

Chapter One

INTRODUCTION

Background

As part of the Navy's Voluntary Education (VOLED) programs, the Program For Afloat College Education (PACE) offers both undergraduate courses from accredited colleges and universities as well as academic skill development courses to naval personnel aboard ships anywhere in the world. As originally designed in the mid-1970's, PACE consisted of courses taught at sea by college professors under contract. As technology improved, the Navy expanded PACE to include stand-alone, self-paced course offerings using computer based instruction (CBI) delivery methods. At first technology delivered courses were offered only in submarines where it was impractical due to space constraints to provide a professor. In the mid-1990's, technology delivered courses were expanded and are currently available in most ships throughout the Navy. Likewise, instructor taught courses are now widely available in all but the smallest of ships.

Need for the study

Since its inception, PACE has accounted for tens of thousands of courses taken by naval personnel throughout the world. The program has grown into a multimillion dollar contractual agreement between the Navy, its prime and subcontractors, and a consortium of six major colleges and universities in the United States. In an empirically based study of the Department of Defense (DOD) Tuition Assistance Program (Boesel and Johnson, 1988), it was shown that VOLED programs can have a strong, positive effect on both promotion and retention of service men and women. Nevertheless, with budget outlays for DOD shrinking rapidly in the past few years, quality of life programs such as PACE in all the services are being carefully reviewed to ensure they are both effective and necessary. From this review will inevitably come a series of policy decisions that will determine the future direction for PACE in terms of funding, scope, and economies of scale. The present study is intended for use by the Director of the Navy's VOLED

programs office as one means of assessing PACE program effectiveness in this era of declining resources.

Purpose of the study

In a cost-effectiveness study of PACE, Dunlap, Best, Green, and Knight (1989) determined that the Navy's investment in PACE was economically sound. Further, the study found that the program met the objectives of DOD and the Navy Department of improving mission performance and preparing naval personnel for positions of greater responsibility. More specific findings of the only comprehensive PACE study conducted to date were that “the Armed Services Vocational Aptitude Battery (ASVAB) composite arithmetic reasoning and paragraph comprehension score (AR + PC) is a good indicator of performance” (Dunlap et al., 1989, p. 32). Likewise, the study concluded that “individuals with a composite ASVAB (AR + PC) score of less than 100 are not likely to successfully complete (PACE) courses” (p. 8); and that “computer-interactive video technology is the most effective of the micro-computer based delivery modes in terms of successful course completions” (p. 8).

Based upon these conclusions the Navy expanded PACE as noted earlier, to include the acquisition and development of computer-interactive video hardware and courseware to equip every ship in which PACE was offered. In addition the Navy established a more rigid screening process for its PACE applicants, requiring them to have achieved a score of 100 on the composite ASVAB (AR+ PC), and began to offer academic skills courses (pre-college) in areas such as reading, writing, and mathematics for those deficient in these prerequisite skills.

In order to better determine how and where to allocate scarce resources within its VOLED programs, the Navy has asked for a follow-on study to assess the current effectiveness of PACE in light of the many program changes that have been implemented since 1989. Specific questions of interest to the Navy about its PACE program are:

1. Does participation in PACE courses influence reenlistment decisions and if so, to what extent?

2. Does participation in PACE courses influence enlisted promotions, and if so to what extent?
3. What elements of the PACE program contribute most to successful outcomes for Sailors enrolled in college level undergraduate courses?

While the present study will focus primarily on finding answers to the third question, the review of literature will address relevant theories and prior research into all three of these questions.

LITERATURE REVIEW

VOLED: Reenlistment And Advancement

In assessing the extent to which participation in PACE influences reenlistment decisions and promotion rates for Sailors, it is necessary to determine what, if any, conclusions have been reached in previous studies pertaining to these questions. A secondary avenue of inquiry is to determine what evidence may exist for expecting to find a relationship between participation in PACE courses and retention and promotion in the Navy.

In the earliest comprehensive report found in the literature of off-duty VOLED programs in the Navy, Gitchens and Wilcove (1977) conducted a series of 12 studies to assess the relationship between VOLED and recruiting, performance, and retention. In addition to PACE, the studies examined two additional VOLED programs: Tuition Assistance and the Contract for Degree. In one of these descriptive studies, 464 PACE participants between 1974 and 1975 were compared to 464 matched non-participants to determine any difference in their advancement and reenlistment rates (p. 12). Controls were established between groups to ensure that those in the cohort had equal opportunity for advancement and reenlistment during the period under study. Non-VOLED participants were carefully matched with participants to include consideration of pay grade, time remaining in current enlistment, and eligibility for promotion prior to expiration of current enlistment (p. 14). In addition, pairs were matched by ability, educational level, occupational specialty, marital status, years of service, race, and gender (p. 14).

Results obtained showed that PACE participants were *not* as likely to reenlist as non-participants and that there was no difference in promotion rates. Promotion rates for PACE

participants (51% promoted, 49% not promoted) were statistically no different at the .05 level from non-participants (53% promoted, 47% not promoted) (p. 15). As for reenlistment, PACE participants left the Navy in significantly greater numbers (63%) than non-participants (54%) (p. 15). Compared to PACE, participants in Tuition Assistance and Contract For Degree programs were more likely to be promoted than matched non-participants, and like PACE participants, those enrolled in Tuition Assistance were less likely to reenlist (p. 15). A complete summary of the results of this comparison study is found in Table 1.

In speculating as to the cause of lower reenlistment rates for VOLED participants, Gitchens and Wilcove suggested that “personnel who know they will be leaving the Navy participate in these programs to better prepare themselves for a civilian vocation” (p. 14). This observation is reasonable in view of the historically high separation (discharge) rates which normally occur among service members at the end of their first or second enlistment terms. Also, surveys of new recruits have consistently shown that men and women who join the service rank educational opportunity as an important reason for enlisting. For example, Mallison (1994) reported that recruits ranked continuing education as their top motive for joining the Navy. Also, in a worldwide survey of naval personnel, Wilcove (1992) found that 30% of enlisted and 22% of officers reported they had taken some form of VOLED sponsored by the Navy. With such large interest and participation in VOLED throughout the Navy, it would not be surprising to find a correlation, in this case negative, between VOLED and promotion and retention.

In another major study of VOLED programs in the military, Boesel and Johnsen (1988) examined the Department of Defense (DOD) Tuition Assistance (TA) program in the Army, Air Force and Navy. This study produced a strikingly different picture of reenlistment rates for VOLED participants than was found by Gitchens and Wilcove eleven years earlier. From a randomly administered DOD wide survey in 1985, a stratified sampling of service men and women responding to the survey was selected as subjects for this study. Records of those sampled who had participated in TA were then merged with other databases to provide

information on promotion and reenlistment activity. The resulting sample size was 10,718 service personnel with Air Force comprising more than half of these.

Boesel and Johnson compared promotion and reenlistment rates for these service members with non-TA users from the 1985 DOD survey and found that separation rates for non-TA users (35.8%) were twice as high as TA users (18.6%). They also found that promotion rates for TA users (53.1%) were significantly higher than for non-TA users (39.1%) (p. 38-39). Next they used LOGIT and Ordinary Least Squares (OLS) regression analysis procedures to determine whether these findings were valid when other factors were taken into account. Promotion category and reenlistment category were regressed on demographic, educational, and career variables for both TA and non-TA users from the 1985 survey. These additional analyses indicated that TA participation was strongly and positively related to both reenlistment and promotion for enlisted service members even when all other variables were statistically controlled (pp. 41, 43). The authors concluded that:

Regardless of the causal relations between TA participation, personal commitment, and retention, it is very clear that TA users are less likely than non-users to leave the military in a given period of time. The evidence directly contradicts the contention sometimes heard that the effect of participation in TA is to hasten the departure of military members for jobs in the private sector (p. 45).

These conclusions must be interpreted cautiously with regard to PACE participants because PACE was not taken into account directly in this study. Nevertheless, since PACE, like TA, is a major element of the Navy's VOLED programs, this study seems to suggest that there might also be a correlation between PACE participation and reenlistment and retention. The next study reviewed provides limited additional insight into this question.

