Moral Development in the Military: The Efficacy of ROTC Morality Training at the Sophomore Level

Paul A. Gillett
Embry-Riddle Aeronautical University - Daytona Beach

Follow this and additional works at: https://commons.erau.edu/edt

Part of the Applied Behavior Analysis Commons, and the Social Psychology Commons

Scholarly Commons Citation
https://commons.erau.edu/edt/72

This Thesis - Open Access is brought to you for free and open access by Scholarly Commons. It has been accepted for inclusion in Dissertations and Theses by an authorized administrator of Scholarly Commons. For more information, please contact commons@erau.edu, wolfe309@erau.edu.
Moral Development in the Military: The Efficacy of ROTC Morality Training at the Sophomore Level

By

Paul A. Gillett
B.S. Biology, Norwich University, VT

Embry-Riddle Aeronautical University
Daytona Beach, Florida
Summer 2013
Moral Development in the Military: The Efficacy of ROTC Morality Training at the Sophomore Level

By

Paul A. Gillett

This thesis was prepared under the direction of the candidate's Thesis Committee Chair, Dr. Shawn Doherty, Assistant Professor, Daytona Campus, and Thesis Committee Members Dr. Kelly Neville, Associate Professor, Daytona Campus, and Dr. Alan Pratt, Associate Professor, Daytona Campus, and has been approved by the Thesis Committee. It was submitted to the Department of Human Factors & Systems in the College of Arts and Sciences in partial fulfillment of the requirements for the degree of Master of Science in Human Factors & Systems.

Thesis Committee:

Shawn Doherty, Ph.D.

Committee Chair

Kelly Neville, Ph.D.

Committee Member

Alan Pratt, Ph.D.

Committee Member

Christina Frederick-Recascino, Ph.D.

MS HFS Program Coordinator

Scott Shappell, Ph.D.

Department Chair

William Grams, Ph.D.

Dean, College of Arts and Sciences

Richard Heist, Ph.D.

Chief Academics Officer

Date: 8/8/2013
Abstract

The Reserve Officers Training Corps (ROTC) is an accessions program designed to produce quality commissioned Officers for operational units, in support of strategic Department of Defense (DoD) objectives. The traditional program length of 4 years coincides with the average number of years required to obtain a baccalaureates degree in the United States, in part because a degree is required for program completion. The program goals are to develop candidates physically, mentally, and morally in order to ensure they can be entrusted with the highest levels of leadership required of a US citizen.

This study aimed at assessing the moral development aspect of sophomore Naval ROTC students, specifically with regards to the efficacy of ROTC training. Navy ROTC, Air Force ROTC, and traditional (i.e., no military affiliation) sophomore students were asked to complete the online version of James Rest’s Defining Issues Test (DIT) version 2. Students were asked to complete 3 iterations: a pretest at the start of the Fall 2012 semester, a mid-test at the start of the Spring 2013 semester, and a post-test at the end of the Spring 2013 semester.

On the basis of high attrition levels of participants among traditional student participants, that group was excluded from the final analysis. Both as compared to themselves over the three iterations, as well as compared to Air Force ROTC students across iterations 1 and 3, Navy ROTC students showed no statistically significant difference in the indices of moral interest (i.e., P score and N2 score). The results suggest that Navy ROTC training at the Sophomore level does not significantly increase moral development as measured by the DIT-2. Additionally, Navy ROTC training does not appear to have any greater efficacy in moral development than Air Force ROTC training, despite service-specific differences in training approaches.
# Table of Contents

List of Tables ........................................................................................................... v
List of Figures ........................................................................................................... vi
List of Abbreviations ................................................................................................. vii
Introduction .............................................................................................................. 1
  Development of the Student .................................................................................... 3
  Morality Training ..................................................................................................... 8
  Measuring Morality ................................................................................................. 10
  Kohlberg and the six stages .................................................................................... 11
  The Defining Issues Test ....................................................................................... 15
  Alternative Methods ............................................................................................... 18
Methods ..................................................................................................................... 21
  Hypothesis ............................................................................................................... 21
  Participants ............................................................................................................. 21
  Materials/Apparatus ............................................................................................... 21
  Design .................................................................................................................... 22
  Procedure ............................................................................................................... 22
  Data Analysis ....................................................................................................... 23
Results ...................................................................................................................... 23
  NROTC across 3 trials ........................................................................................... 27
  NROTC and AFROTC across 2 trials ..................................................................... 30
Discussion ............................................................................................................... 34
References ............................................................................................................... 44
Appendix A .............................................................................................................. 49
List of Tables

Table 1: Kohlberg’s Six Stages .................................................................................. 14
Table 2: Summary of testing results ........................................................................ 26
Table 3: Raw data for selected NROTC participants .............................................. 27
Table 4: Means for NROTC participants in select categories ............................... 27
Table 5: Raw data for NROTC and AFROTC trials 1 and 3 ................................. 30
Table 6: Means for AFROTC and NROTC selected indices on trials 1 and 3 .... 31
Table 7: Means, standard deviations, and sample sizes for DIT-2 participants by education level. .... 37
List of Figures

Figure 1: P score means with standard deviations.................................................29

Figure 2: N2 score means with standard deviations............................................29

Figure 3: P score means with standard deviations for NROTC and AFROTC on trials 1 and 3………32

Figure 4: N2 score means with standard deviations for NROTC and AFROTC on trials 1 and 3……33
## List of Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Full Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>DIT</td>
<td>Defining Issues Test</td>
</tr>
<tr>
<td>ERAU</td>
<td>Embry-Riddle Aeronautical University</td>
</tr>
<tr>
<td>MIDN</td>
<td>Midshipmen</td>
</tr>
<tr>
<td>MJI</td>
<td>Moral Judgment Interview</td>
</tr>
<tr>
<td>MJT</td>
<td>Moral Judgment Test</td>
</tr>
<tr>
<td>P score</td>
<td>Post-Conventional Score</td>
</tr>
<tr>
<td>ROTC</td>
<td>Reserve Officers Training Corp</td>
</tr>
</tbody>
</table>
Introduction

The *Encarta Dictionary* defines “morality” as the “standards of conduct that are generally accepted as right and proper” (Microsoft Encarta Dictionary, n.d). It further defines “ethics” as the “study of moral standards and how they affect conduct” (Microsoft Encarta Dictionary, n.d). While people are most likely born with genetic programming that contributes to their ultimate morality, a certain degree of that same morality is also affected by their interaction with the environment (Kohlberg, 1976). Liaquat (2011) argues that “moral judgment competence is enhanced by ample opportunities of role taking and guided reflection while educational environment that does not promote autonomous critical thinking leads to stagnation, even regression in moral judgment competence” (p.60). Essentially, given education and opportunities to use that education, an individual can be developed properly in a moral sense. Courses in ethics, morality, and even leadership attempt to capitalize on this fact, with the goal being to develop an individual capable of greater moral reasoning than when he began the course. Those organizations such as the United States Navy (or the US Military in general) that offer these types of courses have an additional goal: to guide the moral development of the individual towards ideals which are consistent with the goals of the institution. For Naval Reserve Officer Training Corps (NROTC) midshipmen nation-wide, one mandatory class (in accordance with the 2012 NROTC curriculum) that should help meet this criterion is taught during the first semester of their sophomore year, and is entitled *Leadership and Management* (Naval Service Training Command, 2012). Additionally, the classroom instruction received during this course is augmented by guided training received outside of the classroom. While attempting to avoid defining what moral standards are best or most in accord with the goals of the institution, the aim of this experiment is to determine whether the course and additional instruction produces any
measurable change in moral reasoning among those who take it as compared to those who do not participate in the instruction.

By virtue of the nature of their mission, and the importance of that mission to the nation as a whole, military members are placed in a special position of trust and authority by society. In a democracy, the goals of the military are determined by that society, so individuals with moral deviations that are inconsistent with the goals of the organization ultimately betray public trust. With a standing force of almost 2.5 million people (who are by their nature imperfect), the US Military unfortunately is the scene of a great number of small deviances. As a rule, only the extreme cases of behavior (e.g., the My Lai massacre, the Abu Ghraib prisoner of war torture scandal) are known to the general public, and this knowledge generally excites within the public an attitude of shock and disbelief. It also causes the public to question how such a thing could have occurred.

