain away or diminish.

We cannot conclude this discussion of the path analysis phase of our study without adding the following disclaimer: All of the above findings are tentative. In fact, the whole of our model of academic achievement must be considered highly provisional. We must recall, first of all, that the model as thus far developed misses many variables theory hypothesizes as critical to a proper understanding of academic progress. Secondly, the model's arrows, representing the flow of causality, are simply our best, most current "educated guesses" as to how the academic processes are structured. Small changes in linkage based on a refined understanding of achievement causality might have dramatic impacts on the performance and informative character of the model.

[16] See, for example, *The ALANA Minority Student Retention and Transfer Program,* Program Evaluation PE97-1 (October 1996).

### *Findings from Cluster Analysis*

In this last report section, we will review the results of analyzing our data by means of a radically different approach to modeling academic achievement — cluster analysis. As we mentioned already, a path model of student achievement provides a look at the causal forces at work only in terms of *abstract* central tendencies. Given the perfect path model, in fact, it would be possible in principle to take the absolutely typical student (he or she who had *mean* scores on *all* predictor variables), and work along all of the paths, multiplying unstandardized forms of the *p*-values by corresponding scale values, and sum the results into a proportion identical to the known achiever proportion for the whole student cohort. We reasoned, however, that many important insights might emerge from an analysis approach based on *concrete* study career patterns — how student careers tend to vary and how that variation relates to student achievement.

To accomplish this, we would have to break down our entering student cohort according to stable varieties of study career, modeled in terms of the same process variables used in the regression and path analyses. Cluster analysis is the statistical technique specifically designed to sort a population into groupings (called *clusters*) which reflect the divisions inherent in how members distribute across a range of attributes. To use it for our ends was mostly a substantive matter of specifying which variables to enter into the analysis.

First, we decided to make the selection so that the resulting clusters represented *pure* study career types. This meant dropping background variables from the cluster processing list, but retaining them as possible *post-facto* correlates of study career type. It also meant *not* entering ACHVFLG, the achiever classifier, directly into the clustering process, but saving it for later correlation. Second, we decided to employ the full 10-variable regression list of process variables rather than the 9-variable path list which replaced TRSSEEK and PERSMOTV with a single scale called REGOBJ. This was done to maximize the richness of the data to enable the best cluster solution to emerge.

Thereafter, we needed only to make a few technical decisions. The first was to settle on the *K-Means* form of cluster analysis, allowing us to reduce the set of clusters produced to a manageable number. Second, we chose criteria to judge the acceptability of a cluster solution: (1) That the resulting set of study career clusters be readily interpretable by academic process theory and common sense; (2) That the set include no small "odd ball" clusters, the result of trivial data differences; (3) That the study types identified by a cluster run be well-correlated with academic outcomes, providing a *prima facie* case for the substantive validity of the results, and a set of results meaningful in the context of a study of academic achievement.

The cluster model we settled on presented ten different student career clusters, each made up of cohort members experiencing the same unique journey through the PGCC academic process. The student career types were easy to interpret, through an examination of the pattern of each cluster's mean factor scores, and to produce a summary characterizations in the form of a cluster "nicknames." Furthermore, the $Eta^2$ correlation[17] between student career type and achiever classification, with the former as the predictor, came in at .381; considering that cluster analysis always results in some loss of the information contained in the variables it summarizes as a group typology, this represents a remarkable level of variance explanation — fully 80 percent of the achievement variance explained by the regression and path models was preserved in the cluster model. Table 4, below, embodies the model.

The table displays the ten student career clusters, labeled by nickname and in Percent Achiever order, as data columns containing cluster means (in indexed format) for the variables used in the cluster sort and interpretation. Indexing is a technique for reformulating multi-variable group averages to make comparison easier. Group averages are expressed as proportions of total population means, multiplied by 100. For example, if a cluster's Achievement Classification mean were 75.5 percent and the whole cohort mean were 31.2 percent, as is the case with a student career cluster we labeled "Dean's List," than that cluster's achievement index score would be 100 $\times$ (75.5 $\div$ 31.2), or 242 with rounding. This index score would indicate a cluster mean was nearly two and a half times the size of that for the population as a whole (always set to 100 in indexing format); index scores for group means less than the population mean would fall below 100.

The table's rows show indexed values across the ten clusters for each particular factor scale or attribute variable, plus the raw mean value for the entire cohort (always a mathematically set 50 in factor cases -- see table note – but a found percentage in non-factor cases); 11 factors are listed (the ten on which the clusterization was based, and REGOBJ for additional reference), seven achievement indicators (including the overall Achiever Classifier ACHVFLG), and nine background variables (the seven original regression indicators plus the two path analysis factor scales). In our discussion of the ten student career clusters we will also make use of supplementary data not shown in Table 3 but appearing in Appendix Tables A1-A4.

We will review the student career clusters by broad academic achievement groupings:

---

[17] $Eta^2$ is the appropriate statistic for gauging how much of the variance of a two-category variable can be explained by placement within a typology; it is highly analogous to the $R^2$ statistic used in linear models like regression and causal path analysis.

Table 4. Student Career Clusters within Cohort 1990 (Achievement $Eta^2 = .381$)

| Factors | (Raw) COHORT | DEAN'S LIST | SCHOLARS | COLLEGIATES | TRUE GRIT | PRAGMA-TISTS | FULL-TIME STRUGGLE | PART-TIME STRUGGLE | VANISHERS | UNPREPARED | CASUALS |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Student Career Clusters (Index Values) | | | | | | | | | |
| Cluster % | 100.0 | 9.8 | 6.6 | 14.3 | 9.9 | 4.4 | 5.6 | 10.6 | 7.0 | 15.5 | 16.2 |
| Cluster (N) | (2,386) | (233) | (158) | (342) | (236) | (106) | (134) | (254) | (168) | (369) | (386) |
| TRSSEEK | *50.0* | 104 | 110 | 118 | 103 | 103 | 99 | 85 | 101 | 90 | 95 |
| PERSMOTV | *50.0* | 83 | 87 | 70 | 83 | 172 | 97 | 162 | 75 | 96 | 107 |
| REGOBJ | *50.0* | 110 | 113 | 126 | 109 | 63 | 100 | 58 | 115 | 98 | 92 |
| PREPARED | *50.0* | 120 | 119 | 129 | 114 | 79 | 91 | 90 | 105 | 30 | 127 |
| LAUNCH | *50.0* | 126 | 137 | 155 | 80 | 117 | 102 | 116 | 54 | 74 | 62 |
| ATTITUDE | *50.0* | 174 | 132 | 87 | 154 | 83 | 109 | 106 | 68 | 68 | 63 |
| SUPPORT | *50.0* | 93 | 222 | 86 | 91 | 88 | 210 | 81 | 82 | 81 | 75 |
| EFFORT | *50.0* | 127 | 144 | 150 | 84 | 105 | 117 | 49 | 138 | 107 | 66 |
| PERSIST | *50.0* | 162 | 117 | 92 | 159 | 99 | 107 | 125 | 64 | 73 | 50 |
| PERFORM | *50.0* | 126 | 131 | 130 | 94 | 131 | 87 | 126 | 125 | 61 | 61 |
| PROBLEMS | *50.0* | 58 | 71 | 67 | 137 | 57 | 170 | 120 | 61 | 140 | 97 |
| ACHVFLG | 31.2 | 242 | 219 | 212 | 139 | 97 | 79 | 56 | 36 | 3 | 2 |
| *Transfers* | 13.4 | 202 | 208 | 325 | 89 | 63 | 44 | 21 | 36 | 2 | 4 |
| *Awards* | 8.6 | 340 | 253 | 122 | 172 | 121 | 43 | 55 | 14 | 0 | 0 |
| *Trs. or Awds.* | 18.9 | 244 | 217 | 254 | 123 | 89 | 47 | 35 | 28 | 2 | 3 |
| *Soph Status* | 12.3 | 237 | 221 | 145 | 163 | 107 | 128 | 85 | 49 | 4 | 0 |
| *Continuing* | 9.8 | 118 | 52 | 54 | 264 | 67 | 132 | 177 | 55 | 66 | 29 |
| *Dropout* | 63.7 | 36 | 48 | 51 | 67 | 107 | 103 | 114 | 135 | 146 | 152 |

(Continued)