In their cost effectiveness study of PACE Dunlap et al. (1989) also examined the extent to which PACE (1) prepared Sailors for positions of greater responsibility or promotions in the Navy, and (2) contributed to retention (reenlistment). Unlike the more comprehensive TA study just

reviewed, this study used no actual performance or retention data. Rather, these assessments were made using survey data from a questionnaire that was distributed to PACE users on both the East and West coasts. The study does not describe the process by which surveys were distributed so random selection cannot be assumed. Of 550 surveys distributed, 151 were returned for a 27.5% return rate (p. 20). Among the questions in the survey, respondents were asked to check whether they agreed, disagreed, or were neutral to statements asking whether the PACE program benefited them by improving chances for promotion and in providing them an incentive to reenlist.

In responding to the promotion question, 51% agreed, 13% disagreed, and 45% were neutral (p. 26). To the reenlistment question, 14% agreed, 37% disagreed, and 49% were neutral (pp. 26,28). Further, it was found that of those who said they planned to reenlist (42%), 22% of those agreed that participation in PACE was an incentive to reenlist and 10% reported that PACE was an important factor in that decision (p. 28). Finally, when asked to rank order the most important benefits of PACE, the respondents' first choice was "increased opportunities for jobs", followed by "education upon completion of their Navy career" (p. 26). Dunlap et al. concluded that: "PACE appears to have a small but significant [positive] impact on retention" (p. 28) based on the ten percent who reported PACE as an important factor in their decision to reenlist.

These results are speculative at best and must be interpreted with great caution because of the apparent lack of random generation. Also, with so few respondents and some 45 survey questions to answer, the ratio of respondents to questions was only 3:1 which is far below the commonly accepted standard of at least ten respondents for every variable in the survey.

In summary, since 1989, there have been no additional studies to shed new light on the questions of what influence, if any, PACE has on retention and advancement. The literature provides outdated, contradictory, incomplete, and unreliable data at best. As of this writing the Center for Naval Analyses has been tasked to conduct a thorough assessment of the relationships

between PACE participation and reenlistment and advancement. The literature review will now turn to the question of paramount interest to this researcher: determining those elements of the PACE program that contribute most to successful outcomes.

Comparing Instructional Delivery Modes - Theory and Practice

Because PACE has a dichotomous delivery system, one of the first questions raised by the Director of the Navy's VOLED programs was "Do Sailors learn better with technology or instructor delivered courses?" (Kelly, F., personal communications, December, 1995). In reviewing the literature, the first area to be examined is the efficacy of conducting a media comparison study in response to this question.

For much of the twentieth century, educators and behavioral scientists have been interested in comparing instructional effectiveness between different types of delivery systems, or media. At first glance, it would seem a logical inquiry in view of the changes in instructional media over the years, particularly with regard to computer mediated instruction. While there are many different types of media comparison studies in the literature, the most common approach taken has been to compare traditional or teacher-mediated learning with technology-based devices as either a substitute for or supplement to the teacher. In the United States prior to World War II film and radio were the focus of many early comparison studies (Moore, D. M., personal communications, Spring, 1997). With the war and the subsequent race for space between the two superpowers, extensive government funding for research and development fueled a renewed interest in media comparison studies. About this time researchers began to assess not just the equipment but its capabilities or attributes in improving instructional effectiveness, (Schramm, 1977). The focus on specific media and its attributes has continued with improvements in technology. In recent years the emphasis has moved from television and programmed instruction to personal computers and distance education as the media of choice for comparison studies.

Maddux (1995) described three stages of modern comparison research into the efficacy of computers in education. Prior to 1980, there were more descriptive studies and papers written

than experimental studies comparing computer-delivered instruction with traditional delivery modes. According to Maddux (1995), this trend shifted in the 1980's. Researchers and educational software developers became interested in establishing proof of cause and effect relationships between computer and non-computer delivery modes. Many turned to scientific experimentation as the study method of choice. However, as will be seen shortly, many of these studies suffered from problems of internal and external validity due to specification and design error. "The failure to control for possible interactions among teaching and learning variables, and the failure to adequately describe control group activities made it difficult or impossible to interpret many stage two activities" (Maddux, p.8). In the 1990's, Maddux observed a shift toward improvement in the internal design of computer-delivery experimental comparison studies which he attributed to better controlled laboratory research. However, he opined there still remained a major flaw in much of the research due to its limited generalizability beyond the confines of the laboratory (p. 8).

Throughout these developmental stages of media comparison research, there has been an intense debate in the literature as to whether media alone influence performance outcomes in learning. In the lead of those who believe that media do not and will never influence learning has been Dr. Richard E. Clark, Professor of Educational Psychology at the University of Southern California. Clark has clearly articulated his views in numerous papers and journal articles. His main arguments have been predicated on learning theory espoused by leading educational theorists such as Gagne, Briggs, and Wager (1992) which have been widely accepted in the literature for many years.

Principally Clark has argued that it is not media per se that influence learning. Rather, "learning is caused by the instructional methods embedded in the media presentation" (Clark, 1994b, p. 26). As cited by Clark (1994b, p. 23), Saloman defined instructional method as "any way to shape information that activates, supplants, or compensates for the cognitive processes necessary for achievement or motivation". Gagne et al. (1992) also specified those events which

are necessary to learning as comprising both internal, or cognitive information processing events such as temporary storage, encoding, and retrieval; and external instructional events such as gaining attention, drill and practice, and feedback which are necessary to activate and support the internal processes (p. 202). Gagne et al.(1992) treat media not as an external instructional event, but rather as a “vehicle for the communications and stimulation that make up instruction” (p.205). Since media themselves are merely the conveyors of instructional methods and content, Clark has contended that they do not directly influence learning in any way. Furthermore, his replaceability challenge to his critics got right to the heart of his main argument: “We need to ask whether there are other media or another set of media attributes that would yield similar learning gains. If a treatment can be replaced by another treatment with similar results, the cause of the results is in some shared (and uncontrolled) properties of both treatments” (Clark, 1994b, p. 22).

Opposing Clark is Dr. Robert Kozma, director for the Center For Technology and Learning, SRI International. His main argument has been that media and methods are inextricably interconnected. According to Kozma, both are part of the instructional design. “Media must be designed to give us powerful new methods, and our methods must take appropriate advantage of media’s capabilities” (1994, p. 16). Further, Kozma has maintained that learning from media can be thought of as a complementary process within which representations are constructed and procedures performed, sometimes by the learner and sometimes by the medium (p. 11). By means of its technological capabilities, symbol system, and processing capabilities, Kozma argues that “a particular medium can be described in terms of its capability to present certain operations in interaction with learners who are similarly engaged” (p. 11).

In the view of this researcher, Clark has made a stronger case in the literature than has Kozma in defending his position. First, his arguments that media are mere vehicles for instructional methods are in agreement with learning theory as it is presently hypothesized and widely accepted. Until that theory is revised to include media as an external instructional event, or as an instructional method, the two must be treated in research as separate yet complementary

components of an instructional process. Even more appealing to this researcher is Clark's argument that many very different media attributes (can) accomplish the same learning goal (Clark, 1994b). Until such time as circumstances arise that any single medium becomes *necessary* for learning, which is not likely, media will continue to serve only as a conveyor of the processes which contribute to learning. This is not to say that some media are not vastly superior to others as conveyors of methods and content. Technology has made great strides and is rapidly outpacing other modes of delivery in terms of cost and efficiency as will be seen in the reviews of research that follow. Nevertheless, as long as there is another medium capable of conveying similar methods of instruction in any given instructional setting resulting in similar learning outcomes, then Clark's argument that the cause of the learning must be something other than the media itself, is compelling indeed. Kozma's retort (1994) to this position is that if two treatments yield a similar outcome, it does not mean that they resulted from the same cause. Again, Clark is not claiming that there necessarily is a single causal factor for learning to occur with different media. He is simply arguing that as long as similar learning occurs with different media that there must be some cause(s) other than the media itself.

In practice, such a fundamental disagreement in the literature, as well as a lack of understanding of the key issues in the debate, have for years left researchers confused and perhaps ill-informed regarding the efficacy of conducting media comparison studies. This, along with rising numbers of educational software companies and renewed government interest in education within the past 15 years has resulted in a plethora of such studies. Indeed they are so numerous that meta-analyses of media comparison studies have become common in the literature, including many dissertations of this type. In view of the Navy's initial interest in comparing delivery modes for its PACE courses, a review of the major findings of such studies, including meta-analyses, in both military and non-military settings is warranted. At the same time, a close look at the issues of research design quality, results, and interpretations of these results is also warranted in light of the theoretical and still controversial arguments just cited. Following a

review of experimental media-comparison research, a similar sampling of non-experimental research will be conducted in both settings.