This public knowledge of great deviances in the Armed forces has not always been the standard practice, at least from a US Navy standpoint. It is likely that the actions in places like the Falkland Islands (1831), Qualla Battoo (1832), and Panama (1903) would be poorly received by the US public of today (see Symonds, 1995, for details on these events). One big difference between this earlier time and now is the level of visibility of the actions of the armed forces, a change that occurred with unprecedented speed during the US involvement in Vietnam (approximately 1954-1975). Prior to this point, visibility of military actions was hampered by slow methods of communication and inferior recording devices. Military personnel were expected to act according to their understanding of their orders, but even failure to do so was unlikely to be widely or immediately known. Today, a group of US Marines might desecrate bodies of slain enemies, and this information can be visible to the whole world in a manner of
moments. This greater degree of visibility means that more emphasis must be placed on the ethical training of those military men and women sent to represent the country. Additionally, a paper by Lind (2000b) suggests that insufficient and premature termination of ethical training results in a situation where the individuals “usually show not only a lower level of moral competence but also signs of competence erosion afterwards” (p.3). This was based on a longitudinal cross-sectional study on adolescents in Germany, with the Moral Judgment Test (MJT discussed later) as the measuring tool. Those adolescents who graduated from middle school (i.e., around age 15) and entered the workforce showed signs of moral regression as compared to their peers who continued to attend schooling, suggesting a link between higher education and moral development. This means that ethical training must not only be emphasized as important, but must also be continued for a sufficient amount of time to prevent moral regression. A training program must therefore be of sufficient length, and provide adequate instruction in morality to allow for proper moral development of the student.

**Development of the Student**

The NROTC program (both at Embry Riddle and elsewhere) is designed to be a 4-year program which prepares young men and women to enter Naval military service as officers. This goal is believed to be best served by instructing the students in numerous ways, including ethics and moral preparation. While each year the Naval ROTC program shares training similarities with previous years, each year is also designed to develop the ROTC student in its own unique way. For NROTC sophomores, military development at Embry-Riddle is comprised of two components: a classroom component (i.e., NSC201 Leadership and Management and NSC202 Navigation) and supplemental military instruction in the form of physical training sessions and military leadership lab. With respect to the classroom, the leadership and management course
and the navigation course are unique when compared to many other college courses in that they are standardized: the basic tenets of each course are the same, whether an individual is taking the classes at Norwich University NROTC in Northfield, VT or at Embry-Riddle NROTC in Daytona Beach, FL. Instructors are given a curriculum guide which details what lessons are required to be covered, as well as the learning objectives for each lesson. Sample bullets are provided in these books that help guide the preparation of lesson plans, and additional material is available on Navy Knowledge Online (NKO) for use.

With respect to training outside the military classroom, Midshipmen (MIDN) classroom instruction is enhanced by additional military instruction. Students meet at least three mornings a week as a battalion (i.e., all the MIDN assigned to Embry-Riddle) from 0545 to 0700 for physical training (e.g., running, pushups, situps). Additionally, the battalion meets from 1545 to 1800 (i.e., 3:45 to 6:00 pm) on Thursdays, and additional instruction is given which ranges from marching and ceremony to general military knowledge presentations. During these supplemental periods, another factor which has changed since the student’s freshmen year becomes more important: leadership roles. Following completion of the indoctrination period in the Fall, freshmen MIDN are integrated into the battalion, but are assigned no real responsibilities. As returning sophomores, they are assigned junior leadership billets in the battalion, which will require them to do at least three things: figure out what the standards are so they can meet them, act as good examples for their subordinates, and enforce the policies of the unit to the extent they are able. This will apply not only from the standpoint of rules and regulations, but also morality. These MIDN will be expected to ascertain what moral values are consistent with the organization and abide by them, as well as teach them to their subordinates.
Classroom instruction with supplemental instruction in guided events is not unique to Embry-Riddle NROTC; the Air Force ROTC (AFROTC) uses this method as well, although its time periods and the nature of its instruction are somewhat different. Sophomore AFROTC students begin taking on junior leadership roles, and they will be learning and enforcing the goals of the organization. One of the more substantial differences between NROTC and AFROTC appears to be the classroom portion of instruction; whereas NROTC students are covering leadership and management (1st semester) and Navigation (2nd semester), AFROTC students are covering Air Force history (both semesters), a class that covers parallel material which is already addressed in the 2nd semester of an NROTC student’s freshmen year. This is an important distinction in that any differences between AFROTC and NROTC students might be more readily attributable to classroom instruction, since both AFROTC and NROTC students receive similar forms of instruction external to the classroom.

The day-to-day operations of a non-ROTC affiliated sophomore may not necessarily be too different from his or her ROTC counterparts. Such a student may opt to wake up early and work out. He or she may also engage in forms of moral studying (e.g., enrollment in the course titled “Values and Ethics”), and may occupy positions of junior leadership in organizations with which they are affiliated, such as in student organizations. They are unlikely to wear uniforms, but then again they may have an affiliation with an entity where uniforms are required such as for their work on campus or for a sports team. What can be stated as a certainty is that they will not engage in battalion activities in the same fashion as the Midshipmen, nor will they take the leadership and management class or the navigation class. This is an important distinction from the standpoint of evaluating the combined efficacy of classroom instruction and external guided instruction; whereas neither AFROTC nor traditional students take the NROTC classes, the
traditional students also do not do the same external guided instruction as AFROTC or NROTC. If, then, a greater difference is found between NROTC and traditional students than between NROTC and AFROTC, that additional difference might be attributable to this external instruction.

Since neither the AFROTC students nor the traditional students are engaged in the same form of training as their NROTC counterparts, there is a distinct possibility that their moral development over this period will be different from the NROTC students. This experiment is oriented towards investigating that possibility. A few already published studies give reason to suspect a difference (e.g., Bridston, 1979; Panowitsch, 1975; Piwko, 1975; Schlaefi, Rest, & Thoma, 1985; Wattendorf, 1981). Panowitsch (1975) studied the moral development of 152 undergraduates, split between an applied ethics course and a logic course, each taught for a quarter (i.e., 10 weeks). The students in applied ethics showed statistically significant gains in post-conventional reasoning (i.e., the P-score, Kohlberg’s stages 5 and 6) on the Defining Issues Test (DIT) as compared to their logic counterparts. Piwko (1975) studied 68 undergraduate students over the period of a quarter, with the participants split between a moral development workshop emphasizing moral values and commitment, and a basic human development class. Those students in the moral development workshop showed significant gains in p-scores as compared to their counterparts, who had no significant changes. Bridston (1979) studied 69 first year nursing students in two conditions: the fall semester and the spring semester. Both classes were oriented towards student participation and discussion of ethical dilemmas, with discussions of health policies included. Both classes also showed significant gains on the DIT p-score as compared to their control counterparts.
While the previously mentioned studies are valid from the standpoint of evaluating undergraduate students in general, participants in a study by Wattendorf (1981) are more closely aligned to the groups of interest in this study. The Wattendorf (1981) study used the DIT and compared Naval ROTC undergraduate student scores at various years of study to the scores of their traditional college counterparts. The results indicate that Freshmen NROTC students had higher levels of conventional (i.e., Kohlberg stages 3 and 4) and principled (i.e., Kohlberg states 5 and 6) reasoning than their traditional counterparts, but that NROTC upperclassmen (juniors and seniors in this study) were not statistically different in terms of principled reasoning from their traditional counterparts. Additionally, upperclassmen scored higher than freshmen in principled reasoning. According to Lawrence Kohlberg’s theory of moral development (discussed later), principled reasoning indicates a higher stage of moral development, so the results seem to indicate that both traditional and ROTC students showed moral development while at college, with both groups showing effectively the same level of moral development by the time they were upperclassmen (Wattendorf, 1981). This would seem to indicate that “ROTC training appears to have no discernible effect on principled level of moral reasoning beyond that of a normal college education” (p.1). This conclusion, however, may be premature, because the Wattendorf (1981) study was not longitudinal. The only real conclusion that can be drawn is the relationship of conventional and principled scores of one group to another. It could well be that the ROTC freshmen measured in this study became more principled than their traditional counterparts by the time they were upperclassmen. Conversely, they may have showed some form of moral regression and become less principled. A better approach to determining the accuracy of the conclusion above should therefore involve measuring a group of students (or several groups) over a period of time during which they are receiving moral instruction. This
would allow the actual moral development of the students to be determined, which could then be used to determine the quality of the instruction being received.

