Chapter Two:

METHODOLOGY

Concept of The Study

Using the most current and relevant information from available data, the study will
provide definitive answers to each of the research questions. The approach taken consists of an
ex-post-facto descriptive study employing a variety of descriptive statistical procedures and a
Logistic (LOGIT) multiple regression analysis model. Logistic regression is based on the
assumption that the data follow a binomial distribution. Since all end of course grades in this
study result in one of two possible outcomes, pass or fail, the binomial assumption is met making
LOGIT an appropriate model for the regression. Under this procedure, the dichotomous
dependent variable, PASS/FAIL, will be used as the criterion variable in place of a numerical
grade. The regression analysis will determine the extent to which the independent or predictor
variables collectively explain variance in end of course pass/fail outcomes. Although the results
will not establish cause and effect relationships among the variables, they will present an accurate
depiction of the extent to which they explain variance in end of course outcomes. If significant
predictor variables are found to exist in the model, the resulting predictive equation might then be
used very effectively to improve college course screening and to predict probability of pass/fail
outcomes for Sailors enrolled in PACE.

Assumptions:

Given the ex-post-facto nature of the study, it is assumed that the course completion data
recorded by both contractors, and the demographic data recorded by the Navy are correct as
delivered. Maximum and minimum frequency ranges will be examined to confirm that the data
conform to reasonable limits and ranges. Also, it is assumed the data follow a binomial
distribution due to the dichotomous nature (PASS/FAIL) of possible course outcomes.

Population Sampling:

Each of the contractor firms who administer college level PACE courses for the Navy maintain an electronic database containing student enrollment and course completion record data. These records include information for all Sailors who completed PACE courses from 1992 through mid 1996.   In addition the Navy maintains an electronic Enlisted Master File (EMF) containing a wide range of demographic, educational, and career history information for every Sailor. Using a cross sectional approach to sampling the entire population of Sailors enrolled in PACE between 1992 and 1996, all end of course record data Navy wide will be collected for approximately a ten month period beginning in July, 1995 and ending in May, 1996. This time period corresponds to four terms of instruction for courses taught by CTC and two semesters, Fall 1995 and Spring, 1996, taught by MRC. Preliminary estimates are that the sample size for this cohort is approximately 8,000 enlisted records.

Course completion record data from Fall, 1995 through Spring, 1996 were selected as a representative cohort sample using the following objective criteria:

1. Large, representative, but reasonably manageable sample size. This cross section comprises nearly one fourth of all the Sailors in the population which for purposes of this study is defined as those who have taken PACE college level courses throughout the Navy since data were electronically recorded beginning in 1992 by both contractors. A larger sample size would make it impractical to enter and manipulate the data by exceeding the time and resources available to this researcher in conducting the study.

2. Currency and diversity of the sample. This cross section is the most recent portion of the population, and contains the largest number of course completion record data for any other period. Likewise, it also represents the most numerous and diverse mix of ships and ship types in which PACE courses have been offered since the data were first electronically recorded in 1992.

3. Consistency in administrative procedures and record keeping across the two contracts. Data from this cohort are the most stable and consistent in terms or record keeping between the two contractors. In July, 1995, the Navy awarded its current contract to MRC. As prime contractor, MRC was tasked for the first time with overall responsibility for contract administration, including both technology and instructor delivered courses. Prior to 1995, the Navy contracted with CTC and MRC separately, resulting in separate administrative and record keeping systems.

4. Most current hardware and courseware technology. Hardware and courseware used by MRC have been continuously upgraded as improved technology became available. By 1995, all equipment in use had been upgraded to high speed microprocessors (386MHZ or better) and interactive courseware was widely used in PACE technology delivered courses. After May 1996, some courses taught by MRC were significantly upgraded with enhanced interactive capability. Because of this radical new software design, results from these courses would not be appropriate for inclusion in this study since most of the population would not have been exposed to the newest design.

Specification of Variables:

The variables of interest in the present study fall into two categories: those for descriptive analyses and those for use in LOGIT regression. As discussed earlier, many are suitable for both descriptive and regression analysis. Descriptive variables include course grade (GRADE), gender (SEX), age (AGE), marital status (MS), race/ethnicity (RACETH), mental ability (ASVAB), educational level (EDU), semester hours of recent college courses passed (RECENTCO), years of active service in the Navy (YOS), paygrade (PGDE), Navy occupational rating (RATING), ship in which the course was taken (SHIP), ship satisfactory college course completion rate (COMPRATE), ship type (TYPE), homeport (HP), coast on which homeported (COAST), number of attempts (ATTEMPT), and delivery mode (MODE). LOGIT regression variables

include course outcome (PASS/FAIL), age (AGE), gender (SEX), race/ethnicity (RACETH), mental ability (ASVAB), marital status (MS), years of service (YOS), paygrade (PGDE), educational level (EDU), semester hours of recent college courses passed (RECENTCO), ship satisfactory college course completion rate (COMPRATE), ship type (TYPE), coast on which homeported (COAST), and delivery mode (MODE). Table 8 identifies each of these variables as continuous or categorical and the ranges/types of data expected to be found for each. Other key variables that are well documented in the literature of educational achievement research were considered for inclusion in the study but could not be used given the limits of the available data and the ex-post-facto nature of the study. These included high school grade point average and class standing, teacher effects, socioeconomic status, motivation, and time spent on homework.