Media Comparison Studies: Experimental Research

In analyzing the meta-analyses of a 30% sampling of 500 experimental studies conducted in the 1970's, Clark (1985) found numerous examples of achievement gains for technology-delivered instruction which he described as "overestimated due to often uncontrolled for but robust instructional treatments embedded in computer-based-instruction (CBI) treatments" (p.249). Specifically he noted that many of the instructional methods built into CBI such as corrective feedback and individual pacing were not used by teachers in comparison treatments (p. 250). He described this phenomenon as "confounding", a term which has permeated the literature ever since. In essence, Clark determined that "in confounded research, causes and effects cannot be unambiguously identified. When causes are confounded, it is still possible to determine the size of the effect of the treatment on the outcome measure. However it is not possible to determine which of many uncontrolled - yet correlated - parts of the treatment contributed to the differences" (p. 252). Overall, Clark concluded that 75% of the studies he examined had serious design flaws, fifty percent failed to control for time on task, and only half controlled for instructional methods. In only two of fifteen studies in which instructional methods were controlled were there significant differences in favor of CBI (Clark, 1985, p. 259).

In a more recent meta-analyses of 12 studies involving computer-aided-instruction (CAI) in Adult Basic and Secondary Education between 1984 and 1992, Rachal (1993) found similar design flaws to include lack of specification of the size of control and experimental groups, and time on task. Also, he reported numbers of subjects were small and treatment periods too short, and the abilities of the individual instructors assigned to teach the control groups were never assessed. He concluded that while results in at least one well controlled study found significant differences favoring CAI over non-CAI groups, expectations of CAI becoming a miracle cure for adults in these educational programs are naïve and unrealistic (p. 172).

In a paper dealing with problems of Learner Control Research, Reeves (1993) held that “much of the research in the field of CBI is pseudoscience because it fails to live up to the theoretical, definitional, methodological, and/or analytic demands of the paradigm upon which it was based” (p.39). He recommended that graduate students and researchers “develop an improved understanding of contemporary philosophy of science, a topic largely ignored in many research methodology and statistics courses (p. 43). In this way, researchers would be exposed to a wider range of options in conducting experimental and non-experimental designs.

In a report on effectiveness of technology in schools (Sivin-Kachala & Bialo, 1994) the authors summarized the findings and conclusions of 133 media comparison studies, 58 of which were published and 24 were doctoral dissertations. Sponsored by the Software Publishers Association, the authors boasted that “educational technology has demonstrated a significant positive effect on achievement” (p. 2). As with several reviews of multiple studies in the literature, the authors failed to report, if indeed they examined, the research methods in these studies for adequacy. This researcher examined at random six of the studies cited by Sivin-Kachala & Bialo and found significant design flaws in at least five of them as shown in Table 2. A brief summary of the findings for each of the six studies so examined follows.

In a study of CAI effects of math achievement on black students seeking admission to teacher education programs in the U. S., Reglin (1985) found significant differences in CAI versus traditional instruction. However, he failed to control for prior knowledge, ability, learning style, instructor effects, learner familiarity with technology, and method of instruction. Based upon lack of adequate controls, there simply is no way to verify that the differences he found in pre-and-post tests were the result of the media used in the experimental group.

Clements (1991) studied enhancement of creativity in computer environments using LOGO software versus a non-LOGO CAI user group. While he too found a significant difference in pre-and-post test scores favoring those using LOGO, he failed to adequately explain any differences in instructional methods other than one group used LOGO and the control group did

not. Also, he failed to report controls for ability, learner style, and learner familiarity with the technology. Also, students in the control group had less time on task than those using LOGO.

Skinner examined the effects of CBI on achievement of college students by using a single subject alternating treatment design where 36 college students were rotated among three instructional delivery treatments: Text/CBI/Guided Tutorial, Text/CBI/Solo, and Text-Only. Skinner also used a matched pair procedure for assignment of subjects into two groups to control for possible confounding effects of differing unit difficulty levels. He also controlled for ability by dividing students into levels of achievement based on previous course performance. Results indicated that achievement in test scores was significantly better for the modules in which students used the CBI delivery modes. Performance averages for Text-Only were lower. Although he controlled quite effectively for individual effects across all three delivery modes, he failed to control for instructional method since the Text-Only modules probably lacked many of the methods employed by CBI such as drill and practice, feedback, etc. Also, time on task varied, particularly in the SOLO delivery mode since learners were able to do as little or as much as they chose.

In an excellently designed experimental dissertation, Bair (1990) found a significant difference in the effects of word processing on writing output as measured by productivity, time on task, and quality versus pencil and paper techniques for essays written by 242 emotionally disturbed students. He effectively planned controls for each of the effects listed in Table 2, including writing method. However, there were slight flaws in the execution phase of his study. Some teacher effects were uncontrolled resulting from diffusion of teachers administering the study across 17 schools. Also there was some contamination of the control group once these students learned that others were using word processors. In the opinion of this researcher it would be interesting to repeat this study with better controls in execution to be able to confirm the results with absolute certainty. Such a finding would provide genuine empirical evidence of the efficacy of technology assisted writing for improving product quality. However, since there was

no actual instruction provided or assessment of learning outcomes resulting from technology mediated delivery systems in this experiment, it provides little if any new knowledge concerning the present issue as to whether or not media alone influence learning.

Gardner, Simmons, and Simpson conducted an experiment in which they determined that combining CAI with hands-on science activities significantly increased elementary student's learning outcomes. They administered the same pre-and-post test to control and experimental groups. Although they controlled for time on task and instructor effects, they failed to adequately control for method of instruction, prior knowledge, ability and learner familiarity with technology. As with many other flawed media comparison studies, there was no way to determine whether the improvement in achievement was due to the difference in delivery mode or some other uncontrolled factors.

Finally, among those studies reported by Sivin-Kacbala & Bialo as proving the positive effects of computers on learning achievement, Funkhouser (1993) investigated the influence of problem solving software on student attitudes about math. He determined that the use of problem solving software resulted in more positive attitudes among students about themselves as learners of math and about math as a discipline (p. 339). While he did control for prior knowledge and attitudes toward math through the use of a pretest, there were no other experimental controls, nor was there any control group established. All 40 students were administered the treatment in the absence of any control group. Tests of significance to measure attitudinal changes were based on a comparison of pre-and-post test scores using the same standard assessment test both before and after treatment. With no control group and lack of controls for random selection and assignment of subjects, this procedure was seriously flawed from the outset. There is no way to determine with certainty whether the changes in student attitudes were caused by having taken the same pre-and-post tests, as a result of the instructional methods presented in the software, or due to any other factor such as familiarity with the software program.

In military settings, meta-analyses and individual reports of experimental media comparison studies reviewed by this researcher have likewise found significant differences favoring technology over traditional modes of delivery. Not surprisingly, serious design flaws were found in many of the individual studies reviewed in military settings as shown in Table 3. A brief overview of the conclusions of the meta-analyses of military media comparison studies precedes a more detailed review of individual studies selected at random.

In a paper presented at the Symposium on Military Value and Cost-Effectiveness of Military Training, Orlansky, (1985) assessed the value of major innovations to training such as flight simulators, CBI, and maintenance training simulators in terms of cost and instructional effectiveness. Orlansky reviewed 22 studies and determined that flight simulators cost less to operate than aircraft and that “every hour in a simulator saves about one-half hour in an aircraft” (p. 12). Orlansky then reviewed 30 studies since 1968 that compared conventional to CBI delivery modes. He selected studies in which the control and experimental groups used nearly identical course content, allowed equivalent time on task between delivery modes, and used similar performance outcome measures (p. 17). According to Orlansky, “the only difference was that one group was instructed conventionally and the other group used CBI” (p. 17). Conspicuously absent from his criteria were more explicit controls to preclude confounding in critical factors such as instructional method, prior knowledge, and ability. In 40 courses where CAI delivery was compared with conventional delivery, he found student achievement the same in 24, superior in 15, and inferior in one (p. 17). He concluded overall that there were no significant differences between delivery modes from a practical perspective. For example, in 12 of 13 maintenance training comparisons, he found that “students trained with simulators achieved the same or better test scores than those trained with actual equipment” (p. 26). This is not surprising in view of the self-imposed limits that Orlansky placed on control mechanisms in the research he cited.