Table 4. (Continued)

| Factors | (Raw) | Student Career Clusters (Index Values) | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | COHORT | DEAN'S LIST | SCHOLARS | COLLEGIATES | TRUE GRIT | PRAGMA-TISTS | FULL-TIME STRUGGLE | PART-TIME STRUGGLE | VANISHERS | UNPREPARED | CASUALS |
| ADVANTGD | *50.0* | 126 | 90 | 121 | 94 | 108 | 64 | 102 | 106 | 75 | 102 |
| HISES | *50.0* | 115 | 94 | 116 | 96 | 99 | 82 | 99 | 104 | 88 | 99 |
| *White*+ | 47.9 | 150 | 114 | 148 | 77 | 122 | 39 | 100 | 127 | 40 | 97 |
| *HS Qual** | 25.6 | 256 | 102 | 196 | 74 | 114 | 31 | 62 | 96 | 43 | 78 |
| TRADSTU | *50.0* | 114 | 107 | 123 | 97 | 81 | 109 | 63 | 109 | 107 | 86 |
| *Immed Entry* | 58.1 | 127 | 116 | 146 | 94 | 68 | 118 | 39 | 108 | 108 | 63 |
| *Young* | 64.4 | 100 | 90 | 113 | 77 | 51 | 88 | 30 | 94 | 97 | 91 |
| *Single* | 87.8 | 109 | 102 | 113 | 98 | 88 | 108 | 69 | 109 | 109 | 92 |
| *Female* | 57.8 | 104 | 123 | 89 | 108 | 114 | 31 | 124 | 82 | 90 | 78 |

NOTE: In the Student Career columns, all figures are indexed group means (Index = 100*(raw group mean/raw whole population mean)). In the Whole Cohort column, unitalicized figures are percentages of all cohort students in the variable criterion category; figures in italics (e.g. *50.0*) are transformed factor scale score means. In their original format, factor score whole population means are always 0, with scores below the mean indicated by negative numbers. This format does not permit indexing because indexing requires division by the population mean and mathematics forbids zero division. The transformation formula (Index = 50 + (20*cluster score mean)) resets the population factor mean to 50, with a constant multiplier (20) which has the effect of creating a factor score case range of between 0 and 100.

* Percent of students scoring 3 or 4 on the 4-point *HS Qual* Scale

40

A solid majority of the students in each of the first three clusters classified as achievers by the end of their fifth year of study, most of them the traditional way — either by earning PGCC awards, transferring to four-year colleges and universities, or both. The student bodies of these clusters, together comprising around a third of the cohort, were all characterized by disproportionate tendencies to give standard academic reasons for attending college and to choose transfer rather than occupational study majors; to arrive at college with good basic skills and to attend regularly and perform well during the critical early semesters; to undertake full-time study loads each term, to pass most courses taken while at PGCC with high grades, and to avoid problem patterns threatening their academic good standing. There were, however some interesting differences among them:

Dean's List (10 percent). Over three out of four (76 percent) cohort members making up the Dean's List cluster classified as achievers, the best record of any group; they were also more successful at earning awards than students in any other study career cluster. They tended to be very motivated (highest ATTITUDE score mean) and to persevere (highest PERSIST score mean). Their most unusual feature, compared with the two other top performing clusters, was their tendency to take evening session classes — only 24 percent were exclusively day students; among the top clusters, Dean's List was also the only one with an important percentage of part-time students (30 percent). Another top group distinction was the somewhat greater proclivity of Dean's List students to choose business, allied health and paralegal study majors. Demographically, Dean's List members were predominantly white students from the socio-economically highest ranked neighborhoods who came to PGCC from the County's most outstanding secondary schools (highest *HS Qual* index value of any cluster). Also, they were mostly "traditional students" — young, unmarried and straight from high school.

Good Scholars (7 percent). Nearly seven in ten (68 percent) Good Scholar students made it into the academic success category. Academically, they differed from Dean's List students by being more likely to attend PGCC full-time, to take mainly day classes and to receive some form of institutional support. However, their collective attendance persistence record, while still good, fell considerably short of first cluster's. Good Scholars were also somewhat more likely to major in non-occupational fields. The main contrast, however, was demographic: Good Scholars were the least likely of students in any of the top clusters to enter college from advantageous social and educational backgrounds. Fully a third came from the County's poorest Census tracts, almost two-thirds were non-white, and only 26 percent were alumna of top-ranked secondary schools. Also, a discernable minority of them failed to fit the "traditional student" label.

Collegiates (14 percent). While nearly matching (66 percent) the achievement records of Dean's List and Good Scholar students, the Collegiates contrasted with them by their strong transfer tendencies (44 percent — highest of any cluster and half again as much as the rate for the other two top clusters), a finding consistent with their fairly low mean PERSIST score. Compared with those in the other top scholastic groups, students here showed the most or second most pronounced tendency toward full-time study, day class enrollment, placing out of remedial programs, and majoring in transfer programs generally, and Arts and Science programs in particular. No surprise then that in the aggregate, those in the Collegiate camp also best fit the description of "traditional student." Collegiate students, collectively, had the second most socio-educationally advantaged backgrounds.

The second study career set (14 percent of the cohort) included two clusters exhibiting mid-level rates of academic success. Besides their modest achievement records, the main things they seemed to have in common were that both groups contained far fewer full-time students than did the top performing clusters (around 60 percent part-time to 40 full-time in both instances), and that both included disproportions of students taking courses at extension centers and above-cohort average proportions of adult learners. Otherwise, their profiles diverged:

True Grit (10 percent). Somewhat over two-fifths of the True Grit students won their way through to Achiever status (more often to awards rather than transfers), but not before overcoming a host of difficulties. Although True Grit whole-career course performance was about average for the cohort and college preparation level even a bit above the mean, the students in this cluster typically began their academic careers at PGCC on the wrong foot (fourth poorest LAUNCH index value of any group) and in the later semesters often stumbled badly (third highest PROBLEMS index value). What brought them a measure of success was tenacity (second highest PERSIST index value) bolstered by strong motivation (second highest ATTITUDE score). In fact, over a quarter are still striving to meet their PGCC objectives (26 percent continuing students, by far highest proportion of any cluster).

Solid majorities of True Grit students tended to be African Americans, from the moderate middle class areas of the county, with average high school educations. The also were significantly more likely to be delayed entry students (second highest percentage of any group) and showed a stong inclination to take evening classes.

Pragmatists (4 percent). The success rate of the Pragmatist cluster (30 percent) fell just one point below that of the cohort as a whole. Almost the opposite of True Grit students in study career patterns, the Pragmatists arrived at PGCC with poor basic skills and needing a great deal of remedial work (second lowest PREPARED index value of any cluster), but began well (LAUNCH index above average) and continued well (tied with the Good Scholars for highest PERFORM score; lowest PROBLEMS score). In fact, the academic record of the Pragmatist is so good as to be paradoxical, considering the achievement results. The following are key to understanding the Pragmatists: an only average PERSIST mean score, the highest mean score on PERSMOTV and the most consistent majoring in occupational fields of any cluster. This pattern suggests, as does the label, that the Pragmatist cluster is disproportionally made up of "positive" dropouts, students who worked hard at strictly personal, usually job-related objectives, who withdrew from college once these were met. Demographically, Pragmatists were predominantly older, middle class people. Unlike True Grit, a plurality were white and a disproportion had attended the more prestigious secondary schools in the county.

The third student career set (16 percent of the whole cohort) contained two under-achieving clusters. These consisted mostly of students starting college with sub-standard basic skills, who after getting through the Launch Period in average fashion (almost all survived as students into the second year), tended to find the remaining terms an uneven struggle:

Full-Time Strugglers (6 percent). Only a quarter of this predominantly full-time group (78 percent undertook at least 9 credit hours on average during a major term) managed to make it to the Achiever category after five years, and two-thirds of those qualified only in the weaker sophomore in good standing-sense. Full-Time Strugglers were fighters (above average ATTITUDE score). Ninety-two percent of them required some sort of remediation when they began their studies, but they worked hard as a group to overcome this deficit and 23 percent of them completed all their developmental work (second highest of any cluster), while the remainder kept plugging at it term after term (above average PERSIST score; 18 percent are still continuing students), helped by the school, both financially and tutorially (second highest SUPPORT score). Most, however, fell behind (highest PROBLEMS score) due to course difficulties (third lowest PERFORM score). Demographically, Full-Time Strugglers tended strongly to be African American students straight from high school, living in working class parts of the county (over half from the poorest Census tracts), most having graduated from lower-ranked secondary institutions.

Part-Time Strugglers (11 percent). Fewer than one in five (17 percent) of the Part-Time Strugglers reached Achiever status, and most of these were sophomores in good standing. Like the Full-Time Strugglers, students in this group exhibited below-par college preparation but demonstrated above-average motivation and attendance (fourth highest PERSIST score). But unlike the former, Part-Time Strugglers, when they were admitted to credit courses, tended to do well as a group (above average PERFORM score). Besides their tendency to get stalled in their remedial work, the academic difficulties the students fell into (fourth highest PROBLEMS score) had more to do with scattered effort; not only were they mostly part-timers, but also "stopping out" rates tended to be high here. Low participation in student services programs (low SUPPORT score) may also have been part of the problem. Demographically, Part-Time Struggle's students were the oldest of any cluster, from modest middle class neighborhoods, and unlikely to have attended a high-ranked secondary school. Their academic program orientation was clearly occupational and personal motives were strong.