**Morality Training**

Longitudinal studies conducted by Kohlberg led him to believe that humans develop in a stage progression fashion from childhood through adulthood (Kohlberg, 1976). The Lind (2000b) study indicates that premature termination of moral training will likely result in moral stagnation and regression. Lind concluded that moral development, like other competencies, degraded when not practiced, and those individuals who had not yet reached moral autonomy through sufficient education were likely to avoid morally difficult situations, causing moral regression. On the basis of the findings by Kohlberg and Lind, it therefore seems reasonable to assume that moral training occurs in some fashion, since human beings develop it, and it can be either stopped prematurely or allowed to continue. While there is considerable debate as to the most important venues for moral change (e.g., church, a book, at school, or in the home), one consistent finding is the importance of higher education to moral development; multiple studies have shown a positive correlation between the level of higher education achieved and the attainment of higher moral reasoning (e.g., Liaquat, 2011; Lind 2000b; Rest & Thoma, 1985; Trow, 1976). More importantly, multiple studies geared towards measuring the effect of morality training (i.e., moral intervention studies) indicate that moral instruction produces a measurable change (at least on the DIT) in moral development scores (e.g., Abdolmohammadi, 2005; Schlaefi et al., 1985). For example, Hanford (1980) involved 32 nursing students exposed to moral intervention over a period of 10 weeks. Through the use of discussions about bio-ethics dilemmas, written essays about ethics, and exposure to Lawrence Kohlberg’s theory, the
researcher reported a statistically significant change in moral reasoning scores for the experimental group, which was not the case for the control group (Hanford, 1980).

While direct moral instruction (i.e., classroom work) is important, there are indications that it may be insufficient in and of itself to produce substantial moral development. One key theme of recent moral intervention studies seems to be a greater emphasis on role-playing, which involves giving individuals “opportunities to take over real responsibilities” (Lind, 2000b, p.4). According to Lind (2000b), moral development can proceed in a self-sustaining fashion once an individual reaches moral autonomy, which is aided in part by role-playing. This moral autonomy means that a “person will seek rather than avoid morally difficult situations, and will grow by coping with them” (p.4). Quite apart from creating a self-sustaining level of moral development, there is arguably a more important aspect of this moral autonomy: its relationship to leadership. According to Lind (2000b), “only if a person becomes morally autonomous, he or she is also competent to take over high responsibility for others and for him or herself” (p.6). This is predicated on the idea that morally autonomous individuals seek to view society from a post-conventional standpoint, wherein transcendental rules (e.g., justice, right to life) are created to maximize the rights of all participants. From the standpoint of producing officers—who will be in charge of other people, and will be responsible for ensuring their maximal welfare—it therefore seems critical that role playing be a component of military morality instruction.

Given that the leadership and management class provides some morality instruction, and that the supplemental activities allow for a large degree of role playing, there should be an opportunity for students to develop morally if both forms of training are adequate. In order to assess whether adequate moral training is being provided, there must be a method for measuring morality in the first place. Thus, determining the efficacy of training requires an understanding of
the individual’s baseline performance, as well as how training has moved the individual with regards to the baseline.

**Measuring Morality**

Morality is a difficult theme to measure for a number of reasons. First, a researcher must determine whether he or she wants to review morality of action or morality of thought; the one does not necessarily lead to the other. For example, in *Mein Kampf* Adolph Hitler wrote about a number of actions that a moral person would not find objectionable, from the care of orphaned children to doing what was best for the good of the community (Hitler, 1925). His observed conduct, on the other hand, was a good deal less moral. Mother Theresa, by contrast, is said to have written very dark things in her private journals, but spent her public life caring for the terminally ill (Van Biema, 2007). For the purposes of this study, morality of thought will be reviewed.

A second concern for measuring morality is to determine whether the degree of morality should be measured at the macro or micro-morality level. Macro-morality is how an individual acts with regards to social norms and how they act in relation to society. Essentially, macro-morality “concerns the formal structure of society, as defined by institutions, rules, and roles” (Rest, Narvaez, Bebeau, & Thoma, 1999, p.291). Micro-morality, on the other hand, is more concerned with interpersonal relationships, such as how the individual interacts on a daily basis with other individuals. Micro-morality then “concerns the particular face to face relationships that people have in everyday life” (Rest et al., 1999, p.291).

A third concern for morality measurement is recollection. It would be impractical for a researcher to follow subjects around, waiting for them to perform some act involving morality, and then ask them the reasoning that led to that act. Instead, a researcher may have to probe for
past events of moral dilemmas, and then ask the individual what he or she did and how their reasoning led to that decision. This creates an issue, since it is well documented that people often have poor recollections of their past cognitive processes (e.g., Loftus, 2002; Nisbett & Wilson, 1977; Rest et al., 1999). A more useful approach then might be to use a test that would measure moral competence with respect to hypothetical dilemmas, since all participants can be exposed to exactly the same dilemmas. This is the method to be used for this experiment.

A final concern for morality measurement is scale: how do you rate a man who helps an old lady cross a street in relation to a man who steals a drug to help keep his wife from dying? If morality research is to be done in an objective manner (i.e., in a manner not placing a value of right or wrong on the answer), a good approach would be to evaluate the action in terms of moral development. The question then becomes, “What level of moral development would cause the individual to reason in such a manner?” To answer that question, an accepted standard of moral development must be available. While there has been considerable debate about the validity of this line of reasoning, a generally accepted theory of moral development is Lawrence Kohlberg’s six stages of development.

**Kohlberg and the six stages**

Kohlberg (1958, as cited in Kohlberg, 1976) defined moral reasoning in terms of six stages an individual could occupy. In his model, development is roughly tied to age, with individuals showing “development in adolescence to adulthood, from conventional to post-conventional thinking” (Narvaez, Getz, Rest, & Thoma, 1999, p. 479). This development is said to occur in an invariant sequence (Kohlberg, 1969); effectively a “new stage displaces its predecessor because it provides better cognitive tools to deal with moral problems” (Matarazzo, Abbamonte, & Nigro, 2008, p.667). Unlike Lind’s MJT or James Rest’s DIT, Kohlberg’s theory,
as well as his tool for measuring stage development (i.e., the moral judgment interview), do not seem to allow for the idea of moral stagnation or regression (Lind, 2000b). Kohlberg’s theory also rests on an assumption (and one with considerable controversy): “Knowing what the good is equates with doing the good” (Matarazzo et al., 2008, p.667). Lastly, the theory is said to be more useful at explaining moral development from a macro-morality standpoint than a micro-morality standpoint in that it measures a person’s moral attitudes with respect to society, not with respect to interpersonal relationships (Rest et al., 1999).

Kohlberg’s theory of morality can be defined by six stages, lumped into three major levels of moral reasoning: pre-conventional, conventional, and post-conventional. Each level contains two stages, wherein the “second stage is a more advanced and organized form of the general perspective of each major level” (p.33). The pre-conventional level contains stages 1 and 2, and is primarily concerned with morality and its effects on self (Liaquat, 2011). During stage 1, the individual sees morality as something external to themselves, a series of rules to be followed in order to avoid punishment. At this stage, punishment is something that validates that an action is wrong. In stage 2, there is a transition both in terms of viewing right answers, as well as how punishment is viewed. For right answers, the individual begins to realize that more than one right answer can exist to a problem (i.e., other people have interests they want to pursue), and emphasis should be placed on the answer that meets the needs of the individual (Borgatta & Montgomery, 2000; Kohlberg, 1976). For punishment, the individual no longer sees it as validation of a wrong action, only as a possible consequence/risk of an action.