Regrettably, a finding of no significant difference is often misinterpreted in the literature. As was the case in Orlansky (1995), no significant difference findings were interpreted in the conclusions of his report as meaning that “flight simulators, CVI, and maintenance simulators are *as effective* for training as aircraft, conventional instruction, and actual equipment, respectively” (p.41). In poorly designed research, this may not be a valid conclusion. If the researcher has not carefully controlled, or held constant, the most likely factors which could explain variance in student achievement, then it becomes less likely that one will find significant difference between experimental and control groups since there is no absolute difference affecting only the treatment group. Likewise, in poorly designed research where a significant difference *is* found, it is likely the result of one or more uncontrolled variables such as different methods of instruction that were presented only to the experimental group. This common misinterpretation was also addressed by Clark in his seminal article (1983) on learning from media in which he argued that “no significant difference results simply suggest that the changes in outcome scores (e.g. learning) did not result from any systematic differences in the treatments compared” (p. 447).

Fletcher (1990) conducted a meta-analysis of 47 studies to assess the effectiveness of interactive videodisc (IVD) technology in military, industrial and higher education settings. In a somewhat refreshing approach, he acknowledged Clark’s view that hardware alone does not influence learning, and stipulated his review would “principally concern the instructional capabilities, or functionality’s made available by IVD systems, not the hardware itself” (p. 5-1). With this caveat, he then proceeded to summarize the conclusions of 47 media-comparison studies without reporting in each case the extent to which appropriate controls were used to isolate instructional methods, and other variables that might explain variation in outcomes. Rather, the studies he selected “had to conform to standard criteria for care and control” (p. II-2). Like many meta-analyses, the approach used was to measure effect size for the results of each study and draw conclusions from the aggregate. Effect size, as explained by Fletcher, is a standardized index, measured in standard deviations, of the difference - in this case the difference

in instructional achievement - between a control and a treatment group (p. 2-1). For all three settings and over a variety of instructional approaches such as simulation versus tutorials, and for various outcome measures including knowledge, performance, and retention, Fletcher concluded that IVD instruction increased achievement an average of 0.50 standard deviations over conventional instruction (p. s-2). On balance, Fletcher did acknowledge that there was little in the reviewed studies to indicate how IVD achieved its success and that this unknown factor “remains a proper topic for future research” (p. 5-3).

In one of the most recent reviews of media comparison studies in military settings, Barry and Runyan (1995) focused on the effectiveness of distance learning with regard to cost, efficiency, and instructional effectiveness. They determined that there were significant benefits to the military services in cost savings and throughput of students just in the past few years. For example, they reported cost savings to the Navy between 1989 and 1994 of more than seven million dollars in travel and per diem expenses alone (p. 37). They summarized the findings and conclusions of eleven experimental comparison studies and found that in nearly each case, there were no significant differences in achievement between distance and residential learners (p. 45). Without reporting further on the design methods and controls used in each of these studies, the authors could be misinterpreted as implying that different delivery media work equally well, when in fact that may not be the proper conclusion as noted earlier. In order to have a better understanding of these findings, this researcher examined several of the studies cited by Barry and Runyan, along with other similar studies, to assess their design and the appropriateness of the conclusions reached in military settings. These results are summarized in Table 3 and will be briefly described as follows.

Phelps, Wells, Ashworth, and Hahn (1991) examined the effectiveness of asynchronous distance learning for Army reserve officers using Computer Mediated Communications (CMC). Using CMC, 14 officers assigned to the experimental group met at different times via an electronic classroom for seven months. Their achievement was compared to a residential control

group of 370 officers who had only two weeks of instruction on the same material. The study report failed to mention what controls, if any were established for differences in instructional method. Performance results were mixed between the two treatment groups of officers. In one treatment group there was a significant difference while in the second there were no significant differences with officers in the residential program. This researcher evaluated this study as inconclusive due to its failure to control for instructional method and other critical factors such as prior knowledge and ability.

Winkler and Polich (1990) evaluated the effectiveness of IVD versus hands-on equipment in Army Communications training. The IVD system consisted of an integrated microcomputer, video display, laser videodisc, and courseware. This study was better designed than many and included a very elaborate matched pair computer program method for assigning subjects with similar backgrounds to control and treatment groups. While they adequately controlled for prior knowledge, ability, and time on task, there were uncontrolled differences across both groups in instructional method. The experimental group had nearly twice as many workstations in their classroom as did the control group (p. 28). Instead of working at their desk reviewing equipment manuals when not engaged with actual equipment like the control group, the treatment group used the IVD as a supplement to the equipment which provided an opportunity for differences in instructional method or routines such as drill and practice. There were no significant differences found on knowledge achievement tests, but the students in the experimental group learned quicker and with less effort. This is not surprising in view of the fact that they had more work stations available and spent their time more productively than the control group.

Keene and Cary (1992) studied teleconferencing versus residential learning in the Army for Reserve Components who participated in the Command and General Staff Officer's Course. In this study method of instruction was adequately controlled in that all students received the same lecture and materials. The experimental group participated from remote locations while the

control group was in residence. The study found significantly higher scores in favor of the teleconference group on 3 of 4 outcome measures (p. 101). However, the study suffered from some major design flaws in that there was no random selection or assignment of subjects to control and treatment groups, and a table of background differences for the participants did indicate some rather significant differences between groups in age and educational level. Also, the study did not control for ability or prior knowledge other than to administer a pre-test.

In a self-described quasi-experimental comparison study for the Army, Bessemer (1991) examined the effectiveness of tactical training exercises using a simulator networking (SIMNET) system as an enhancement to the Army Officer Basic (AOB) Course. Student achievement was measured before and after the course was modified to add SIMNET training. Regression analyses were conducted to measure changes in student knowledge and tactical performance during and after the mounted tactical training sequence of the course. Historical student performance data from earlier AOB classes were used for control purposes in comparing observations of the treatment group. Bessemer found that as instructors became more proficient in using SIMNET, there were positive gains in the quality of AOB graduates in performing tactical maneuvers (p. vii). He cautiously concluded that “improvement in SIMNET training... appeared to be responsible for much of the increases in [student] performance, advanced training, and graduate quality” (p. 42).

As acknowledged by Bessemer, there were several inherent limitations in the design of this comparison study, not the least of which was the absence of a properly constituted control group. Subjects were neither randomly selected nor assigned, and there were no controls for differences in instructional method, time on task, etc. Accordingly, Bessemer’s findings and conclusions regarding the role of SIMNET in producing these changes are highly suspect.

Simpson, Pugh, and Parchman (1991) compared student achievement results between live classroom instruction and six levels of video teletraining (VTT) in a Navy schoolhouse. The six levels of VTT ranged from multi-channel two-way video with two-way audio to audio-graphics

only. Due to several constraints of the setting there was no opportunity to use random selection or assignment of subjects, and the experiment was limited to one day out of four total course days for students assigned to the more austere levels of video/audio. Instructional method was the same for all versions of the course, except that instructors varied from week to week over 39 course sessions. Results indicated a slight difference in quiz scores for live classroom students but none of any practical significance among four of the seven delivery modes (p. 15) As with most of the studies reviewed by this researcher, there were sufficient flaws in design with this study to warrant a cautious interpretation of the results.

This small sampling of experimental media comparison research was typical of numerous other studies reviewed in an effort to determine whether a media comparison study would be appropriate for evaluating what elements of the PACE program contribute most to successful outcomes for Sailors enrolled in college level courses. Findings of no significant difference dominated these studies, and for reasons already cited in those instances where there were significant differences found, there were in most cases enough design flaws present to attribute the results to some uncontrolled variables other than the media itself. Next this literature review will examine a sampling of the descriptive research and its conclusions in the field of comparative media analysis.

Media Comparison Studies: Descriptive Research

As with the experimental literature, this researcher investigated findings and conclusions reached through a variety of non-experimental studies in both military and non-military settings regarding the advantages and disadvantages of technology versus instructor delivery modes. Again the literature was replete with papers, articles, and studies on this question, particularly relating to surveys where learners were asked to compare computer interactive technology with conventional delivery modes. In order to provide findings most pertinent to the central question of assessing PACE effectiveness since the last study in 1989, the following discussion will focus mainly on the more recent literature.