The final study career set (39 percent of the cohort) featured a miscellany of three extremely low-performing clusters:

Vanishers (7 percent). The emergence of the Vanishers as a separate study career cluster was unexpected. It had a distinctly odd pattern: Its members exhibited study career patterns that ought to have spelled academic success, at least for a fair proportion of them (average or above average cluster scores on factor scales measuring regular academic orientation, college preparedness, study load, and course performance and very low mean score on PROBLEMS); yet only about one in ten ended up classifying as an Achiever. Other measures showed very low group levels of Launch Period progress and attendance persistence. It was as if the bulk of these students, on the very brink of successful academic careers, just disappeared for no apparent reason. On further reflection, however, we concluded that nothing could be more natural in this chancy world — educational analysts ought to make room in their theories for such outcomes; every term some students will be forced to withdraw from college for a variety of non-academic reasons, from unexpected pregnancy to ill health to job lay-off. The impression that this is the key to understanding the Vanisher group is reinforced by the fact that demographic attributes seem to distribute more or less randomly here, with the exception of an overrepresentation of international students.

Unprepareds (16 percent). In stark contrast to the Vanishers, nothing could be more unambiguous than the study career pattern of students in the Unprepared cluster, or less paradoxical than its connection to the cluster's negligible level of academic accomplishment (Achiever Classification under 1 percent). By *every* standard predictor of success except study load, the Unprepareds ranked lowest or among the lowest of the clusters. Especially evident was the pattern among developmental variables that gave the cluster its name: lowest PREPARED score, lowest mean placement test score, 100 percent placement into at least one developmental program, 76 percent placement into all three developmental areas, 0 percent completion of remedial requirements. Not only were the Unprepareds non-achievers, they hardly got past the starting gate: 44 percent not returning for a second term, 67 percent not returning for a second year, 57 percent *no* regular credits earned, 37 percent *no* regular credit courses taken. Unprepared students were typically racial/ethnic minorities from working class areas, with diplomas mostly from lower ranking secondary institutions. A majority were immediate entry students.

Casuals (16 percent). The last student career cluster that emerged was another apparent oddity — cohort members who, as a group, were seemingly capable of successful academic careers (second highest PREPARED score) but who collectively turned in an academic achievement rate of under 1 percent. In fact, only 30 percent of them enrolled beyond the first term, and over 60 percent quit after taking only one credit course. What is going on here? One possible explanation is that our filter for excluding non-degree-seeking students from our analyses of academic outcomes proved to be too conservative. The original size of the cohort, it will be recalled, was 2,643 students of whom 9 percent were dropped as non-degree-seeking on strict grounds; these were students who attended no term beyond the first year *and* either stated job- or personal enrichment reasons for attending PGCC or had not picked a major by their last term or both. The study pattern of this last group of students seems to imply a non-degree-seeking stance as well, only more loosely. We believe that this cluster represented, if not non-degree-students in the strict sense (never any intention of earning a degree or transfer), then at least enrollees with a very "casual" attitude toward their studies at PGCC. Casuals had the lowest ATTITUDE score, lowest SUPPORT score, and tied Unprepareds for lowest PERFORM score. If we combine this group with the earlier excluded non-degree-seeking group, we come up with something like a maximum estimate for the proportion of students effectively unintegrated with the normative academic process at PGCC — around 24 percent. Demographically, students in the Casual cluster do not seem distinctive, except in one way: a high disproportion were adult learners.

***Summary and Conclusions***

OIRA has been tracking the academic progress of a cohort of first-time students entering PGCC in the Fall 1990 semester, and now has five full years of background, behavioral and outcome data for this group. For the first time we possessed a longitudinal data set on our students complete and comprehensive enough to begin development of a model of the total academic process at Prince George's Community College. Preliminary to this, we thought through the theoretical issues involved in creating any such model, creating an ideal "paper" model which included all of the component elements which researchers in recent years have identified as critical to capturing the complex reality of higher educational institutions.

This theoretical construct broke down the academic process into three major divisions — *input* (exogenous or environmental variables, measuring pre-college student attributes and college-external personal circumstances), through-put (endogamous or process variables, measuring student academic behavior and progress) and *output* (criteria or outcome variables, measuring the fulfilment of formal and informal educational objectives). A comparison of the available data connected with Cohort 1990 and the inventory of elements in the theoretical model made it clear that at this early point we lacked the means for measuring many important aspects of the academic process — particularly those having to do with environmental factors (e.g., social class, financial circumstances, family and job pressures, high school performance, home values). We decided to proceed nevertheless with an *exploratory* modeling effort, to see just how much of the phenomenon of the PGCC academic process we could capture in a structured way with the data in hand. This paper reports the results of that effort.

The research first used regression analysis to gauge how much general power to explain student outcomes rested in our data set, and to work out certain methodological problems, especially *collinearity* among independent variables (the tendency of a set of variables to intercorrelate highly, a situation which indicates unnecessary duplication of measurements and tends to badly muddle the statistical analysis). The student outcome measure was OIRA's standard summary variable —the Achievement Classifier — which divided students into an Achievers (those who earned either academic awards or transfers to four-year schools or reached sophomore in good standing status by their last term of attendance) and Non-Achievers (all others).

Early regression runs of all 90 original independent variables against the Achievement Classifier detected serious collinearity problems. This meant that *data reduction* (the elimination of measurement redundancy) was needed, and we managed to use *factor analysis* (an automatic scaling technique) to consolidate our set of achievement predictors down to just 10 scales representing the various underlying dimensions of the academic process, without sacrificing too much achievement explanatory power. The 10 new scales represented the following process dimensions uncovered by factor analysis: transfer vs. occupational program orientation, personal (non-academic) motives for attendance, college skills preparedness and remediation need, early term survival and progress, typical study load carried, course performance and academic status, attendance persistence and continuity, the presence of structured patterns of remediation difficulties and stalled academic progress, level of study motivation and commitment, and participation in college financial aid and special academic support programs.

These 10 academic process factor scales, plus 7 student background indicators (race, age, gender, marriage status, home Census tract socio-economic status, recency of high school graduation, and scale measuring quality of secondary school experience) were run against Achiever Classification in a last regression. The resulting $R^2$ statistic, measuring the amount of statistical variance of the achievement variable explained by all the predictor variables, was .47 (forty-seven percent of the variance explained). This level of statistical explanation is more than respectable compared with past similar research in the field, but in an absolute sense indicated that our data could not account for over half of the total phenomenon of student outcomes, doubtlessly because so many theoretic aspects (particularly those dealing with environment factors) were left unmeasured. Nevertheless, we concluded that it was worthwhile proceeding in the development of a model if we kept firmly in mind that any such emerging from this research was to be considered the *highly provisional first effort*.

The second phase of the research involved actually creating an academic process model explaining student outcomes through the application of *causal path analysis*. Unlike regression analysis, the utility of which lies mainly in its efficiency as a tool for preliminary data exploration, path analysis was designed specifically for the construction and testing of complex models. Its use allows the modeler to arrive at a construct which best reflects the underlying causal structure of phenomenon for a given set of data; concretely, it results in a model diagram which explicates the causal links (*paths*) among predictor variables (represented graphically by labeled boxes with connecting arrows), marked with probability weights (*path coefficients*) indicating the associative power of each link. The path analysis output also provides diagnostic statistics for gauging the quality of the performance of the entire model.

Our causal path model resulted in the following findings:

- The total path model explained almost exactly the same amount of achievement variance as the regression equation (47 percent), a good sign since path models are more strictly defined and usually account for less variance compared with parallel regression equations; furthermore, advanced assessment statistics rated our model's ability to fit the data representing the student achievement phenomenon as fair or better.

- A central feature of the path diagram turned out to be the existence of two semi-independent "trails" (sequences of paths) of approximately equal probability weight leading to Achiever Classification.

- The first was the "Effort Trail" which linked the following in rough causal sequence: "traditional student" attributes (young, single, immediate from high school), transfer program orientation, level of institutional support, typical term study load, and attendance persistence.

- The second large feature was a "Performance Trail" of student socio-educational attributes (race, social class, quality of high school experience), college preparation level and remedial need, early term survival and progress, course performance, and academic problem syndromes.

- Another prominent feature of the path model was a sort of traffic hub or rotary revolving around study motivation level. Moderate-to-strong paths ran from it to Achiever Classification and to virtually all nodes along the Effort and Performance trails. The centrality of study motivation in student achievement, as represented by its strategic positioning in the model and its very high total probability weight, was perhaps the single most important finding of this study.

- Other key findings were the importance of the *Launch Period* (early term survival and progress), a prime node of the Performance Trail, and the significant role *institutional support* was shown playing in conditioning both Launch Period outcomes and typical study load. These two findings have major implications for academic policy.