The Conventional level (i.e. stages 3 and 4), in contrast to the pre-conventional theme, transitions from a view of self to a view of the whole: morality is no longer an external element imposed by others, but an agreement among members of society (Kohlberg, 1976). In stage 3,
the individual begins to view morality in terms of motives and what is good for the community. The nature of rules is becoming more relativistic, with questions of what is good for the group as a whole beginning to erode the inviolate nature of those rules. In stage 4, the individual will now see themselves as a member of society, and will have effectively transitioned away from the egocentric tendencies of the pre-conventional level. The hallmarks of stage 4 are obedience to authority, respect of the law, and doing one’s duties, all with the overarching goal of maintaining social order.

The Post-Conventional level (i.e., stages 5 and 6) is marked by a transition from simple social order to an ideal standard, or what Kohlberg called a prior-to-society mindset (Kohlberg, 1976). Whereas the conventional level stresses the need for a group mindset and a functioning society, the post-conventional level stresses viewing this society from a more idealistic standpoint: this society is clearly well ordered, but is it good? Stage 5 is marked by a more abstract way of thinking about morality. The individual must mentally remove themselves from their own society and determine what goals a society ought to pursue (Borgatta & Montgomery, 2000; Kohlberg, 1976). While still valuing obedience to law, this introspection allows them to see that certain inalienable rights should exist, and that the pursuit of these rights overrules the authority of laws that impede them. Stage 6 is characterized by universal principles, a concept involving the impartial review of all parties concerned, with the intent of creating rules that respect the basic dignities of the individual. In Kohlberg’s eyes, the underlying theme was justice (Kohlberg, 1976; Rudd, Mullane, & Stoll, 2010), and a stage 6 individual would pursue this justice even if it entailed civil disobedience. Table 1 is a partial copy of table 2.1 from Kohlberg’s (1976) chapter, which summates Kohlberg’s six stages (Kohlberg, 1976, p. 34-35).
Table 1. Kohlberg’s six stages.

<table>
<thead>
<tr>
<th>Level and Stage</th>
<th>What is right</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Level I-Preconventional</strong></td>
<td>To avoid breaking rules backed by punishment, obedience for its own sake, and avoiding physical damage to persons and property. Following rules only when it is to someone's immediate interest; acting to meet one's own interests and needs and letting others do the same. Right is also what's fair, what's an equal exchange, a deal, an agreement.</td>
</tr>
<tr>
<td>Stage 1- Heteronomous morality</td>
<td></td>
</tr>
<tr>
<td>Stage 2- Individualism, instrumental purpose, and exchange</td>
<td></td>
</tr>
<tr>
<td><strong>Level II- Conventional</strong></td>
<td>Living up to what is expected by people close to you or what people generally expect of people in your role as son, brother, friend, etc. Being good is important and means having good motives, showing concern about others. It also means keeping mutual relationships, such as trust, loyalty, respect, and gratitude. Fulfilling the actual duties to which you have agreed. Laws are upheld except in extreme cases where they conflict with other fixed social duties. Right is also contributing to society, the group, or the institution.</td>
</tr>
<tr>
<td>Stage 3- Mutual interpersonal expectations, relationships, and interpersonal conformity.</td>
<td></td>
</tr>
<tr>
<td>Stage 4- Social system and conscience.</td>
<td></td>
</tr>
<tr>
<td><strong>Level III- Post-Conventional</strong></td>
<td>Being aware that people hold a variety of values and opinions, that most values and rules are relative to your group. These relative rules should usually be upheld, however, in the interest of impartiality and because they are the social contract. Some nonrelative values and rights like life and liberty, however, must be upheld in any society and regardless of majority opinion. Following self-chosen ethical principles. Particular laws or social agreements are usually valid because they rest on such principles. When laws violate these principles, one acts in accordance with the principle. Principles are universal principles of justice: the equality of human rights and respect for the dignity of human beings as individual persons.</td>
</tr>
<tr>
<td>Stage 5- Social contract or utility and individual rights</td>
<td></td>
</tr>
<tr>
<td>Stage 6- Universal ethical principles</td>
<td></td>
</tr>
</tbody>
</table>

The Kohlberg theory of development is not without its critics. It has been called sexist (e.g., Gilligan, 1982), culturally biased (e.g., Vine, 1986), a political ideology masquerading as cognitive development (e.g., Rest et al., 1999), and out of touch with everyday morality (e.g., Killen & Hart, 1995). It also has a flaw in equating the knowledge of good to the desire to commit the act; Matarazzo, Abbamonte, and Nigro (2008) determined that “people are more likely to adopt moral values when faced with situations in which the price requested by moral choices is not too high for their concerns” (p.673). This kind of mindset is a form of situational
ethics, which involves “changing the ethical guidelines whenever a good or better reason exists to change them” (Lumpkin, Stoll, & Beller, 2003, p. 42). Despite these criticisms, Kohlberg’s theory has stood the test of time, both based on its own merits, as well as by refuting arguments (see Kohlberg, Levine, & Hewer, 1983). One such proponent of the theory, James Rest, used Kohlberg’s theory to develop an empirical method for collecting morality data, called the Defining Issues Test (Rest, 1979).

**The Defining Issues Test**

The Defining Issues Test (DIT; Rest, 1979), in its original form, is a standard paper and pencil test presented in the form of six moral dilemmas: Heinz and the Drug, Student Takeover, Escaped Prisoner, Newspaper, Webster, and The Doctor’s Dilemma. Each dilemma is presented in paragraph form and followed by three choices as to what should be done. The choices are then followed by twelve generalized statements, each indicating a possible line of reasoning for having made that decision. These statements are presented/worded in such a fashion that “the fragmented nature of the items requires the participant to supply meaning to the items that they are rating” (Rest et al., 1999, p.301). Unbeknownst to the test-taker is that each of these statements is also phrased in such a way so that it closely aligns with the reasoning in 5 of the 6 stages in the Kohlberg theory. For example, in “Heinz and the drug”, statement 1 is “Whether a community’s laws are going to be upheld.” This is an example of stage 4 reasoning, since its key theme is adherence to law for the sake of the community. Statement 3, on the other hand, states “Is Heinz willing to risk getting shot as a burglar or going to jail for the chance that stealing the drug might work?”, which is stage 2 reasoning; a person at stage 2, among other things, views stealing as a risk of certain behavior, not as the justification that the behavior itself is bad. After reviewing the statement, the individual rates each of the twelve statements according to their
level of importance in the decision making process (Great, Much, Some, Little, or No), and ultimately they rate (in order, from 1-4) the top four statements they consider as most important in having helped reach that decision. These top four statements are weighted by the researcher, combined with the choices from each of the dilemmas, and then used to determine at what moral development stage the individual reasons predominantly. Because each statement is tied to 1 of the 6 stages (except stage 1), no matter which four the individual ranks as most important, the researcher can use that data to assess where the individual is at on the Kohlberg scale.

Since its creation in 1979, the DIT has been used in many studies (see Schlaefi et al., 1985, for a brief review of 55 studies), and is tied to a database of over 500,000 test takers (Rest, Narvaez, Thoma, & Bebeau, 1999). Due in part to critiques by reviewers, as well as additional research by Rest, the test has also been modified in a few ways and is now available in an updated version, the DIT-2. The DIT-2 is a modification both in terms of the number of stories (5 versus 6), as well as the type of stories, the types being more modern themes versus Vietnam-era moral dilemmas (Rest et al., 1999). Additionally, the nature of the measured indices has changed, where the “index is the overall score by which a participant is characterized” (Rest, Thoma, Narvaez, & Bebeau, 1997, p. 498). Initially, while researchers were capable of determining at what level of moral reasoning the individual was at, they were most interested in the extent of post-conventional reasoning, which they termed the p-score. This led to a tendency to collect data only on the answers that indicated post-conventional reasoning, which resulted in a high prevalence of data on p-scores at the expense of other measures (e.g., type indicator, utilizer score). The DIT-2, on the other hand, generates at least eight indices (per test) upon which an individual can be evaluated; three are related to the three main levels of Kohlberg’s theory, two are tied to comparisons (either post-conventional to personal or dilemma solution to
items rated most highly), two are comparisons to expert ratings of certain items (Humanitarian and liberalism score or anti-social score), and one is tied specifically to religion (Rest et al., 1997). In addition to a variation of indices, there is also greater variance preserved in the sample, since the DIT-2 purges fewer participants during answer reliability checks than the original DIT (Rest et al., 1999). Lastly, in keeping with the modern age, the DIT-2 can be administered online, and the online version seems to have the same validity and ease of use as the paper and pencil version (Xu, Iran-Nejad, & Thoma, 2007).