Martin and Rainey (1993) surveyed student attitudes after having participated in an experiment involving a satellite delivered high school science course. Secondary school students who participated in this study from seven public schools were generally very positive about the technology delivered lessons. Favorable attitudes outnumbered unfavorable by a margin of 2:1 among the nine subjects in the experimental group (Martin & Rainey, Table 3). Disadvantages of technology delivery cited were no direct contact for students with teachers, it was not helpful for average or below average learners, and there was difficulty contacting the teacher, and difficulty keeping attention focused on the television (Martin & Rainey, Table 3). The most substantive advantages cited were that it makes students more responsible for their work; their time was used more wisely; more material was covered than in regular classes; students were able to see the lab results better; and it allowed individual's to work at their own pace (Martin & Rainey, Table 3). In their conclusions, Martin et al. recommended that since distance delivery might not be appropriate for all students, criteria should be established for student participation in distance classes.

Janda (1992) conducted a quasi-controlled experiment to measure differences in achievement and attitudes among college students enrolled in an introductory course in American government and politics at Northwestern University. He lectured to the entire class three days a week and on the remaining two days, he divided the class into three groups employing a different technique of instruction for each group. Groups were led by teaching assistants using one of three modes: traditional group discussion of the assigned readings and the lecture; responses to questions about the readings and/or lecture followed by discussion of two CAI programs that students in that group were required to use along with the text; and discussion of the lecture plus discussion of questions the students had encountered after completing a multi-media hypercard-based videodisc unit (p. 342-343). Not surprisingly, he found no significant differences overall for the three groups in achievement, yet he obtained some very positive feedback from the multi-media treatment group. The majority (93%) agreed that video realism helped them understand

complex events more than just reading about them. Likewise, 89% said the video was worth the time it took, and 99% agreed the technology was easy to use (p. 348). Janda concluded from these mixed results that just because students like the technology delivery, they do not necessarily learn any better or worse from it in a social science setting (p. 352). To reiterate a point made by this researcher earlier in the present review of literature, Janda's conclusion regarding what he perceived as equally effective delivery modes may be erroneous since he failed to adequately control for differences in instructional method across the three groups.

Stone (1988) conducted a post-hoc descriptive study to assess variation in grades among graduate engineering students in the U. S. based on age, gender, major, and instructional modality. He compared results for the following delivery modes: on campus degree seeking; on campus non-degree; off campus degree seeking using non-interactive video; off campus non-degree seeking using non-interactive video; off campus degree seeking taking interactive video classes; and off campus non-degree seeking with interactive video classes (p. 10). He received data totaling more than 8400 records from 8 of 50 universities surveyed and used a main effects factorial model to test for significant differences among the groups by delivery mode, age, and gender. He found that off-campus students performed better than on-campus, and that off campus students using non-interactive video did better than their interactive video counterparts (p. 12). He concluded that off campus graduate engineering students "do not suffer from the inability to talk back to faculty in real time", and that "distance students performed better where they control not only where but when learning occurs" (p. 16).

Mitra (1994) assessed instructor effect on attitudes toward technology assisted teaching among college students. The Focus-Question-Analyze-Recommend (FCAR) evaluation method was used as a means of gathering in-depth attitudes and opinions from the students. There were five focus groups of seven students who evaluated two teachers, both of whom taught the class for half a semester using technology in the classroom. Results from the focus group discussions indicated that "perceptions of effectiveness of a tool (such as animated pictures or

demonstrations), and attitudes toward the use of multimedia in teaching a course, were significantly dependent on the way the instructors used the tools and the way the students perceived the instructors (p. 16). The author concluded that in technology assisted learning, the role of the instructor made a significant difference in perceived effectiveness of the technology.

Hansford and Baker (1990) assessed 38 college student's perceptions of the acceptability of instructional processes and communications via compressed two-way interactive video (IV) in a two-week teaching trial in Australia (p. 290). Differences in IV transmission speed were injected into the study, as were differences in instructional format (lecture versus tutorial) and instructional setting (large lecture hall versus small conference rooms). Subjects completed a Likert-type evaluation instrument including open-ended questions as the course evolved. Results indicated that the technology was not equally effective over the different teaching and learning situations (p. 306). Students preferred the small conference room IV presentations to lectures from the larger auditorium. Also, during the course, 80% of students rated the technology delivered course as useful, and slightly fewer perceived it as enjoyable. However, compared to face to face instruction, only 4% said that they would prefer IV (p. 297). In a final end of course critique the following areas were rated as unacceptable by the students: clarity and helpfulness of instructional aids; opportunities for questions and discussion; enjoyment; comparison with face to face instruction; and general usefulness of the IV class overall. Because of differences among student attitudes regarding differential effectiveness of technology delivery based on instructional setting, the author concluded that his results "suggest that the effectiveness of a delivery system is unlikely to be constant over [a variety of] settings and applications" (p. 303). If one were to consider setting as part of the instructional delivery system, than these results could be interpreted as weakening Clark's argument that delivery systems alone do not influence learning.

Turning to descriptive studies regarding technology versus instructor delivery in military settings, Phelps et al.(1991) found that Computer Mediated Communications (CMC) for Army Engineers was a more cost-effective delivery mode than resident training. However course completion rates

were much better for the residence (control) group (p. 12). Also, consistent with research reported the following year by Janda (1992) was a finding that in self-ratings, students in the treatment group evaluated their gains in level of knowledge to be much greater (33% versus 12%) than did the resident group. In spite of this strong attitudinal endorsement for technology delivery, achievement test scores between the groups were statistically equivalent (p. 12). The authors concluded that CMC was cost effective as a means of distance learning but that the extent of cost savings was dependent upon the course completion rate. As more students in technology delivery learning drop out, the cost per student rises (p. 16). They recommended that “a means of predicting which students are likely to succeed or drop out of distance study will help educators and trainers maximize resources, both by selecting appropriate students and by providing assistance to those at high risk” (p. 16).

Bramble and Martin (1995) investigated the efficacy of delivering military occupational specialty courses to remote locations via two-way compressed video from a civilian community college in Florida. They measured performance outcomes and student perceptions of 275 Army and Navy students but were unable to compare these results with resident delivery modes. Students in all four of the courses measured showed significant gains in achievement based on pre-and-post test scores (Bramble & Martin, Table 2). Student perceptions were also very positive with 80% having reported the technology delivery mode to be as effective as live instruction (Bramble & Martin, Table 3). The authors concluded that teletraining was a reliable and highly effective delivery mode based on student achievement and attitudes.

Returning to the only comprehensive descriptive study of the Navy’s PACE program, Dunlap et al. (1989) investigated course completion rates and academic achievement across both instructor and technology delivery modes. They found that “the withdrawal rate for [technology delivered courses] was more than twice the withdrawal rate for [instructor delivered] courses” (p. 30). In addition, they compared student achievement data (grade distributions) by mental ability or ASVAB score (composite of arithmetic reasoning and paragraph comprehension), type of ship,

and delivery mode (p. 19). As can be seen in Table 4, student performance as measured by grades alone was far superior in the instructor delivery mode compared with technology delivery (p. 30). Likewise, grade distributions by ASVAB group and delivery mode were found to be statistically higher for instructor delivery both within and between categories as shown in Table 5 (p. 31). Finally, grade distributions by type of ship in which the PACE courses were taken were found to be higher for instructor delivery in ships offering both delivery modes as shown in Table 6 (p. 34). The relationships found in this study between mental ability (ASVAB) and grade performance were consistent with earlier research reported by Lavin (1965) on intelligence and academic performance. However, the wide differences noted in achievement across delivery modes are inconsistent with much of the descriptive literature previously reviewed. It would be very useful to update findings in each of these grade distributions in view of the enhanced screening and eligibility requirements that were imposed on Sailors taking technology delivered PACE courses beginning in the early 1990's.

From this brief review of descriptive studies it appears that technology delivered courses offer significant advantages over instructor delivery in terms of cost and efficiency, particularly in distance learning applications. However, except for the results found in the 1989 PACE study, student achievement overall was about the same regardless of delivery mode. Most learners indicated a favorable perception of technology and a willingness to take additional technology courses again.

What has become clear to this researcher in light of the both the experimental and descriptive literature reviewed thus far is that there are a wide variety of variables that can explain academic achievement. Attempts to isolate any one variable such as delivery mode in an empirical study in order to prove causation have been difficult for most researchers to do and seldom produced any significant conclusion. Conversely, if truly significant differences were to be found in a properly designed experiment, their usefulness would be very limited because of the extent of artificial controls required to produce such a result. Before reviewing the literature

further for alternative study methods that should help determine those elements of PACE which contribute most to successful outcomes, a closer examination of PACE itself is necessary to understand the complexities and dynamics of this major educational enterprise.