- Finally, in this brief review, we should mention how the model depicted the specific way student background variables operated in the overall causal network conditioning student outcomes. Past research on the correlates of academic achievement often found student background factors like race and socio-economic status as having little impact on college success. The path analysis model, however, suggests that this lack of discovered correlation may have been due more to poor methodology than to the truth of the matter. Past studies typically looked only at the *direct effects* of student background, controlling for all other predictors, which naturally wiped out their apparent power because background variables, almost by definition, impact mostly *indirectly* on achievement. This is just what the path model revealed: variables measuring various forms of socio-educational advantage were strongly predictive *locally* in the model, especially affecting level of college preparation, while "traditional student" attributes proved to have a good deal to do with program orientation, level of institutional support, and study load.

The last phase of the research was the attempt to produce a second, alternative model of the academic process, complementary to the path model but less abstract. To the academic policy maker, a drawback of the path analytic approach is that everything is based on radical averaging and underlying any path model is something like the notion that reality can best be grasped by studying the behavior of a hypothetical single super-case, here the absolutely "typical student." But we all know that the typical student is a myth. Much can be gained, we reasoned, by modeling student achievement concretely, in terms of the *actual* set of *varying* study careers experienced by PGCC attenders.

*Cluster analysis* is a statistical procedure which automatically sorts cases into discrete groups (called *clusters*), according to their similarities and dissimilarities across a set of attributes and behaviors. We reasoned that in applying cluster analysis to the cohort data measuring academic attributes and behaviors (the 10 factor scores), we would in effect be generating a model of the student body based on stable study career patterns, which could then be correlated with the Achiever Classifier and with student background attributes to produce a very rich picture of how things work academically at PGCC. The clusterization of the data resulted in a 10 cluster model of study career, which explained 37 percent of the achievement variance, a very strong performance for a cluster model. The nicknamed study career clusters, in academic success order, can be described briefly as follows:

- **Dean's List.** Mostly strongly motivated white attenders from advantaged social and educational backgrounds with sterling academic careers; predominantly "traditional students," nevertheless many typically took evening classes. Nearly four-fifths ended up in the Achiever column.

- **Good Scholars.** With study careers almost as exemplary as those of Dean's List, these were mostly strongly motivated African American "traditional students" from the middle socio-educational ranks, disproportionately participating in institutional support programs. Over two-thirds finished in the Achiever category.

- **Collegiates.** Highest ranked socio-educationally, these mostly white very "traditional students" (the youngest and most straight-from-high-school group) strongly favored transfer programs, especially in the Arts & Sciences, and also had superior study careers. Around two-thirds were Achievers by their last term, showing the greatest tendency to transfer early rather than to finish degree work.

- **True Grit.** Many in this essentially African American middle class cluster of older students, often taking evening classes, experienced significant problems with remedial programs and credit courses, but over two-fifths eventually became achievers through drive and persistence.

- **Pragmatists.** Like True Grit students, the Pragmatists tended to be middle class adult learners, but unlike them were more likely to be white, part-time, oriented to occupational courses, and were very likely to give job-related reasons for attendance. Most arrived at PGCC poorly prepared, but did fairly well as a group in course performance. However, only around three in ten became Achievers, because of stalled academic progress and difficulties completing all required remediation. It seems probable that many dropped out early, having satisfied personal agendas.

- **Full-Time Strugglers.** Mostly working class African American full-time students straight from lower prestige high schools, who entered PGCC somewhat unprepared. Full-Time Strugglers, however, showed good attitude and persistence, and participated in support programs at far above the average rates. Nevertheless, only around a quarter became Achievers by their last term.

- **Part-Time Strugglers.** This older, more job-oriented group compounded the problems caused by high remediation needs and low study loads by attending only irregularly. Their persistence and decent course pass rates did not make up for this, and fewer than one in five ended classifying as Achievers.

- **Vanishers**. The study career of the Vanishers was peculiar — an excellent initial course performance record followed shortly by withdrawal — as if study had been cut short by some personal emergency like ill-health or financial collapse. Hardly more than one in ten made it into the Achiever category.

- **Unprepareds**. Arriving at PGCC needing the most intensive academic remediation, most of the students in this working class African American group did not survive the first year of study, and less than 1 percent became Achievers.

- **Casuals**. Mostly well-prepared, part-time students from middle and upper-middle class neighborhoods, many explicitly giving job and personal enrichment reasons for attending, who took very few courses and exerted little effort to get good grades in those they did take. Again, less than 1 percent became Achievers.

We found several aspects of this roster interesting:

- First, it taught us that top performing students are *not* necessarily "traditional students" from the better neighborhoods and high schools who leave for four-year institutions at the first opportunity (the equivalent of the Collegiate cluster). Of the three most academically successful clusters, one consisted mainly of evening students and a second of lower-middle class African Americans, and both had high award earned rates.

- Second, a goodly proportion of our students actually fell outside the regular parameters of college study: Around 7 percent of the cohort's members "vanished" in the midst of successful study careers, probably due to personal emergencies (the first time we have ever been able to pin a number to this phenomenon), and fully 16 percent proved to be "casual" course-takers, not serious about pursuing a degree or transfer.

- Third, the cluster model identified another 16 percent of the cohort as so unready for college work that they were beyond the best efforts of our developmental teachers and counselors to help in any real way; this one-sixth may represent something like an estimate of the proportion of students entering PGCC each major term who are almost destined to fail because of an effective inability to deal with college-level material.

- Fourth, and perhaps most important, is the observation that among clusters which exhibited high concentrations of less socio-educationally advantaged, adult, part-time and job-oriented students, those which accomplished the most academically had in their study career profiles high scores on either level of personal motivation or level of financial/academic support receipt or both. Sheer attendance persistence, often present, did not seem to be enough.

Although our path and cluster models were highly tentative and not complete, nevertheless their construction and examination was well worthwhile for the important insights they offered. From path analysis we learned about the critical importance of personal motivation and the Launch Period in conditioning achievement probabilities. And from the cluster model we discovered the importance of taking student career differences seriously. These core findings suggest that the college's recent efforts to establish academic support programs which reach students early in their careers at PGCC, are designed to build confidence and esprit as well as develop academic skills, and which can be customized to individual educational needs and objectives, are right on target.