Exclusions and Grouping of Data

It was decided to keep nearly all of the data from the cohort period in the study, including records of multiple courses taken by individual Sailors, in order to obtain the most complete and accurate analyses for external validity. The only exceptions to this will be records of more than two attempts for the same course, and less than 100 records of courses for Sailors in technology delivered courses who were permitted to extend their course into the follow on semester. Records indicating a second attempt will be retained in order to measure any difference in grade between Sailors who took a course only once and those who took the same course twice. Records of more than two attempts to take the same course, if any, would serve no useful purpose for the study and will be dropped. Likewise, records of technology delivered courses in which Sailors were permitted to extend their study time into a subsequent semester will be dropped since additional time on task might skew the regression analysis.

To improve the accuracy of the regression analysis and reduce error in prediction, all course records remaining in the cohort will be categorized into six course groupings with separate regression analyses run on each group. A review of the course titles and frequencies for each course found that there are six broad categories of college level courses offered by both CTC and

MRC. These groupings are Business, English, History, Math, Social Science, and other. Table 9 contains a detailed listing of the individual courses and the enrollment frequencies observed for each course within the cohort selected for this study.

Data Analysis:

Course completion record data obtained from the contractors, along with demographic data from the Enlisted Master File will first be analyzed for completeness and accuracy. Decisions to retain or discard incomplete or obviously incorrect data will be made on a case basis after every effort has been made to obtain correct and complete data. Any data discarded will be noted with an explanation as to why it was discarded. With regard to the data collected for end of course grade, it will be necessary to convert letter grades to a single interval scale for comparison purposes. The consortium of universities and colleges under contract with MRC assigns grades using plus and minus points with letter grades, whereas CTC uses only whole letter grades. Accordingly, an interval scale of 0.0 through 4.0 will be established with 0.0 corresponding to a grade of "F" ascending to 4.0 corresponding to a grade of "A". In order to account for the practice by CTC and MRC discussed earlier of assigning a grade of "W" in place of an "F" for Sailors who do not complete once the course has begun, both will be treated as an "F" with a value of 0.0 assigned for data analysis purposes. Course record data from CTC and MRC will then be merged by social security number with demographic and educational data from the Navy Enlisted Master File (EMF). Data thus compiled will include both continuous and categorical variables for each Sailor in the cohort having completed a college level PACE course.

The following procedures and tests will be performed on the data:

1. Data will be entered into a computer based statistical program (SPSS v.7.51) for both descriptive and regression analyses.

2. To find answers to research questions 1-7, descriptive statistical analyses to include frequency distributions, calculation of means, standard deviations, correlations and contingency tables, and significance testing of the means between variables and

among groups and among the correlations will be performed, analyzed, and reported. A .05 significance level will be established for hypothesis testing.

3. To find answers to research question 8 pertaining to main and interaction effects, a two factor 3x2 design analysis of covariance (ANCOVA) as depicted in Table 10 will be performed. Taking ASVAB scores into account, the ANCOVA will determine the extent to which end of course grades differ for each delivery mode by type of ship. An assumption of ANCOVA is that grade data are normally distributed. Residual error testing will be performed to determine the extent to which the data are normally distributed. In the event grade data are not normally distributed, the data will be transformed using the most appropriate function based on analysis of residual testing. Main and interaction effects will then be determined through ANCOVA. There is no need to include submarines in this analysis because there is only one PACE course delivery mode available in submarines as noted earlier. A diagram of the ANCOVA will be provided along with a full interpretation of results.

4. Prior to conducting LOGIT regression analysis (research question 9), a mean numerical grade for each Sailor who took more than one course within any course group will be calculated and converted to pass/fail for regression analysis. A value of 1 will be assigned to PASS outcomes, and a value of 0 will be assigned to FAIL outcomes. This will produce a single PASS/FAIL outcome for every observation in the regression thereby eliminating the potential bias which might result from regressing multiple grades for the same individual.

5. The data will then be sorted by course group for regression analysis as discussed earlier.

6. Next, multicollinearity diagnostics will be performed on each data set in order to determine the extent to which multicollinearity may be present among the predictor

variables. Tests for variance inflation factors and variance components will be performed to determine the presence of linear dependence, or shared variance. If present, the diagnostic tests will identify and measure the extent of linear dependency. Variables so identified may then be combined as a single predictor variable or eliminated to reduce multicollinearity in the model. The objective of this procedure is to produce a series of predictor variables which are as independent of one another as possible and improve prediction accuracy in the regression model.

7. Stepwise LOGIT regressions will then be conducted for each course group.

8. At this point, to determine the amount of variance in course outcomes explained by the predictor variables, a measurement of percent of observations correctly classified will be calculated using a value of .5 probability as the classification criteria for PASS. All such findings will be reported and explained fully.

9. Based on analysis of the percent of observations correctly classified by the regression model, recommendations will be formulated regarding the potential usefulness of the model for predicting probability of pass/fail outcomes for screening of Sailors seeking to enroll in future PACE college courses.

Results of these tests will be reported and thoroughly discussed. Implications as to their potential usefulness for improving PACE effectiveness and efficiency will also be presented, along with recommendations for further research where appropriate.