Overview of PACE

As presently configured, PACE consists of three levels of courses provided to officers and enlisted personnel in ships for only the cost of their textbooks. Academic skills courses are designed for students in need of elementary or high school level remediation in basic mathematics, reading, and writing. Pre-college courses prepare students in need of limited remediation for college level studies. At the college undergraduate level a wide variety of liberal arts, business, and technical courses are taught in one of two delivery modes: instructor or technology delivery. Officers and enlisted personnel aboard ships are encouraged but not required to participate in PACE. Course offerings are made available by the Navy only at certain times of year on a ship by ship basis depending on factors such as ship operating and maintenance schedules. Courses are usually offered in conjunction with an extended deployment or operation at sea but may be conducted in the ship's homeport as well.

Under the current contract with the Navy, Middlesex Research Center, Inc. (MRC) as prime contractor develops a PACE Education Plan (PEP) for each ship designated to participate in PACE. Central Texas College (CTC), the sub-contractor responsible for providing instructor delivered courses, participates in the PEP development. In order to register for a college level PACE course, students must have attained a minimum composite score (AR + PC) of 100 on the ASVAB exam which is administered to all personnel upon enlistment in the Navy. Exceptions are made for applicants who have a satisfactory college placement exam result (American College Testing Program ASSET test) for the area they wish to study. They also must be current in all their assigned at-sea watch stations. Once qualified, applicants are interviewed and screened by the contractor's representatives to determine course sequencing and the appropriate delivery mode. Factors considered are type of ship (submarines do not have space for visiting

professors); the applicant's qualifications; school experience; and personal preference.

Additional screening by the ship's command authorities is strongly encouraged for students requesting to take technology delivered courses. However, the extent to which Sailors are screened by supervisors in their chain of command prior to enrolling in PACE courses varies widely from one ship to another.

In the instructor delivery mode, Central Texas College (CTC), a fully accredited two-year community college provides an instructor embarked in the ship for the duration of the course. Students meet as a class for 48 contact hours. Classes last six or eight weeks per term depending on which coast the ship is homeported. West Coast ships typically use six-week terms for the instructor taught (CTC) courses because they tend to remain at sea longer and can typically complete a course before making port calls in the Pacific. In the Atlantic and Mediterranean operating areas, East Coast ships typically make port calls sooner and therefore require additional time for the same number of class sessions due to the interruptions of the port visit. A minimum of ten students is required to establish a class, and once the course begins, all students enrolled are required to complete the course (Yeonopolis, J., personal communications, March, 1997). Students failing to complete the course are assigned a grade of "F" or "W". Withdrawals are permitted only when circumstances beyond the control of the student inhibit course completion as certified by the ship's command authorities. Students earning a grade of "F" in any class are generally permitted to take the class again, whereas those who withdraw without good reason cannot retake the course.

CTC offers a wide variety of courses worldwide, most of which can be taught aboard ship as long as the class size minimum has been met. Professors are even more numerous because they are hired for one term at a time. Some teach only the minimum of two courses before returning to their home campus or moving on to another shipboard PACE course. Occasionally a CTC professor will teach a series of classes in the same ship for several months at a time but this is more often the exception than the norm.

Technology courses on the other hand are offered by a consortium of five universities and colleges through the Navy's prime contractor for PACE, Middlesex Research Center, Inc. (MRC). Participating colleges are George Washington University, Coastline Community College, Richland College, University of Oklahoma, and the University of Maryland. For college level courses, students earn full credit from the college or university offering the course, and like the instructor delivered courses, pay only for the cost of their textbooks. Unlike the CTC six to eight week terms, semester lengths for technology delivered courses are approximately twelve weeks. Instruction is self paced and once a student begins the course, he or she must complete it or receive an "F" or "W" depending on the circumstances of the withdrawal. (Wilburn, B., personal communications, January 1997). Technology course applicants are personally interviewed by a contractor representative to evaluate their motivation and willingness to complete the course in view of the many distractions that occur daily when working and living in confined quarters such as those found in warships at sea. Also, consent is encouraged from the student's supervisor, taking into account the individual's performance record and demonstrated stamina in self-directed situations. Students accepted for technology delivered courses are provided familiarization training with the equipment, and a team of officers assigned to the ship is trained to proctor exams and monitor student progress. There is usually no direct interface with a professor until the course materials are turned in at the end of the term and evaluated by the respective college or university professor. One unique feature of technology delivered courses is the provision for a student to be extended into a subsequent twelve-week semester provided they are making satisfactory progress and have reasonable grounds for requesting an extension. This policy is not available to students enrolled in instructor delivered courses through CTC.

As for the hardware and courseware, the equipment used in MRC courses has evolved along with advances in technology since the PACE program began in 1977. Initially, the primary instructional delivery method was videotaped lectures supported by a computer for student accountability and presentation of material. Computer interactive video was first introduced in

1992 in conjunction with prerecorded video. Practice exams were in use by 1994 along with drill and feedback routines. The three primary systems in use as of July 1995 when the most recent PACE contract began are listed in Table 7 (Wilburn, B., personal communications, April 14, 1997). Approximately 40% of the computer equipment in use in 1995 and later were Pentium based 75 MHZ and the remainder a mix of 75 and 33 MHZ 386 and 486 processors. Courseware is currently produced by MRC in cooperation with each of the participating universities and colleges using Authware development software. Courses in use by MRC between 1995 and 1996 included video-only, computer interactive video, and computer courses without video. Of these options, approximately 85 to 90 percent of courses taught in that year used computer interactive video for delivery (Grubb, A., personal communications, September 8, 1997). Beginning in mid 1996, developmental courses in compact disc (CD) format were successfully pilot tested and MRC expects to have these new versions of technology delivered courses generally available by the end of 1997 (Wilburn, B., personal communications, April 14, 1997).

Alternative Research Methods

From the forgoing description it should be clear that the PACE program is quite extensive in terms of size, scope, and complexity. Differences in delivery mode are just one of many variations which any earnest attempt to assess PACE instructional effectiveness would have to take into account. Accordingly this researcher has concluded that a PACE media comparison study is neither appropriate nor feasible in view of the conclusions reached earlier from the review of literature that such studies produce largely meaningless results. In any event a media comparison study would most likely result in a finding of no significant difference and be of little use to the Navy in its assessment process. The review of literature will now turn to alternative types of research in educational settings that may offer a more effective avenue in determining what elements in the PACE program contribute most to successful outcomes.

According to Clark (1985), "as long ago as 1924, Freeman recommended multivariate studies with more careful controls" (p. 250) for assessing learning outcomes as an alternative to

media comparison studies. Multiple regression analysis is a statistical tool used frequently in descriptive behavioral studies “for relating variation in a set of predictor variables to variation in a criterion variable” (Wolfe, 1979, p. 319). The procedure provides for explanation of variance in the dependent variable by each of the independent variables while controlling for or holding constant all other independent variables. Unlike experimental research, such tools are not capable of determining cause and effect relationships. Multiple regression does not provide empirical explanatory evidence as to why something happens. Rather, its purpose is to establish correlation and to infer or predict outcomes based upon these correlations.

A small sampling of recent assessments of learning outcomes using multiple regression analysis demonstrates its potential usefulness as a method for assessing factors leading to successful outcomes for Sailors enrolled in PACE. Wolfe and Johnson (1995) used multiple regression analysis to identify the best combination of variables from high school grade point average (GPA), total scholastic aptitude test (SAT score), and 32 personality traits for predicting college GPA. They obtained high school records for 201 college students and gave these students a series of personality inventory tests (p. 179). Cumulative college GPA was regressed on each of the predictor variables resulting in a finding that approximately one-third of variance in college GPA is explained or accounted for by high school GPA (19%), organization (a self control personality variable) (7%), and SAT score (5%) (p. 182). Their findings show that self-control plays a major role in predicting college GPA in addition to the more common indicators used by college admissions officials.