Whether one uses the DIT or DIT-2, the basic question arises as to whether either of these is an adequate measure of moral development. This depends to a certain extent on whether Kohlberg’s theories were right, given that Rest built his model around Kohlberg’s ideas of moral development. Kohlberg himself mentions a .7 correlation between results from Rest’s method and results from the moral judgment interview, Kohlberg’s own form of moral development assessment (Kohlberg, 1976). Kohlberg viewed moral development as proceeding along similar lines as other forms of cognitive development; an individual interacts with the environment and begins building a mental framework in order to better understand those interactions (Kohlberg, 1969). Kohlberg used his research to begin constructing stages (i.e., distinct levels of development), and he validated the usefulness of these stages by testing for (and finding) their presence in cultures around the world (see Kohlberg, 1969). Like Kohlberg’s initial stage research, the DIT has been used worldwide, and is available in multiple languages (Ahmed & Gielen, 2002). In fact, the DIT is said to be “the most widely used measure of moral judgment development” (Thoma, 2002, p.225). Given the data in support of Kohlberg’s work, as well as the strong correlation between the DIT and Kohlberg’s form of assessment, it seems reasonable to conclude that the DIT is an acceptable means for assessing moral development.
The DIT, despite its widespread acceptance, is not without its weaknesses. For example, although it was intended to be able to assess moral reasoning according to the Kohlberg scale, a 12-year-old reading level is required to take it (Rest et al., 1999; Rudd et al., 2010). This age also happens to be where some transition is occurring from pre-conventional to conventional reasoning, making the test less effective at measuring stages 1 and 2. A second weakness is length; reading the six dilemmas (5 for DIT-2), rating the twelve statements, and then choosing the four most important statements is time consuming. Rudd, Mullane, & Stoll (2010) capture this dilemma succinctly by stating “too many long, complicated scenarios may cause respondents to fatigue and provide unreliable responses” (p.68). This concern has been addressed to some extent in the DIT-2, which has fewer dilemmas, but the concern as a whole is still very real. A third weakness of the DIT-2 is in terms of calculations. Whereas DIT was a pen and paper test that could be reviewed and scored by a human, the complexity of the DIT-2 is such that “only a computer should be put through the amount of calculation necessary to produce N2 and new checks” (Rest et al., 1999, p. 652). The weaknesses of the DIT have led some researchers to develop other testing forms which they believe are superior in collecting morality data. Some of these alternative methods are discussed below.

Alternative Methods

While the DIT appears to be empirical, and is based in part on the work done by Kohlberg, Kohlberg himself did not fully endorse the use of the DIT. In fact, Kohlberg jokingly referred to the test as a form of alchemy, likening it to medieval scientists trying to turn lead into gold (Rest et al., 1999). For him, the DIT was a quick and dirty method, with the Moral Judgment Interview (MJI) being a superior way to assess the level of a person’s reasoning. In the MJI, the interviewer uses the same moral dilemmas as described in the DIT, and asks open ended
questions related to the dilemmas (e.g., “why shouldn’t you steal from a store?”); (Kohlberg, 1976, p.36). The interviewer guides subsequent questions from the responses of the interviewee, and the session is tape recorded. In order to maintain inter-rater reliability, the MJI requires multiple individuals to rate the answers of the interview, with the goal of assessing at what stage the individual reasons. For example, to the question of why you shouldn’t steal from a store, an answer might be “It’s not good to steal from the store. It’s against the law. Someone could see you and call the police” (Kohlberg, 1976, p.36). Kohlberg would rate this as the 2nd stage of pre-conventional thinking since it constitutes an “individual considering his own interests and those of other isolated individuals” (p.36). While this form of testing allows for greater room on the part of the individual to express themselves (versus the DIT’s “Great, much, some, little, no” responses to prepackaged statements), there are a number of perceived weaknesses with the MJI method. The first is the argument that participants might be unable or unwilling to really be candid with their moral reasoning while being interviewed. The second is the decidedly poor ability of humans to adequately understand their own cognitive processes, as shown by multiple experiments (Nisbett & Wilson, 1977, Uleman and Bargh, 1989). A third concern is that this form of moral measure has been revised so many times to bring it into consistency with age and invariant stage progression that it no longer really measures what it was intended to measure (Lind, 2008). A final concern involves the researcher; if he or she doesn’t fully understand the levels as expounded by Kohlberg, he or she will consistently rate the interviews incorrectly, resulting in inaccurate data.

Another form of morality measure is the Morality Judgment Test (MJT), developed by Georg Lind between 1975 and 1977 (Lind, 2000a). While sharing some similarities with the DIT from the standpoint of presenting moral dilemmas (two dilemmas in this case), it differs
markedly in what the respondent is supposed to do, as well as the manner in which the results are scored. The dilemmas presented have already had an action; the stories’ protagonists have made their choices. Participants rate the action taken by the protagonist on a -3 to +3 integer scale in terms of their agreement with that course of action. Participants are then directed to review 12 arguments (6 in favor of the action chosen and 6 against), and then rate them on a 9 point scale (-4 to +4) in terms of how much they accept that line of reasoning (+) or reject it (-). As with the DIT, the arguments are tied to the levels of Kohlberg’s theory. Ultimately, a score (C-score in this case) is calculated, and it is said to reflect the level of moral judgment competence, which the inventor defines “as the ability of a subject to accept or reject arguments on a particular moral issue consistently in regard to their moral quality even though they oppose the subject's stance on that issue” (Lind, 2008, p. 200). A bonus of this test is that, like the DIT-2, it can be administered online. Another bonus feature is that it can be administered to individuals with a 5th grade reading level, versus the DIT requirement for a high school reading level (Lind, 2000a).

The main weakness of this test (as compared to the DIT) appears to be its research base: it has a base of 40,000 as compared to the base of half-million for the DIT (Lind, 2000a). Its other weakness is the measure of interest; whereas the MJT measures moral judgment competence (i.e., the ability to judge arguments on the basis of their moral quality), the DIT measures moral orientation and attitudes (Lind, 2000a). In the Lind (2008) study, Georg Lind mentions that while moral attitudes and moral competency have a high degree of correlation, they are not the same thing (Lind, 2008). Since there is an interest in this experiment for both the level of moral development as well as the moral attitudes of the participants involved; and since the DIT seems to require a degree of education in greater proximity to the groups of interest (i.e., 12th grade reading level vice 5th grade), the DIT appears to be a superior test for this experiment.
Hypothesis

The intent of this experiment was to determine whether Navy Sophomores (who took a Leadership and management course and a Navigation course, as well as receiving supplemental role playing instruction) showed changes in moral reasoning as compared to another ROTC unit on campus (Air Force ROTC) and traditional (i.e., non-military) college peers. Moral reasoning (as well as its possible change) was assessed using James Rest’s Defining Issues Test Version 2, which is based on the 6 moral development stages advocated by Lawrence Kohlberg.

Methods

Participants

Participants were comprised of Embry-Riddle Daytona Beach sophomore undergraduate volunteers from 3 groups: a population of Naval ROTC students, a population of Air Force ROTC students, and a population of traditional college students without prior military experience and not affiliated with any ROTC programs. Academic sophomores were considered to be those students with 30 or more credits and at least 2.5 years left in their degree progression. To standardize age, students were excluded if they were older than 22 or younger than 18. Participants were required to list whether English was their first language on the survey, and were excluded from analysis if it was not. The survey was voluntary, and the online site started with a consent form to ensure participants were made fully aware of the goals of the study.