Karl Boughan
Supervisor of Institutional Research

## A * P * P * E * N * D * I * X

Table A1. Cohort Sub-bodies by Social Background and Academic Outcome Variables

| Student Attributes | COHORT | Student Sub-Body Raw Statistics | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | DEAN'S LIST | SCHOLARS | COLLE-GIATES | TRUE GRIT | PRAGMA-TISTS | FULL-TIME STRUGGLE | PART-TIME STRUGGLE | VANISHERS | UNPRE-PARED | CASUALS |
| SES | 50.0 | 61.0 | 45.1 | 61.9 | 47.2 | 50.2 | 33.5 | 48.9 | 54.7 | 41.9 | 48.7 |
| % White | 39.5 | 80.1 | 36.7 | 58.5 | 28.8 | 42.5 | 7.5 | 40.6 | 42.8 | 16.8 | 35.8 |
| % Black | 49.5 | 26.2 | 42.4 | 24.6 | 60.2 | 40.6 | 79.9 | 48.8 | 35.1 | 79.7 | 51.6 |
| % Latino | 2.2 | 1.3 | 2.5 | 4.1 | 2.1 | .0 | 1.5 | 3.1 | 4.2 | 1.1 | 1.3 |
| % Asian/Pacific | 3.1 | 4.3 | 3.2 | 5.3 | 1.7 | 1.9 | 3.7 | 2.4 | 3.0 | 1.1 | 3.6 |
| % Native American | .5 | .8 | .6 | .6 | .8 | .9 | .0 | .0 | .0 | .0 | .8 |
| % International | 7.3 | 7.3 | 14.6 | 7.0 | 6.4 | 14.2 | 7.5 | 5.1 | 14.8 | 1.4 | 7.0 |
| % Age 17-19 | 64.2 | 79.4 | 72.2 | 90.6 | 61.9 | 40.6 | 70.9 | 24.4 | 75.6 | 69.6 | 50.0 |
| % Age 20-24 | 17.2 | 11.2 | 10.9 | 7.0 | 18.2 | 21.7 | 19.4 | 21.3 | 19.6 | 19.5 | 23.3 |
| % Age 25-29 | 7.0 | 1.3 | 10.1 | 1.5 | 9.7 | 17.9 | 5.2 | 13.8 | 1.2 | 5.1 | 9.6 |
| % Age 30 Up | 11.7 | 8.1 | 7.1 | .8 | 10.2 | 19.6 | 4.5 | 40.6 | 3.6 | 5.7 | 16.6 |
| % Female | 57.8 | 59.7 | 71.5 | 51.5 | 62.7 | 63.2 | 59.0 | 71.7 | 47.6 | 51.8 | 52.8 |
| % NeverMarried/Separated/Div. | 87.8 | 95.3 | 69.9 | 99.4 | 86.0 | 77.4 | 94.6 | 61.0 | 95.9 | 95.7 | 80.3 |
| % Immediate Entry from HS | 58.1 | 73.8 | 67.1 | 65.1 | 54.7 | 39.6 | 66.7 | 22.8 | 62.5 | 62.8 | 41.2 |
| % Private HS Diploma | 6.9 | 14.6 | 3.8 | 16.4 | 5.1 | 11.3 | 2.2 | 2.0 | 7.1 | 1.6 | 4.8 |
| % Top PG Public HS Diploma | 25.4 | 40.8 | 22.9 | 36.8 | 23.7 | 20.6 | 11.9 | 16.1 | 24.4 | 24.4 | 22.5 |
| % Exurban Elite-1 (*PG-TRAK90*) | 13.5 | 19.9 | 6.3 | 22.4 | 9.2 | 10.4 | 4.2 | 15.8 | 13.7 | 7.0 | 15.8 |
| % Black Enterprise-2 | 6.7 | 5.9 | 6.6 | 9.0 | 8.7 | 6.3 | 1.7 | 6.5 | 10.2 | 6.4 | 5.8 |
| % Beltway Havens-3 | 5.5 | 11.8 | 3.5 | 9.0 | 5.9 | 5.2 | 2.5 | 4.6 | 5.5 | 3.5 | 2.6 |
| % Upwardly Mobile-4 | 11.4 | 14.0 | 14.7 | 9.7 | 10.4 | 11.5 | 5.1 | 12.0 | 12.3 | 9.0 | 14.4 |
| % Black Middle America-5 | 12.1 | 10.4 | 9.8 | 10.2 | 13.7 | 10.4 | 12.7 | 19.4 | 10.8 | 15.7 | 7.6 |
| % Rural Development-6 | 12.2 | 16.3 | 7.0 | 11.2 | 14.1 | 17.7 | 6.5 | 9.8 | 6.9 | 12.6 | 13.9 |
| % Fort George-7 | .3 | .0 | .0 | .3 | .0 | .0 | .8 | .0 | 1.4 | .0 | .6 |
| % Cosmopolitans-8 | 2.6 | 2.2 | 2.9 | 5.0 | 1.8 | 5.2 | 1.7 | 2.4 | 4.1 | .8 | 2.0 |
| % Town & Gown-10 | .8 | 1.4 | .7 | 2.1 | .0 | 1.0 | .8 | .0 | .0 | .3 | .8 |
| % Minority Comers-11 | 5.4 | 2.7 | 9.8 | 3.4 | 4.5 | 5.2 | 6.8 | 3.3 | 6.8 | 6.4 | 6.1 |
| % Old County-12 | 3.1 | 2.7 | 6.3 | 3.7 | 3.2 | 3.1 | 2.5 | 2.4 | 5.5 | 2.0 | 2.0 |
| % Afro Blue Collar-13 | 10.4 | 4.1 | 10.5 | 3.4 | 11.9 | 6.3 | 23.8 | 11.6 | 6.2 | 14.6 | 11.5 |
| % Latino Mix-14 | 5.4 | 4.5 | 12.6 | 4.7 | 5.1 | 6.3 | 3.4 | 2.4 | 6.2 | 8.1 | 5.3 |
| % City Line-15 | 10.6 | 4.1 | 10.5 | 5.8 | 11.4 | 9.4 | 25.4 | 11.1 | 4.1 | 15.4 | 11.1 |
| % Achievers | 31.2 | 75.5 | 68.4 | 66.1 | 43.2 | 30.2 | 24.6 | 17.3 | 11.3 | .6 | .5 |
| % Transfers Only | 10.4 | 17.2 | 18.4 | 37.7 | 8.5 | 6.6 | 5.2 | 2.0 | 4.2 | .3 | .5 |
| % Transfer & Award | 3.0 | 9.8 | 9.5 | 5.8 | 3.4 | 1.9 | .7 | .8 | .6 | .0 | .0 |
| % Award Only | 5.6 | 19.3 | 13.3 | 4.7 | 11.4 | 6.5 | 3.0 | 3.8 | .6 | .0 | .0 |
| % Soph./Good Standing/Exited | 7.6 | 19.3 | 23.4 | 14.0 | 8.1 | 6.5 | 7.5 | 3.1 | 3.6 | .0 | .0 |
| % Soph./Good Standing/Cont. | 4.7 | 9.8 | 3.8 | 3.8 | 11.9 | 4.7 | 8.2 | 7.5 | 2.4 | .5 | .0 |
| % Continuing Only | 5.1 | 1.7 | 1.3 | 1.5 | 14.0 | 1.9 | 9.7 | 9.8 | 3.0 | 6.0 | 2.8 |
| % Unexplained Exit | 63.7 | 22.7 | 30.4 | 32.5 | 42.8 | 67.9 | 85.7 | 72.8 | 85.7 | 93.2 | 96.6 |