In a similar study, Palmer and Wright (1996) examined whether age has a differential influence on academic performance as measured by GPA for students enrolled in an MBA program. They studied 86 graduate students as both an entire sample and divided by age into two groups consisting of younger ($n = 50$) and older ($n = 36$) students (p. 76). Through multiple regression analysis they found that age was a significant negative predictor of GPA for the sub-sample of older students (over 31 years of age), and for students who scored relatively low on the

verbal portion of the Graduate Management Admissions Test (GMAT) (p. 76). The authors speculated from these findings that older students frequently have demands on their time such as family and career obligations that may produce such a result (p. 78). However, any such inference about causation lies beyond the regression model itself and can be offered with no certainty.

Finally, Smittle (1995) found that a variety of academic and non-academic variables are useful in predicting academic performance among community college students. She effectively used multiple regression analysis to examine the collective and individual usefulness of Computerized Placement Test (CPT) scores, high school GPA, high school class rank, senior year absences, multiple skill deficiencies, race, and gender in predicting college GPA. The sample consisted of a cohort of 412 students who had enrolled in college within a one-year time frame (p. 40). College GPA was regressed on the independent (predictor) variables in both full and reduced models. In the full model, all variables were taken into account and for the reduced models gender, race, and CPT score were the independent variables (p. 40). Correlation coefficients for each variable with college GPA were calculated and showed that high school GPA had the greatest correlation (Pearson's $r = .52$) (p. 41).

Not surprisingly, senior year absences and multiple skill deficiencies yielded a negative correlation to college GPA (p. 41). The multiple regression analysis of the full model produced a combined explanation of variance in college GPA of 30%. However, in both the full and reduced models, race and gender were not significant predictors at the .05 significance level (p. 42).

The DOD Tuition Assistance study reviewed earlier (Boesel & Johnson, 1988), along with these few studies and hundreds more like them in the literature show multiple regression analysis to be a very powerful method for examining relationships among and between a multitude of categorical and continuous data variables. Given the wide variety of academic, environmental, demographic, and methodological factors likely to influence successful academic achievement for Sailors enrolled in PACE, this researcher has concluded that a descriptive study

employing both descriptive and regression statistical techniques is the most appropriate approach to answering the Navy's third question. Moreover, these procedures could provide a very useful tool in predicting those Sailors most likely to pass or fail PACE college courses. This capability would permit the Navy to screen and select applicants more effectively and to provide assistance or remedial instruction in advance to those at greater risk.

Variables of Interest In This Study

This researcher has identified several factors associated with the PACE program as potential variables of interest in an effort to determine which lead to successful outcomes for Sailors enrolled in college courses. A brief description and rationale for each is provided below. In some cases, the rationale may be predicated on personal observation and experience through close association with the PACE program in various ships for nearly thirty years. In others, the literature provides a theoretical basis for including these variables in the study.

The most practical criterion variable for descriptive analysis of performance outcomes for Sailors enrolled in PACE college courses is end of course grade. Not only is grade the most common criterion variable used in academic performance research (Lavin, 1965, p. 14), it is also the most readily available criterion variable to this researcher. Standardized tests were considered early in the research process as a criterion variable. However these are not routinely administered to Sailors taking PACE courses and would therefore have to be taken separately for research purposes. With Sailors assigned throughout the world, and many having taken PACE who are no longer in the Navy, this option would only work in a controlled setting under the most artificial circumstances. Further, as pointed out by Lavin (1965, p. 21), standardized tests would eliminate from the analysis of variance certain interactions between Sailors and teachers or between Sailors and technology which might help explain course outcomes.

In spite of the arguments just presented for using grade as the criterion variable, there is at least one serious control problem with grade that will undoubtedly limit its comparability in this study. Teachers have different grading criteria and use different test instruments based on

personal preference. Grading criteria will be a source of extreme variation for both CTC and MRC (technology) delivered courses since ultimately there is a human professor in each type of course making value judgements of Sailors' performance. Because of the large turnover in professors described earlier, there is no practical way to control for teacher effects other than in an artificially controlled experiment where generalizability of the findings would be limited to those participating in the research.

There are many potential variables that will be examined in this study. These include demographic factors such as gender, race, age, and marital status. Academic factors include mental ability, education level at time of accession into the Navy, number of semester hours of college courses taken and passed in the previous four years, number of attempts in cases where a Sailor took the same course twice, and PACE course delivery mode. Career and environmental factors are occupational specialty or rating, paygrade, years of active naval service, type of ship, specific ship, ship satisfactory completion rate for PACE courses, Coast of the U.S. to which the ship is assigned, and homeport. Each of these will now be described more fully and classified for use in simple descriptive analysis, regression analysis, or both.

Gender is a categorical variable of interest for both descriptive and regression analysis. Men and women have served in ships for several years, and in the past three years most all ships, less submarines, have become gender-neutral for assignment purposes. Approximately ten percent of most gender-neutral ship's crews are women. Witt, Dunbar, and Hoover (1994) studied interaction effects of adolescent boys and girls on academic achievement and found that there were differences between them. Girls performed better in language arts than boys, and boys outperformed girls in math and science (p. 241). These results have been corroborated in other studies of adolescent youth using standardized tests, according to Witt et al. (1989). However, in a study of graduate engineering students, Stone (1988) found no significant difference between men and women in success or failure rates (p. 13). Because of these contradictory findings in the

literature, gender should be examined in this study to determine whether or not it makes a difference in grade outcome within the PACE program.

Race/ethnicity is commonly found in social science literature as a categorical variable of interest and will be included for both descriptive and regression analysis. Because there are no quotas or restrictions of any sort regarding race or ethnicity in personnel assignments in the Navy, in all likelihood there is a normal distribution of Sailors in ships by race/ethnicity. Smittle (1995) found no significant differences in academic performance for this variable which was limited to an examination of white and black community college students. Accordingly, it is expected that there will be no significant differences by race/ethnicity in grade outcome within the PACE program. Data available for the present study will permit race/ethnicity to be categorized as White, Hispanic, Asian-Pacific, and Other.

Age has been shown earlier to make a difference in academic achievement and will be included for both descriptive and regression analysis. Stone (1988) found significant differences in performance among graduate engineering students in 26-30 and 31-35 year-old cohorts (p. 15). Conversely, as noted earlier Palmer and Wright (1996) found a negative correlation with academic performance for students 32 years of age and older. Based on personal experience the anticipated distribution of PACE students by age is expected to be from 18-38 with the majority under age 30. Because of contradictory findings in the literature, age is included as a variable of interest to determine its potential relationship to end of course grade.

Marital status, a categorical variable, will be included as a variable of interest for both descriptive and regression analysis. Married Sailors are generally older, more mature, and more focused than their single counterparts aboard ship in this researcher's experience. In other settings, family and career obligations could detract from their academic performance as was noted earlier. However, since most PACE courses are taken when Sailors are at sea, there should be fewer family-related distractions among married Sailors aboard ship. Accordingly, it is

expected that there will be a positive relationship between married Sailors and end of course grades.

Turning next to academic predictors, mental ability has been a key variable of interest in most studies involving academic achievement and will be included as a variable of interest for both descriptive and regression analysis. As a measure of mental ability, the ASVAB Composite (AR + PC) provides a standardized assessment and is readily obtainable for most Sailors. As reported earlier, the Navy has established a minimum ASVAB composite score of 100 in order to take PACE college-level courses based on the findings of Dunlap et al. (1989). Also, Lavin (1965) reported that as early as 1949, Cronback had found that on the college level, ability tests correlate about .50 with GPA (p.51). Based on these well-documented findings, ASVAB will likely be found to be a key predictor variable in that a positive correlation is expected to be found between ASVAB and course grade. ASVAB tests are administered at time of accession into the Navy and may be retaken during one's enlistment. In order to capture the most current ASVAB score available, data extracted from the Enlisted Master File will be up to date as of three months prior to the start date of PACE courses for each Sailor in the cohort.

Educational level is a continuous naturally ordered variable of particular interest in this study for both descriptive and regression analysis. In a review of studies measuring the impact of time on school and learning, Fredrick and Walberg (1980) reported that "in 54 surveys conducted between 1949 and 1971 involving 80,000 respondents, results clearly showed that higher levels of schooling were correlated with more knowledge" (p. 183). They reported that a 1966 survey showed similar results. As educational level at time of accession increases for Sailors enrolled in PACE, the present study will determine if, as expected, there is a positive correlation with end of course grades. Educational levels in this study are defined as less than high school; high school; associate degree; and higher than an associate's degree.

Number of semester hours of college-level courses passed within the previous four years is a continuous variable suitable for both descriptive and regression analysis. As in the case of

educational level, it is hypothesized that increasing hours of college courses passed within the previous four years will correlate positively with end of course grade.