Materials/Apparatus

Participants required access to a laptop or personal computer with access to the internet. On the basis of their groups, as well as the test period, participants were directed to a specific URL where they completed the Defining Issues Test version 2. The site provided instructions on how to complete the test.
The NROTC class (i.e., leadership and management) follows the attached syllabus (i.e., Appendix A). The six textbooks (as cited in the syllabus) are 1) *Leadership: Enhancing the Lessons of Experience 6th Edition*, 2) *MCDP 5 Planning*, 3) *MCDP 6 Command and Control*, 4) *The Armed Forces Officer*, 5) *Leadership Embodied*, and 6) *Leadership and Management*. Presentation style involves consistent use of Powerpoint and lecture. Group activities and student planned lessons are used to keep material varied.

**Design**

The quasi experiment was a 3 x 3 mixed fully factorial design. Participants were comprised of the groups they volunteered from: Naval ROTC, Air Force ROTC, and Traditional (i.e., Non-ROTC) students. Adequate student participation in their respective classes (i.e., by grade and attendance) was ensured by the respective teachers in accordance with Embry-Riddle and ROTC policies. Each group took an initial test at the beginning of the Fall semester, another test at the beginning of the spring semester, and a final test at the end of the Spring semester. The test data provided several dependent measures, of which the two used for further analysis were the P score and N2 score (discussed later on). To track possible confounds from previous ethics courses, the demographics section of the survey questioned participants about whether they had taken a previous ethics class. The results are summarized in the data analysis section.

**Procedure**

The Defining Issues Test is presented using 5 moral dilemmas. For each dilemma, participants specify what their course of action would be. Additionally, they rate the general importance of twelve listed reasoning statements in helping them reach that conclusion (i.e., great, much, some, little, none). They then identify the four of those statements which were the most important in assisting the decision, and rank those four statements against each other (i.e., 1
through 4 in order of importance). The Test’s creator, James Rest, tied each of these twelve statements to a Moral development level as defined by Lawrence Kohlberg’s 6-stage theory of moral development, with differing weights given to the top four choices. Since the students completed the survey outside of class time, they had as much time to finish the survey as they required; mean survey completion times were not tracked. According to the center for the Study of Ethical Development, the survey takes approximately 20-30 minutes to complete on average. The computerized raw data for each of the testing periods was sent to the Center for Ethical Development for review and scoring. Following receipt of scores, statistical analysis was performed to determine if differences in the indices of interest were statistically significant.

Data Analysis

Based on high levels of attrition among traditional student participants between trials (see table 2), that group was excluded from the final analysis. Two overall analyses (comprised of several smaller comparisons) were performed: an analysis of p-scores and N2 scores for NROTC students across all 3 testing trials (N= 7), and an analysis of p-scores and N2 scores for NROTC students (N= 13) and AFROTC students (N=4) who completed just trials 1 and 3. For the first analysis, 3 students indicated previous exposure to an ethics class (either in college or high school). For the second analysis, 4 NROTC and 1 AFROTC student indicated previous exposure to an ethics class. Results for the analyses are summarized below.

Results

In order to facilitate an understanding of the results, a review of the categories of interest is necessary. The DIT-2 raw data provides the necessary numbers to create multiple categories, of which the following were of interest during this study: Stage 2/3 score, Stage 4 score, p-score, and N2 score. A brief of those indices follows.
1. Stage 2/3 represents the personal interest score, wherein the actor is predominantly concerned with simple exchanges that favor themself (i.e., I’ll do it as long as I benefit somehow), as well as the fostering of good simple relationships with other individuals. It represents the composite of Kohlberg’s stage 2 and stage 3 development levels. The focus is egocentric, with the individual predominantly concerned with themself. There is a prevailing concept that morality is something externally imposed rather than internally driven, and the individual seeks extrinsic rather than intrinsic motivation for desired behaviors. Scores in this category range from 0 to 100, with 0 representing no preference for this stage, and 100 representing almost exclusive preference for this moral stage.

2. Stage 4 represents the preference for maintaining the existing system, with an emphasis on obedience to rules for the system’s sake. It represents a shift away from viewing morality as an externally imposed system (as in Stage 2/3), with the person now seeing morality as internally driven. Scores in this category range from 0 to 86, with 0 representing no preference for this stage, and 86 representing almost exclusive preference for this moral stage.

3. The P score represents the preference for post-conventional reasoning, with a focus on pre-societal ideals such as basic human rights and justice. Morality is again viewed as external to the individual, but whereas Stage 2/3 is viewed as externally imposed, post-conventional reasoning is viewed as a series of external mutual agreements (in which the actor has a say) between groups with differing interests. The score can range from 0 to 95, with 0 representing no preference for this stage, and 95 representing almost exclusive preference for this moral stage. College students typically score in the high 30’s to low 40’s in this category.
4. The N2 score has two parts: the extent to which post-conventional items are prioritized (i.e., the P score) as well as the extent to which lower stages (e.g., such as stages 2 and 3) receive lower ratings. The two parts are combined to form one composite score. A simplistic way to view this is as a weighted P score. Because of the relatively new nature of this measure (i.e., it was proposed in 1999), this weighting is useful because it allows p-scores obtained on older studies to be compared with N2 scores on new studies, providing some research continuity.

Online links were created for the survey using the Survey Monkey website, with each group having its own link for each of the three test periods. Each link had its own data collector (i.e., a website administered cache allowing for the responses specific to that URL to be collected independently of responses on the other URLs), with this collector indicating the number of individuals who had logged into the URL to complete the survey. Following completion of the testing window, access to the websites were closed, the data were downloaded into Microsoft excel format, and then sent off to the Center for Ethical Development for evaluation. Data were returned from the center by email as PDF files. The data were further scrutinized, to control for such factors as age (between 18-22 only), level of education (sophomore only), and whether or not English was the primary language of the survey participant (non-native speakers were excluded). For NROTC, the data review (both by the center and internally) resulted in a 26%, 16%, and 12% reduction in viable candidates across the 3 trials respectively. For AFROTC, the reductions were 22%, 16%, and 50%. For traditional students, the reductions were 44%, 50%, and 90%. These reduction numbers represent the total number of participants (in percentage form) who were listed by Survey Monkey as having started the survey, but who were ultimately removed (either by the Center or the researcher) from the list of accepted surveys for that trial,
either due to failure to complete, due to a reliability check of the DIT survey itself (e.g., high “meaningless item” score), or due to the language and age controls listed below. These participants were removed for the listed trial only; attempts by the same participant on other trials were accepted as long as they completed the survey and passed the reliability checks. The results are summarized below in Table 2.

**Table 2. Summary of testing results.**

<table>
<thead>
<tr>
<th>Session 1</th>
<th>Start</th>
<th>End</th>
<th>Done</th>
<th>Cleared</th>
<th>Accepted</th>
</tr>
</thead>
<tbody>
<tr>
<td>NROTC</td>
<td>7-Sep-12</td>
<td>12-Oct-12</td>
<td>34</td>
<td>29</td>
<td>25</td>
</tr>
<tr>
<td>AFROTC</td>
<td>27</td>
<td>23</td>
<td>21</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TRAD.</td>
<td>43</td>
<td>32</td>
<td>24</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Session 2</th>
<th>Start</th>
<th>End</th>
<th>Done</th>
<th>Cleared</th>
<th>Accepted</th>
</tr>
</thead>
<tbody>
<tr>
<td>NROTC</td>
<td>13-Jan-13</td>
<td>2-Feb-13</td>
<td>18</td>
<td>16</td>
<td>15</td>
</tr>
<tr>
<td>AFROTC</td>
<td>6</td>
<td>5</td>
<td>5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TRAD.</td>
<td>6</td>
<td>3</td>
<td>3</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Session 3</th>
<th>Start</th>
<th>End</th>
<th>Done</th>
<th>Cleared</th>
<th>Accepted</th>
</tr>
</thead>
<tbody>
<tr>
<td>NROTC</td>
<td>10-Apr-13</td>
<td>1-May-13</td>
<td>17</td>
<td>17</td>
<td>15</td>
</tr>
<tr>
<td>AFROTC</td>
<td>12</td>
<td>7</td>
<td>6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TRAD.</td>
<td>10</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: “Done” represents the number of participants Survey Monkey recorded as logging onto the group specific URL. “Cleared” represents the number of surveys accepted and graded by the center from the exported excel sheet data. “Accepted” represents the number of surveys cleared by the center that were accepted by the researcher after controlling for age and native language.