54

BEST COPY AVAILABLE



55

Table A2. Cohort Sub-bodies by Academic Variables

| Student Attributes | COHORT | Student Sub-Body Raw Statistics: DEAN'S LIST | SCHOLARS | COLLE-GIATES | TRUE GRIT | PRAGMA-TISTS | FULL-TIME STRUGGLE | PART-TIME STRUGGLE | VANISHERS | UNPRE-PARED | CASUALS |
|---|---|---|---|---|---|---|---|---|---|---|---|
| PROGRESS | 50.0 | 80.4 | 73.7 | 63.6 | 69.9 | 55.0 | 57.1 | 54.5 | 37.0 | 30.4 | 29.9 |
| CREDITS | 50.0 | 70.3 | 65.7 | 65.8 | 52.1 | 41.1 | 40.8 | 37.7 | 31.8 | 41.5 | 45.9 |
| REQURMTS | 50.0 | 68.3 | 62.8 | 69.1 | 59.5 | 57.0 | 50.8 | 43.8 | 54.2 | 28.2 | 31.9 |
| % Mainly Largo Campus Classes | 84.9 | 79.8 | 90.5 | 98.6 | 75.0 | 79.8 | 99.3 | 60.6 | 95.2 | 98.3 | 74.6 |
| % Mainly Day Classes | 53.4 | 23.6 | 43.7 | 77.2 | 36.8 | 60.4 | 65.7 | 31.1 | 78.6 | 70.7 | 45.1 |
| % Reason/4Year Transfer | 63.0 | 73.8 | 73.4 | 90.1 | 71.2 | 12.3 | 65.7 | 18.1 | 83.3 | 63.1 | 56.5 |
| % Reason/First Career | 15.6 | 14.6 | 11.4 | 7.3 | 11.4 | 21.7 | 17.9 | 16.9 | 14.3 | 22.2 | 18.9 |
| % Reason/Explore Subject | 14.7 | 6.0 | 8.2 | .9 | 9.7 | 48.1 | 10.4 | 41.3 | 1.8 | 11.7 | 21.0 |
| % Reason/Upgrade Job Skills | 3.6 | 2.6 | 3.8 | 1.2 | 6.8 | 4.7 | 2.2 | 11.0 | .6 | 2.4 | 2.3 |
| % Reason/Personal Enrichment | 3.1 | 3.0 | 3.1 | .6 | .8 | 13.1 | 3.7 | 12.6 | .0 | .5 | 1.3 |
| % Major/Transfer Program | 54.0 | 52.8 | 66.5 | 73.1 | 50.4 | 52.8 | 53.7 | 31.9 | 63.7 | 54.2 | 45.6 |
| % Major/CareerProgram | 33.8 | 41.6 | 29.7 | 20.8 | 41.1 | 38.7 | 41.8 | 43.3 | 28.6 | 36.9 | 26.7 |
| % Major/Other | 5.5 | 4.3 | 3.8 | 4.1 | 3.8 | 5.6 | 4.4 | 10.3 | 4.8 | 8.1 | 4.4 |
| % Major/None | 6.7 | 1.3 | .0 | 2.0 | 4.7 | 2.8 | .0 | 14.6 | 3.0 | .8 | 23.3 |
| % Arts & Sciences-*Transfer Major* | 1.9 | 2.1 | 1.3 | 4.4 | 2.5 | 2.8 | .8 | .0 | 3.6 | .8 | 1.3 |
| % Music/Art | 2.3 | 2.1 | 1.9 | 4.1 | 2.5 | 1.9 | 3.0 | .4 | 3.6 | 1.1 | 2.3 |
| % Computer Sciences | 2.4 | 2.6 | 4.4 | 2.3 | 4.2 | 2.8 | 3.0 | 2.0 | 2.4 | 1.6 | 1.3 |
| % Engineering | 2.5 | 3.9 | 4.4 | 4.1 | 2.5 | 1.9 | 2.2 | 1.2 | 1.8 | 1.4 | 1.6 |
| % Business Administration | 7.6 | 6.9 | 10.1 | 12.0 | 6.4 | 3.8 | 11.9 | 4.7 | 10.1 | 7.9 | 4.9 |
| % Education | 2.6 | 6.4 | 5.1 | 5.0 | 2.1 | 1.9 | 2.2 | .4 | 1.2 | 1.1 | 1.6 |
| % General Studies | 35.1 | 29.6 | 41.1 | 41.8 | 29.7 | 39.8 | 31.3 | 22.8 | 41.7 | 41.2 | 32.9 |
| % Business Studies-*Career Major* | 10.9 | 13.3 | 5.1 | 10.5 | 11.4 | 10.4 | 14.2 | 19.7 | 8.0 | 10.0 | 8.3 |
| % Computer Technologies | 4.4 | 3.9 | 3.6 | 1.8 | 3.0 | 6.6 | 6.2 | 7.1 | 4.2 | 6.6 | 3.6 |
| % Technical Trades | 4.5 | 2.6 | 3.2 | 1.5 | 3.4 | 2.8 | 3.0 | 6.1 | 7.1 | 7.6 | 6.0 |
| % Nursing | 6.1 | 6.9 | 7.6 | 4.4 | 10.2 | 11.3 | 7.5 | 7.9 | 4.2 | 5.4 | 2.3 |
| % Allied Health Technologies | 2.6 | 4.7 | 2.5 | 1.5 | 2.5 | .0 | 7.5 | 2.0 | 3.0 | 2.7 | 2.1 |
| % Child Care/Management | 1.6 | 1.3 | .6 | .9 | .8 | 1.9 | .0 | 3.5 | 3.0 | 2.4 | .8 |
| % Office Technologies | .8 | .0 | 1.9 | .3 | .4 | .9 | .0 | .8 | 1.2 | 1.6 | .8 |
| % Paralegal/Justice Technologies | 6.7 | 12.0 | 6.3 | 4.1 | 13.6 | 7.5 | 6.7 | 5.5 | 4.2 | 5.4 | 4.7 |
| % Misc. Traditional Trades | 1.0 | .4 | .6 | .0 | .0 | .9 | 1.5 | 2.0 | .0 | 2.2 | 1.8 |
| Average Devel Test Score | 48.3 | 53.5 | 51.5 | 53.1 | 46.3 | 49.1 | 41.3 | 46.7 | 52.4 | 39.8 | 50.4 |
| % Developmental Required | 55.6 | 32.2 | 39.9 | 36.0 | 66.5 | 53.8 | 91.6 | 66.9 | 36.9 | 100.0 | 33.2 |
| # Areas Placed/Dev. Studentx | 2.0 | 1.6 | 1.8 | 1.4 | 2.1 | 1.7 | 2.4 | 2.1 | 1.6 | 2.3 | 1.5 |
| % Completing Developmental | | 12.4 | 12.7 | 17.8 | 27.5 | 2.8 | 23.1 | 12.6 | 3.0 | .0 | 7.5 |
| # of Major Terms Attended | 3.4 | 5.7 | 4.7 | 4.0 | 5.1 | 3.7 | 4.6 | 4.0 | 1.7 | 2.0 | 1.4 |
| % Enrolled beyond Fall 1 | 77.1 | 100.0 | 100.0 | 100.0 | 100.0 | 91.5 | 95.5 | 97.6 | 45.8 | 55.8 | 29.8 |
| % Enrolled beyond Spring 1 | 61.9 | 98.7 | 87.3 | 78.1 | 98.7 | 69.8 | 81.3 | 85.4 | 23.2 | 32.5 | 13.2 |
| % Good Start -All 3 Early Terms | 56.0 | 76.8 | 78.5 | 72.5 | 45.8 | 54.6 | 72.7 | 54.4 | 11.7 | 34.1 | 9.6 |
| % No Stopping Out (3 Terms+) | 44.0 | 53.6 | 62.0 | 64.0 | 31.4 | 48.5 | 57.0 | 39.5 | 9.1 | 29.3 | 7.0 |
| Cumulative GPA | 2.0 | 2.7 | 2.6 | 2.5 | 2.3 | 2.4 | 1.5 | 2.7 | 2.2 | .7 | 1.2 |
| % Course Pass Rate <.75 | 33.7 | 7.3 | 8.2 | 7.3 | 22.0 | 14.2 | 44.8 | 17.3 | 28.0 | 80.0 | 61.1 |
| % No Credit Courses Attempted | 8.3 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 5.2 | 2.4 | 0.0 | 36.6 | 12.2 |
| % No Credit Hours Earned | 17.4 | 0.0 | 0.0 | 0.0 | 4.7 | 0.0 | 9.7 | 3.5 | 0.0 | 56.9 | 44.3 |
| % Passing English 101 | 45.6 | 87.1 | 72.8 | 82.8 | 58.0 | 54.7 | 43.3 | 34.6 | 45.8 | 5.7 | 12.2 |

56 57

Table A3. Cohort Sub-bodies by Social Background and Academic Outcome Variables

| Student Attributes | % COHORT | Student Sub-Bodies Index (100*[Group Raw/Cohort Raw]) | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | DEAN'S LIST | SCHOLARS | COLLE-GIATES | TRUE GRIT | PRAGMA-TISTS | FULL-TIME STRUGGLE | PART-TIME STRUGGLE | VANISHERS | UNPRE-PARED | CASUALS |
| SES | 60.0 | 122 | 90 | 124 | 94 | 100 | 67 | 98 | 109 | 84 | 97 |
| % White | 39.6 | 213 | 98 | 156 | 77 | 113 | 20 | 108 | 114 | 45 | 95 |
| % Black | 49.6 | 53 | 86 | 60 | 122 | 82 | 161 | 99 | 71 | 161 | 104 |
| % Latino | 2.2 | 59 | 114 | 186 | 95 | 0 | 68 | 141 | 191 | 50 | 59 |
| % Asian/Pacific | 3.1 | 139 | 103 | 171 | 66 | 61 | 119 | 77 | 97 | 35 | 116 |
| % Native American | .5 | 180 | 120 | 120 | 180 | 180 | 0 | 0 | 0 | 0 | 180 |
| % International | 7.3 | 100 | 200 | 96 | 88 | 196 | 103 | 70 | 204 | 19 | 96 |
| % Age 17-19 | 64.2 | 124 | 112 | 141 | 96 | 63 | 110 | 38 | 118 | 108 | 78 |
| % Age 20-24 | 17.2 | 65 | 63 | 41 | 106 | 126 | 112 | 124 | 114 | 113 | 138 |
| % Age 25-29 | 7.0 | 19 | 144 | 21 | 139 | 266 | 74 | 197 | 17 | 73 | 137 |
| % Age 30 & Up | 11.9 | 63 | 61 | 8 | 89 | 172 | 39 | 353 | 31 | 47 | 141 |
| % Female | 57.8 | 103 | 124 | 89 | 108 | 109 | 102 | 124 | 82 | 90 | 91 |
| % NeverMarried/Separated/Div. | 87.8 | 109 | 102 | 113 | 98 | 88 | 108 | 70 | 109 | 109 | 91 |
| % Immediate Entry from HS | 58.1 | 127 | 115 | 146 | 94 | 68 | 118 | 39 | 108 | 108 | 71 |
| % Private HS Diploma | 6.9 | 211 | 55 | 237 | 74 | 164 | 32 | 29 | 104 | 24 | 71 |
| % Top PG Public HS Diploma | 25.4 | 161 | 90 | 153 | 93 | 82 | 47 | 64 | 96 | 85 | 89 |
| % Exurban Elite-1 (*PG-TRAK90*) | 13.5 | 147 | 47 | 166 | 68 | 77 | 31 | 117 | 101 | 52 | 118 |
| % Black Enterprise-2 | 6.7 | 88 | 84 | 136 | 130 | 94 | 26 | 82 | 153 | 96 | 88 |
| % Beltway Havens-3 | 5.5 | 214 | 64 | 184 | 108 | 95 | 46 | 83 | 100 | 64 | 48 |
| % Upwardly Mobile-4 | 11.4 | 123 | 129 | 86 | 92 | 101 | 46 | 106 | 108 | 79 | 127 |
| % Black Middle America-5 | 12.1 | 85 | 81 | 84 | 112 | 85 | 104 | 159 | 90 | 128 | 62 |
| % Rural Development-6 | 12.2 | 134 | 57 | 91 | 116 | 145 | 70 | 80 | 73 | 103 | 114 |
| % Fort George-7 | .3 | 0 | 0 | 106 | 0 | 0 | 265 | 0 | 460 | 0 | 189 |
| % Cosmopolitans-8 | 2.6 | 85 | 106 | 192 | 70 | 200 | 66 | 90 | 159 | 33 | 79 |
| % Town & Gown-10 | .8 | 171 | 83 | 266 | 0 | 124 | 89 | 0 | 0 | 40 | 114 |
| % Minority Comers-11 | 5.4 | 51 | 192 | 63 | 84 | 96 | 126 | 61 | 164 | 119 | 114 |
| % Old County-12 | 3.1 | 88 | 203 | 120 | 104 | 100 | 81 | 76 | 178 | 66 | 66 |
| % Afro Blue Collar-13 | 10.4 | 40 | 102 | 33 | 116 | 81 | 231 | 113 | 79 | 142 | 111 |
| % Latino Mix-14 | 5.4 | 84 | 233 | 86 | 94 | 117 | 63 | 44 | 115 | 113 | 99 |
| % City Line-15 | 10.6 | 39 | 100 | 57 | 109 | 90 | 242 | 105 | 39 | 147 | 106 |
| % Achievers | 31.2 | 242 | 219 | 212 | 138 | 97 | 79 | 55 | 36 | 3 | 2 |
| % Transfers Only | 10.4 | 165 | 178 | 365 | 82 | 63 | 50 | 19 | 40 | 3 | 5 |
| % Transfer & Award | 3.0 | 330 | 317 | 193 | 113 | 63 | 23 | 27 | 20 | 0 | 0 |
| % Award Only | 5.6 | 345 | 238 | 84 | 203 | 152 | 54 | 70 | 11 | 0 | 0 |
| % Soph./Good Standing/Exited | 7.6 | 254 | 308 | 184 | 107 | 112 | 99 | 41 | 47 | 0 | 0 |
| % Soph./Good Standing/Cont. | 4.7 | 211 | 81 | 81 | 253 | 100 | 174 | 160 | 51 | 11 | 0 |
| % Continuing Only | 5.1 | 33 | 25 | 29 | 274 | 37 | 190 | 192 | 59 | 118 | 55 |
| % Unexplained Exit | 63.7 | 36 | 48 | 51 | 67 | 107 | 103 | 114 | 135 | 146 | 152 |