Number of attempts is a variable of interest for descriptive purposes in order to determine whether taking the same course a second time makes a difference in course grade. Additional time on task and prior familiarity with the course material have been shown in the literature to improve grade performance. For these reasons it is expected that Sailors taking a course on the second attempt will earn better grades than Sailors taking a course for the first time. This variable is not suitable for regression analysis because of suspected grade bias which additional time on task for the second attempt would likely bring to the analysis.

Delivery mode is a complex categorical variable of interest for both descriptive and regression analysis because of the many differences previously noted in the two delivery systems. In this study, delivery mode implies far more than differences in the media themselves. It includes differences in time on task (6-8 week terms versus 12), methods of instruction, teacher effects, learner style, and student familiarization with technology. It also includes differences in assessed aptitude for college courses. Selection and assignment of Sailors to either mode is accomplished through a screening process that includes consideration of the Sailor's preference, motivation, and prior college history. In essence, it takes into account all the differences noted earlier between CTC and MRC delivered courses. Most of these differences, though important, cannot be measured individually due to the size and complexity of the PACE program. An artificially controlled study might be able to measure these effects individually but, as noted earlier, the results would probably not be generalizable beyond the study group due to the unrealistic conditions necessary to conduct these measurements. Because of the inherent complexity of this variable, and the overwhelming evidence in the literature of the futility of media comparison studies, it is expected that there will be no significant differences found for delivery mode.

There are several career/environmental variables of interest that may contribute to successful outcomes for Sailors enrolled in PACE college courses. As a categorical variable, Navy occupational specialties, referred to as ratings, are a well-defined series of descriptive job codes that identify the skills and duties of each Sailor assigned to that rating. Sailors must first pass through a lengthy apprenticeship before earning a rating. This usually occurs between two and three years of active service. Because the number and variety of rating categories found in the Navy aboard ships are quite extensive, it is anticipated that the variable “rating” would not make a significant contribution to the explanation of course grade Navy wide and is therefore not suitable for regression. However, descriptive statistical analyses of mean course grades by rating categories may be of value to the Navy as an assessment tool in determining which ratings have produced the highest end of course grades.

Years of service is also a continuous variable of interest for both descriptive and regression analyses. As Sailors reenlist for their second, third, or more terms of service, usually in increments of four years each, they acquire more professional knowledge, greater responsibilities as leaders and as professionals, and higher levels of salary. Sailors are not permitted to reenlist unless they have demonstrated excellent past performance and have potential to excel in future assignments. As professional performance qualities of Sailors improve, it is expected that there will be a corresponding correlation in academic performance.

Paygrade is a continuous variable of interest for both descriptive and regression analyses. Paygrade in the military is synonymous with promotion level in the private sector. There are nine levels of paygrade for enlisted personnel: E-1 through E-9. Since Sailors are promoted based on performance factors and accumulated time in service, it is customary that Sailors in higher paygrades will be more knowledgeable and outperform their juniors in professional tasks. Accordingly, it is expected that Sailors in higher paygrades will be more motivated and possess greater self-discipline than those junior to them resulting in achieving higher grades in PACE college courses.

Type of ship is a categorical variable of interest for both descriptive and regression analyses. Not only are Navy ships different by their size and displacement, but they differ significantly in their mission, combat systems, propulsion capabilities, and in the quality of Sailors who are assigned to them. Submarines are all propelled by nuclear reactors and all Sailors assigned to submarines are rigidly screened for intelligence, motivation, and integrity. This screening process varies among other types of ships more or less depending upon the mission and level of sophistication of its systems. As a result of these differences, it is expected that submarine Sailors will outperform Sailors in other ship types. For purposes of this study, ships will be categorized into one of four groups: aircraft carriers, cruiser-destroyers, submarines, and other.

The individual ship in which a Sailor takes a PACE course is a categorical variable of great interest to the Navy in helping to understand what environmental factors lead to successful course outcomes. It is a well-known fact within the Navy that a ship's Commanding Officer (CO) sets the tone for good order and discipline within the command. The CO also sets the priorities within the command, and these may vary extensively from one ship commander to another. The level of attention, support, and personal attention a CO provides for voluntary education can be a major factor in the success or failure of PACE aboard that ship. In order for Sailors to participate and be successful they must be permitted, if not encouraged, to devote the time and energy required for class participation and study. In many ships the working routine is arduous - particularly at sea where most Sailors have watch stations to man 24 hours a day in addition to their regularly assigned training, administrative, and maintenance duties. Making time for personal study in this environment is tough and can be extremely difficult without the full support of a Sailor's supervisors. Because of frequently conflicting priorities up and down the chain of command, it takes an active and supportive CO to set - not only the tone - but the policies to nurture and build a successful PACE program. With more than 270 ships participating in PACE, it is unlikely that this variable will have a significant correlation with end of course

grades Navy wide, and is therefore not suitable for regression analysis. However, it is included as a variable of interest for descriptive statistical analyses in determining which ships have the best grade average. Armed with this information the Navy should be able to learn a great deal about its most successful programs aboard these ships.

Satisfactory PACE college course completion rates by ship is included as a continuous variable for both descriptive and regression analyses. Course completion rates vary widely by ship based on factors such as weather, nature of ship's operations, and the degree of support provided by the CO and the chain of command. As a continuous variable, completion rate is particularly suitable for regression analysis to measure the extent to which command support may explain variance in end of course outcomes. Completion rate will be calculated by dividing the total number of Sailors enrolled in PACE on each ship during the cohort period into the number of Sailors who passed these courses. Both instructor and technology delivered courses will be factored into a single completion rate for each ship. It is expected that completion rates will compare favorably with end of course grades for each variable examined.

Coast is a categorical variable for both descriptive and regression analyses in order to determine the extent to which differences in administration of the PACE program on either coast of the U.S. may explain variance in course outcomes. There is at least one key difference between coasts. In the Atlantic, instructor taught course lengths are eight weeks whereas in the Pacific ships deliver the same course in six weeks. Based on additional time on task, it is expected that ships in the Atlantic will achieve higher course grades than their Pacific counterparts.

Lastly, homeport is a categorical variable suitable for descriptive analyses only. Differences in mean grades by homeport could alert PACE program managers on each coast to possible differences in program oversight or execution within their respective homeports. There are approximately 17 homeports in the continental U. S. and overseas from which PACE program managers plan and direct PACE operations. Due to the large number of homeports, it is not

anticipated that this variable would make a significant contribution to the explanation of course grade Navy wide and therefore would not be suitable for regression analysis.

Research Questions

The following questions pertaining to Sailors enrolled in college level PACE courses, when answered, will provide specific measures for the Director of the Navy's VOLED programs to use in determining what elements contribute most to successful outcomes for Sailors enrolled in college level PACE courses:

1. What is the average end of course grade for Sailors enrolled in PACE by gender, age, marital status, race, ASVAB, educational level, semester hours of recent college courses passed, years of active service, paygrade, rating, ship, type of ship, homeport, coast, number of attempts, and delivery mode?
2. Is there a difference in course grade by gender, marital status, race, educational level, paygrade, type of ship, number of attempts, and delivery mode? For those in which there is a difference, which have a mean difference of at least one whole letter grade?
3. Is there a correlation between course grade and ASVAB, semester hours of college courses passed, years of service, and age?
4. Is there a difference in satisfactory course completion rates (A-D Versus W/F) between gender, age, marital status, race, ASVAB, educational level, years of active service, paygrade, type of ship, coast, and delivery mode? For those in which there is a difference, for which types of ship, delivery modes, ASVAB, etc. do Sailors achieve the highest satisfactory course completion rates?
5. What is the completion rate by ship for Sailors enrolled in PACE who satisfactorily completed college courses during the cohort period of this study?
6. For each delivery mode, what are the percentages of Sailors enrolled in PACE by type of ship, Navy occupational rating, years of service, and paygrade?

7. What are the average ASVAB scores for Sailors enrolled in PACE by educational level, years of service, paygrade, ship type and delivery mode?
8. Taking ASVAB scores into account, to what extent do end of course grades differ for each delivery mode by ship type?
9. To what extent do differences in age, gender, race, ASVAB, marital status, years of service, paygrade, educational level, semester hours of recent college courses passed, coast, satisfactory ship completion rate, and delivery mode explain variation in end of course pass/fail outcomes?