As predicted, respondent levels on the 2nd and 3rd iterations for all groups were lower than the 1st iteration. However, the attrition was so extreme as to present serious analysis issues. Of the 3 groups, the traditional students had the greatest level of attrition; only 1 student of the original 43 participants (i.e., 2%) completed all 3 trials. This presented a significant problem in proceeding with the original intended comparison. Based on the data available, a decision was made to remove the Traditional student scores from analysis, and proceed with the following analysis: a scores comparison of the NROTC individuals who completed all 3 testing trials (n=7), and a comparison of those NROTC students who completed trials 1 and 3 (n=13) with their AFROTC counterparts (N=4). Due to the low sample sizes involved, non-parametric analyses were used to analyze effects between trials and groups individually.
NROTC Across 3 Trials

Data was gathered for the indices of interest (i.e., stage2/3, Stage4, P-score, and N2 score) for the 7 NROTC participants who participated in all 3 trials. The raw data is listed below in Table 3, where the prefix for each category (i.e., 1, 2, or 3) indicates the trial number.

Table 3. Raw data for selected NROTC participants.

<table>
<thead>
<tr>
<th>Individual</th>
<th>Stage23</th>
<th>Stage3</th>
<th>Stage4</th>
<th>P score</th>
<th>N2 score</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>30</td>
<td>30</td>
<td>34</td>
<td>21.26</td>
<td>50</td>
</tr>
<tr>
<td>2</td>
<td>32</td>
<td>22</td>
<td>34</td>
<td>39.15</td>
<td>42.08</td>
</tr>
<tr>
<td>3</td>
<td>26</td>
<td>50</td>
<td>20</td>
<td>25.2</td>
<td>24.36</td>
</tr>
<tr>
<td>4</td>
<td>30</td>
<td>40</td>
<td>28</td>
<td>29.21</td>
<td>33.94</td>
</tr>
<tr>
<td>5</td>
<td>46</td>
<td>48</td>
<td>10</td>
<td>10.17</td>
<td>32.34</td>
</tr>
<tr>
<td>6</td>
<td>16</td>
<td>50</td>
<td>30</td>
<td>37.04</td>
<td>42.45</td>
</tr>
<tr>
<td>7</td>
<td>18</td>
<td>30</td>
<td>42</td>
<td>44.7</td>
<td>43.85</td>
</tr>
</tbody>
</table>

Means and standard deviations for the raw data are summarized below in Table 4.

Table 4. Means for NROTC participants in select categories.

<table>
<thead>
<tr>
<th>Trial</th>
<th>Variable</th>
<th>N</th>
<th>Minimum</th>
<th>Maximum</th>
<th>Mean</th>
<th>Std. Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Stage23 score</td>
<td>7</td>
<td>16.00</td>
<td>46.00</td>
<td>28.28</td>
<td>9.96</td>
</tr>
<tr>
<td>1</td>
<td>Stage4 score</td>
<td>7</td>
<td>22.00</td>
<td>50.00</td>
<td>38.57</td>
<td>11.35</td>
</tr>
<tr>
<td>1</td>
<td>P score</td>
<td>7</td>
<td>6.00</td>
<td>42.00</td>
<td>27.71</td>
<td>11.69</td>
</tr>
<tr>
<td>1</td>
<td>N2 score</td>
<td>7</td>
<td>10.17</td>
<td>44.70</td>
<td>29.53</td>
<td>11.84</td>
</tr>
<tr>
<td>2</td>
<td>Stage23 score</td>
<td>7</td>
<td>14.00</td>
<td>34.00</td>
<td>24.00</td>
<td>6.93</td>
</tr>
<tr>
<td>2</td>
<td>Stage4 score</td>
<td>7</td>
<td>24.00</td>
<td>52.00</td>
<td>39.71</td>
<td>10.79</td>
</tr>
<tr>
<td>2</td>
<td>P score</td>
<td>7</td>
<td>24.00</td>
<td>50.00</td>
<td>32.57</td>
<td>9.07</td>
</tr>
<tr>
<td>2</td>
<td>N2 score</td>
<td>7</td>
<td>24.05</td>
<td>43.85</td>
<td>34.72</td>
<td>8.41</td>
</tr>
<tr>
<td>3</td>
<td>Stage23 score</td>
<td>7</td>
<td>8.00</td>
<td>32.00</td>
<td>20.86</td>
<td>9.01</td>
</tr>
<tr>
<td>3</td>
<td>Stage4 score</td>
<td>7</td>
<td>16.00</td>
<td>68.00</td>
<td>42.29</td>
<td>19.09</td>
</tr>
<tr>
<td>3</td>
<td>P score</td>
<td>7</td>
<td>18.00</td>
<td>44.00</td>
<td>32.86</td>
<td>9.15</td>
</tr>
<tr>
<td>3</td>
<td>N2 score</td>
<td>7</td>
<td>24.36</td>
<td>48.25</td>
<td>37.44</td>
<td>7.87</td>
</tr>
</tbody>
</table>
P score and N2 score are directly related to the scores obtained on Stage 2/3 and Stage 4. In the case of P score, the higher the score for an individual in Stage 2/3 and Stage 4, the lower the ultimate score can be for P score. N2 score, on the other hand, is a ratio; the higher a participant rates Stage 2/3 items (which is reflected in higher Stage 2/3 scores), the smaller the p-score will ultimately be. Since N2 is a ratio between P-score and stage 2/3 items, a higher Stage 2/3 score should result in a lower N2 score. As an example, consider Table 3, trial 1, individuals 5 and 7. Individual 5 has relatively high Stage 2/3 and Stage 4 scores, leading to lower P and N2 scores. Individual 7, on the other hand, has lower Stage 2/3 and Stage 4 scores, leading to a higher p-score and N2 score.

Given that Kohlberg postulated post-conventional reasoning (as represented by the P score) was at the top of moral stage development ladder, as well as the study’s interest in determining moral development in the form of positive stage transition (i.e., from lower to higher stages), the N2 and P score indices were selected for additional analysis. Both P score and N2 score were subjected to non-parametric related samples analysis across all 3 trials.

For p-scores across all 3 trials, the data was subjected to a Friedman’s two way analysis of variance by ranks. The resulting p-value of 0.852 indicated that the null hypothesis should be retained, suggesting that there were no differences between Naval ROTC MIDN as a result of time of testing. Figure 1 graphically represents the mean p-scores and standard deviations for the 3 trials. For N2 scores across all 3 trials, the data was subjected to a Friedman’s two way analysis of variance by ranks. The resulting p value of 0.651 indicated that the null hypothesis should be retained, also indicating no effect of time of testing on N2 scores for Naval ROTC MIDN. Figure 2 graphically represents the mean N2 scores and standard deviations for the 3 trials. Overall, difference between the trial times were determined to be not statistically significant for
P score and N2 score, indicating a high degree of likelihood that any perceived variance between trials is due to chance.

Figure 1. P score means with standard deviations.

Figure 2. N2 score means with standard deviations.
NROTC and AFROTC Across 2 Trials

Data was gathered for the indices of interest (i.e., stage2/3, Stage4, P-score, and N2 score) for the 13 NROTC and 4 AFROTC participants who participated in trials 1 and 3. It is interesting to note here that NROTC participants completing trials 1 and 3 were more numerous than the number completing all 3 trials. Intuitively, one might expect participants to complete 1 and 2, and then not trial 3, but this was not the case. The disparity might be due to being busy during trial 2 (i.e., trial 2 was at the start of the spring semester), better incentivizing, or greater engagement of the students on the part of the researcher. Whatever the cause, the two trials present bigger sample sizes than trials 1 and 2 for both groups. The raw data is listed below in Table 5, where the prefix for each category (i.e., 1 or 3) indicates the trial number.

Table 5. Raw data for NROTC and AFROTC trials 1 and 3.