BEST COPY AVAILABLE

58

59

Table A4. Cohort Sub-bodies by Academic Variables

| Student Attributes | % COHORT | Student Sub-Bodies Index (100*[Group Raw/Cohort Raw]) | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | DEAN'S LIST | SCHOLARS | COLLE-GIATES | TRUE GRIT | PRAGMA-TISTS | FULL-TIME STRUGGLE | PART-TIME STRUGGLE | VANISHERS | UNPRE-PARED | CASUALS |
| PROGRESS | 50.0 | 161 | 147 | 127 | 140 | 110 | 114 | 109 | 74 | 61 | 60 |
| CREDITS | 50.0 | 141 | 131 | 132 | 104 | 82 | 82 | 75 | 64 | 83 | 92 |
| REQURMTS | 50.0 | 137 | 126 | 138 | 119 | 114 | 102 | 88 | 108 | 56 | 64 |
| % Mainly Largo Campus Classes | 84.9 | 94 | 107 | 116 | 88 | 94 | 117 | 71 | 112 | 116 | 88 |
| % Mainly Day Classes | 53.4 | 44 | 82 | 146 | 69 | 113 | 123 | 58 | 147 | 132 | 84 |
| % Reason/4Year Transfer | 63.0 | 117 | 117 | 143 | 113 | 20 | 104 | 29 | 132 | 100 | 90 |
| % Reason/First Career | 15.6 | 94 | 73 | 47 | 73 | 139 | 116 | 108 | 92 | 142 | 121 |
| % Reason/Explore Subject | 14.7 | 41 | 56 | 6 | 66 | 327 | 71 | 281 | 12 | 80 | 143 |
| % Reason/Upgrade Job Skills | 3.6 | 72 | 106 | 33 | 189 | 131 | 61 | 306 | 17 | 67 | 64 |
| % Reason/Personal Enrichment | 3.1 | 97 | 100 | 19 | 26 | 423 | 119 | 406 | 0 | 16 | 42 |
| % Major/Transfer Program | 54.0 | 98 | 123 | 135 | 93 | 98 | 99 | 59 | 118 | 100 | 84 |
| % Major/CareerProgram | 33.8 | 123 | 88 | 62 | 122 | 114 | 124 | 128 | 85 | 109 | 79 |
| % Major/Other | 5.5 | 77 | 68 | 73 | 68 | 100 | 79 | 184 | 86 | 145 | 79 |
| % Major/None | 6.7 | 19 | 0 | 30 | 70 | 42 | 0 | 218 | 45 | 12 | 348 |
| % Arts & Sciences-*Transfer Major* | 1.9 | 113 | 67 | 231 | 134 | 149 | 39 | 0 | 188 | 43 | 68 |
| % Music/Art | 2.3 | 93 | 83 | 178 | 111 | 82 | 130 | 17 | 156 | 47 | 101 |
| % Computer Sciences | 2.4 | 107 | 185 | 97 | 177 | 118 | 124 | 82 | 99 | 68 | 54 |
| % Engineering | 2.5 | 155 | 177 | 164 | 102 | 75 | 90 | 47 | 71 | 54 | 73 |
| % Business Administration | 7.8 | 88 | 130 | 154 | 81 | 48 | 153 | 61 | 130 | 101 | 63 |
| % Education | 2.6 | 248 | 195 | 191 | 81 | 73 | 86 | 15 | 46 | 42 | 60 |
| % General Studies | 35.1 | 84 | 117 | 119 | 85 | 113 | 89 | 65 | 119 | 117 | 94 |
| % Business Studies-*Career Major* | 10.9 | 122 | 46 | 97 | 105 | 95 | 130 | 181 | 55 | 82 | 76 |
| % Computer Technologies | 4.4 | 88 | 86 | 40 | 67 | 150 | 119 | 161 | 95 | 154 | 82 |
| % Technical Trades | 4.5 | 57 | 70 | 32 | 75 | 63 | 66 | 114 | 159 | 169 | 132 |
| % Nursing | 6.1 | 113 | 125 | 72 | 167 | 186 | 122 | 129 | 68 | 89 | 38 |
| % Allied Health Technologies | 2.6 | 182 | 97 | 56 | 98 | 0 | 287 | 76 | 114 | 104 | 80 |
| % Child Care/Management | 1.6 | 80 | 40 | 55 | 53 | 118 | 0 | 221 | 186 | 152 | 49 |
| % Office Technologies | .8 | 0 | 237 | 37 | 53 | 118 | 0 | 98 | 149 | 203 | 97 |
| % Paralegal/Justice Technologies | 6.7 | 179 | 94 | 61 | 202 | 113 | 100 | 82 | 82 | 81 | 70 |
| % Misc. Traditional Trades | 1.0 | 43 | 63 | 0 | 0 | 94 | 149 | 197 | 0 | 217 | 181 |
| Average Devel Test Score | 48.3 | 11 | 107 | 110 | 96 | 102 | 86 | 97 | 109 | 82 | 104 |
| % Developmental Required | 55.6 | 58 | 72 | 65 | 120 | 97 | 165 | 120 | 66 | 180 | 60 |
| # Areas Placed/Dev. Student | 1.1 | 81 | 91 | 71 | 106 | 86 | 121 | 106 | 81 | 116 | 76 |
| % Completing Developmental | 11.5 | 108 | 110 | 155 | 239 | 25 | 201 | 110 | 26 | 0 | 65 |
| # of Major Terms Attended | 3.4 | 168 | 138 | 118 | 150 | 109 | 135 | 118 | 50 | 59 | 41 |
| % Enrolled beyond Fall 1 | 77.1 | 130 | 130 | 130 | 130 | 119 | 124 | 127 | 59 | 72 | 39 |
| % Enrolled beyond Spring 1 | 61.9 | 159 | 141 | 126 | 159 | 113 | 131 | 138 | 37 | 53 | 21 |
| % Good Start -All 3 Early Terms | 56.0 | 137 | 140 | 129 | 82 | 98 | 130 | 97 | 21 | 61 | 17 |
| % No Stopping Out (3 Terms+) | 44.0 | 122 | 141 | 145 | 71 | 110 | 130 | 90 | 21 | 67 | 16 |
| Cumulative GPA | 2.0 | 135 | 130 | 125 | 115 | 120 | 75 | 135 | 110 | 35 | 60 |
| % Course Pass Rate <.75 | 33.7 | 22 | 24 | 22 | 65 | 42 | 133 | 51 | 83 | 237 | 181 |
| % No Credit Courses Attempted | 8.3 | 0 | 0 | 0 | 0 | 0 | 63 | 29 | 0 | 441 | 147 |
| % No Credit Hours Earned | 17.4 | 0 | 0 | 0 | 27 | 0 | 56 | 20 | 0 | 327 | 255 |
| % Passing English 101 | 45.6 | 191 | 160 | 182 | 127 | 120 | 95 | 76 | 100 | 13 | 27 |





BEST COPY AVAILABLE

## Appendix B. *Methodology of Social Background Variable Construction*

The five social background dichotomies are examples of what are called in statistics *dummy* variables, in the sense that they are specifically constructed to *stand in* for other variables the full formats of which are methodologically problematic with respect to regression. The most common problem solved by dichotomization is that posed by *nominal* scales (e.g., White/African American/Hispanic/Asian-Pacific Islander/Native American/International Student), where variable categories have no non-arbitrary quantitative value or even any intrinsic ordering. These are not true numerical scales, but can be forced into a kind of numerality, one acceptable in repression analysis. This is done by singling out one category as key, lumping the remaining categories together, and assuming that the result is a two-fold variable measuring the presence (1) or absence (0) of the quality reflected in the key variable; since [0,1] counts in mathematics as a complete number scale, albeit the shortest possible one, variables in this form may be used in quantitative linear analysis.

Traditionally, dummy versions of variables are re-named to reflect the key quality presence being measured; our *White*$^+$ indicator is a good example of a dummy solution for the nominal scale problem presented by the original six-category Race/Ethnicity variable, named to highlight the *white*/non-white distinction. Two others are *Female*, originally Gender: Female/Male, and Single, originally Marital Status: Now Married and Living with Spouse/Other, except that these, as natural dichotomies, required no category collapsing, only assignment of 1 and 0 values. (The plus in the *White*$^+$ label is an acknowledgment that in this particular case the whiteness of the key category is not quite pure; two underpopulated minority categories Asian/U.S. Resident and International Student actually were assigned to the large white category, on the grounds that prior cross-tabulation proved student achiever classification rates in these groups nearly identical to that of students in the white category, while the rates of the large African American and Hispanic groups were distinctively lower. It seemed to us better to load the dice somewhat towards a higher race-to-achievement correlation than to effectively mask the genuine achievement bias of two small racial groups.)

Two other common problems solved by dummy variables are population skew and curvilinear correlation. In the population skew instance, a variable in good numerical scale format (e.g., years between high school and college) may have so many cases crowded at one extreme end of the scale that the true impact of an independent variable on a dependent variable may be under-represented by association statistics simply because the real effects of the other parts of the scale are obscured due to their underpopulation. The student distribution over our original Years to College Enrollment scale was so slight beyond the within one year range, that we thought it best to create the *Immediate Entry* dummy variable, dichotomizing at the one-year-or-more point.

In the curvilinear correlation instance, it is the shape of an independent variable's causal *relationship* with the dependent variable which causes difficulties e.g., its impact is not steady over the length of its scale but systematically varies. We discovered an example of this when we ran a pre-regression scattergram of our original student age at admission indicator with percent classifying as achievers. What we found was a trend-line of relatively high achievement rates at the 17-20 years old beginning of the age scale, then a sharp drop-off over the 21-24 age interval, and then a leveling off to a gradual decline over the 25-89 range. Clearly, the age variable's association with achievement classification was mostly a phenomenon of youngest students versus all others, and the use of the full-range age interval would obscure its strength by including a long range of relatively uncorrelating values (25-89 year). Therefore, we decided to create a dummy version (*Young*) as a more fairly indicative replacement, by cutting the age variable at 21 years old and collapsing the resulting two halves of the scale into two categories.

The *High School Quality* variable, on the other hand, was a derived scale based on an entirely different sort of construction methodology, created for reasons having nothing to do with the regression-fittedness of the original data. Its creation was theory-driven, an attempt by indirect means to fill an important gap in our direct measurement coverage of the background phenomena relevant to college achievement: the quality of a student's pre-collegiate educational opportunities and experiences. The possibility struck us of using data available on name of high school attended by each student to infer level of prior educational advantage level by assuming a rough but good link between advantage level and quality of specific high school environment, provided that we found some defensible technique for ranking high school educational environments.

We decided on the expert panel approach to accomplish the latter, asking a small group of county resident staff known to be knowledgeable about secondary education in the county to rate the educational reputations of each high school as either excellent, good or less than good. We then rank-ordered county secondary schools according to the mean panel ratings of each, chose reasonable cutting-points, and created a four-category ordinal scale: 4=County Private High Schools (consistently rated the highest)/3=High Reputation County Public High Schools/2=Other County Public High Schools/1=Non-County High School, G.E.D. or No Diploma. The last category, of course, was strictly speaking off the scale, since the panel was not ask to rate non-county high schools nor the G.E.D. program, and the No Diploma category by definition precluded assessment. Our justification for the inclusion of this low-end category was basically a pragmatic one: As an empirical matter, the group of students who came out of the three non-rated educational circumstances (a small minority in the first place) exhibited an achievement rate significantly lower than any of the rated groups, and when we put these with the rated groups as the bottom category of a now four-step educational quality scale, the achievement rates formed a neat lowest to highest trend-line.

U.S. Department of Education
Office of Educational Research and Improvement (OERI)
Educational Resources Information Center (ERIC)

ERIC®

JC 980 044

# REPRODUCTION RELEASE

(Specific Document)

## I. DOCUMENT IDENTIFICATION:

| | |
|---|---|
| Title: Toward a Model of the Academic Process: Summarizing Cohort 1990 Progress and Achievement | |
| Author(s): Karl Boughan | |
| Corporate Source: | Publication Date: March 1997 |

## II. REPRODUCTION RELEASE:

In order to disseminate as widely as possible timely and significant materials of interest to the educational community, documents announced in the monthly abstract journal of the ERIC system, *Resources in Education* (RIE), are usually made available to users in microfiche, reproduced paper copy, and electronic/optical media, and sold through the ERIC Document Reproduction Service (EDRS) or other ERIC vendors. Credit is given to the source of each document, and, if reproduction release is granted, one of the following notices is affixed to the document.

If permission is granted to reproduce and disseminate the identified document, please CHECK ONE of the following two options and sign at the bottom of the page.

| | The sample sticker shown below will be affixed to all Level 1 documents | The sample sticker shown below will be affixed to all Level 2 documents | |
|---|---|---|---|
| [X] **Check here For Level 1 Release:** Permitting reproduction in microfiche (4" x 6" film) or other ERIC archival media (e.g., electronic or optical) *and* paper copy. | PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY ______ Sample ______ TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) | PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN OTHER THAN PAPER COPY HAS BEEN GRANTED BY ______ Sample ______ TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) | [ ] **Check here For Level 2 Release:** Permitting reproduction in microfiche (4" x 6" film) or other ERIC archival media (e.g., electronic or optical), but *not* in paper copy. |
| | Level 1 | Level 2 | |

Documents will be processed as indicated provided reproduction quality permits. If permission to reproduce is granted, but neither box is checked, documents will be processed at Level 1.

*"I hereby grant to the Educational Resources Information Center (ERIC) nonexclusive permission to reproduce and disseminate this document as indicated above. Reproduction from the ERIC microfiche or electronic/optical media by persons other than ERIC employees and its system contractors requires permission from the copyright holder. Exception is made for non-profit reproduction by libraries and other service agencies to satisfy information needs of educators in response to discrete inquiries."*

**Sign here→ please**

| | | |
|---|---|---|
| Signature: | Printed Name/Position/Title: Karl Boughan, Supervisor of Institutional Research | |
| Organization/Address: Prince George's Community College, 301 Largo Road K-231, Largo MD 20774 | Telephone: 301-322-0722 | FAX: 301-808-0960 |
| | E-Mail Address: Kb2@pgstumail.pg.cc.md.us | Date: January 15, 1998 |

## III. DOCUMENT AVAILABILITY INFORMATION (FROM NON-ERIC SOURCE):

If permission to reproduce is not granted to ERIC, *or*, if you wish ERIC to cite the availability of the document from another source, please provide the following information regarding the availability of the document. (ERIC will not announce a document unless it is publicly available, and a dependable source can be specified. Contributors should also be aware that ERIC selection criteria are significantly more stringent for documents that cannot be made available through EDRS.)

| |
|---|
| Publisher/Distributor: |
| Address: |
| Price: |

## IV. REFERRAL OF ERIC TO COPYRIGHT/REPRODUCTION RIGHTS HOLDER:

If the right to grant reproduction release is held by someone other than the addressee, please provide the appropriate name and address:

| |
|---|
| Name: |
| Address: |

## V. WHERE TO SEND THIS FORM:

| | |
|---|---|
| Send this form to the following ERIC Clearinghouse: | Jonathan Kelly<br>ERIC Clearinghouse for Community Colleges<br>3051 Moore Hall<br>Box 951521<br>Los Angeles, CA 90095-1521 |

However, if solicited by the ERIC Facility, or if making an unsolicited contribution to ERIC, return this form (and the document being contributed) to:

Institutional Research & Planning
October 1997