<table>
<thead>
<tr>
<th>Group</th>
<th>1Stage2</th>
<th>1Stage3</th>
<th>1P-score</th>
<th>1N2score</th>
<th>3Stage2</th>
<th>3Stage3</th>
<th>3P-score</th>
<th>3N2score</th>
</tr>
</thead>
<tbody>
<tr>
<td>NROTC</td>
<td>28</td>
<td>38</td>
<td>28</td>
<td>34.56</td>
<td>22</td>
<td>14</td>
<td>60</td>
<td>61.26</td>
</tr>
<tr>
<td>NROTC</td>
<td>30</td>
<td>30</td>
<td>34</td>
<td>21.26</td>
<td>32</td>
<td>34</td>
<td>34</td>
<td>34.8</td>
</tr>
<tr>
<td>NROTC</td>
<td>12</td>
<td>18</td>
<td>54</td>
<td>52.46</td>
<td>34</td>
<td>16</td>
<td>38</td>
<td>47.62</td>
</tr>
<tr>
<td>NROTC</td>
<td>32</td>
<td>22</td>
<td>34</td>
<td>39.15</td>
<td>22</td>
<td>38</td>
<td>38</td>
<td>36.72</td>
</tr>
<tr>
<td>NROTC</td>
<td>28</td>
<td>42</td>
<td>30</td>
<td>36.6</td>
<td>42</td>
<td>34</td>
<td>8</td>
<td>13.77</td>
</tr>
<tr>
<td>NROTC</td>
<td>26</td>
<td>50</td>
<td>20</td>
<td>25.2</td>
<td>32</td>
<td>16</td>
<td>44</td>
<td>32.86</td>
</tr>
<tr>
<td>NROTC</td>
<td>24</td>
<td>34</td>
<td>26</td>
<td>29.33</td>
<td>24</td>
<td>42</td>
<td>24</td>
<td>26.29</td>
</tr>
<tr>
<td>NROTC</td>
<td>30</td>
<td>40</td>
<td>28</td>
<td>29.21</td>
<td>16</td>
<td>38</td>
<td>40</td>
<td>48.25</td>
</tr>
<tr>
<td>NROTC</td>
<td>46</td>
<td>48</td>
<td>6</td>
<td>10.17</td>
<td>14</td>
<td>68</td>
<td>18</td>
<td>24.36</td>
</tr>
<tr>
<td>NROTC</td>
<td>38</td>
<td>12</td>
<td>48</td>
<td>35.01</td>
<td>52</td>
<td>12</td>
<td>28</td>
<td>29.89</td>
</tr>
<tr>
<td>NROTC</td>
<td>20</td>
<td>58</td>
<td>22</td>
<td>30.76</td>
<td>56</td>
<td>28</td>
<td>56</td>
<td>39.35</td>
</tr>
<tr>
<td>NROTC</td>
<td>16</td>
<td>50</td>
<td>30</td>
<td>37.04</td>
<td>8</td>
<td>68</td>
<td>24</td>
<td>41.41</td>
</tr>
<tr>
<td>NROTC</td>
<td>18</td>
<td>30</td>
<td>42</td>
<td>44.7</td>
<td>22</td>
<td>34</td>
<td>32</td>
<td>43.7</td>
</tr>
<tr>
<td>AFROTC</td>
<td>26</td>
<td>38</td>
<td>36</td>
<td>34.25</td>
<td>14</td>
<td>48</td>
<td>36</td>
<td>40.42</td>
</tr>
<tr>
<td>AFROTC</td>
<td>26</td>
<td>26</td>
<td>40</td>
<td>43.05</td>
<td>14</td>
<td>56</td>
<td>30</td>
<td>37.5</td>
</tr>
<tr>
<td>AFROTC</td>
<td>26</td>
<td>18</td>
<td>54</td>
<td>56.19</td>
<td>14</td>
<td>30</td>
<td>50</td>
<td>51.52</td>
</tr>
<tr>
<td>AFROTC</td>
<td>26</td>
<td>52</td>
<td>20</td>
<td>34.86</td>
<td>18</td>
<td>58</td>
<td>24</td>
<td>29.81</td>
</tr>
</tbody>
</table>
Means and standard deviations were determined for the raw data. The results are summarized in Table 6.

**Table 6. Means for AFROTC and NROTC selected indices on trials 1 and 3.**

<table>
<thead>
<tr>
<th>Group</th>
<th>1Stage2</th>
<th>3Stage2</th>
<th>1Stage4</th>
<th>3Stage4</th>
<th>1P score</th>
<th>3P score</th>
<th>1N2 score</th>
<th>3N2 score</th>
</tr>
</thead>
<tbody>
<tr>
<td>NROTC</td>
<td>Mean</td>
<td>26.77</td>
<td>25.08</td>
<td>36.31</td>
<td>36.77</td>
<td>32.00</td>
<td>32.73</td>
<td>36.94</td>
</tr>
<tr>
<td></td>
<td>N</td>
<td>13.00</td>
<td>13.00</td>
<td>13.00</td>
<td>13.00</td>
<td>13.00</td>
<td>13.00</td>
<td>13.00</td>
</tr>
<tr>
<td></td>
<td>Std. Deviation</td>
<td>9.18</td>
<td>13.16</td>
<td>13.68</td>
<td>19.88</td>
<td>12.35</td>
<td>12.91</td>
<td>10.57</td>
</tr>
<tr>
<td>AFROTC</td>
<td>Mean</td>
<td>26.00</td>
<td>15.00</td>
<td>33.50</td>
<td>48.00</td>
<td>37.50</td>
<td>35.00</td>
<td>42.09</td>
</tr>
<tr>
<td></td>
<td>N</td>
<td>4.00</td>
<td>4.00</td>
<td>4.00</td>
<td>4.00</td>
<td>4.00</td>
<td>4.00</td>
<td>4.00</td>
</tr>
<tr>
<td></td>
<td>Std. Deviation</td>
<td>0.00</td>
<td>2.00</td>
<td>14.82</td>
<td>12.75</td>
<td>13.99</td>
<td>11.14</td>
<td>10.22</td>
</tr>
</tbody>
</table>

As indicated above, p-scores and N2 scores are directly related to scores on Stage 2/3 and Stage 4. P-score and N2 scores for each group were selected for analysis. Both P score and N2 score were subjected to non-parametric analysis across the 2 trials.

For p-score, trial 1 p-scores were compared between NROTC and AFROTC. Based on a Mann-Whitney U test p-value of 0.364, the null hypothesis was retained, suggesting no difference between military branches at the start of the study. Trial 3 p-scores were also compared between NROTC and AFROTC. Based on a Mann-Whitney U test p-value of 0.733, the null hypothesis was retained. This outcome reveals that no difference was found in post-conventional reasoning between the two branches at the end of the testing an academic year later. NROTC p-scores were then evaluated using a non-parametric related samples analysis. Based on a Wilcoxon signed rank test p-value of 0.844, the null hypothesis was retained. Therefore, no difference was found in Naval ROTC MIDN in post-conventional thinking between the start of the study and at the third measurement a year later. AFROTC p-scores were then evaluated using
a non-parametric related samples analysis. Based on a Wilcoxon signed rank test p-value of 0.414, the null hypothesis was retained. No difference was found for AFROTC in post-conventional thinking between the start of the study and the third measurement. Figure 3 represents P score means for each group by trial, with standard deviations shown.

![NROTC and AFROTC mean P scores by trial](image)

**Figure 3. P score means with standard deviations for NROTC and AFROTC on trials 1 and 3.**

For N2 score, trial 1 N scores were compared between NROTC and AFROTC. Based on a Mann-Whitney U test p-value of 0.213, the null hypothesis was retained, suggesting no difference between military branches at the start of the study. Trial 3 N2 scores were also compared between NROTC and AFROTC. Based on a Mann-Whitney U test p-value of 0.651, the null hypothesis was retained. This outcome reveals that no difference was found in prioritizing post-conventional reasoning at the expense of stage 2/3 reasoning between the two
branches at the end of the testing an academic year later. NROTC N2 scores were then evaluated using a non-parametric related samples analysis. Based on a Wilcoxon signed rank test p-value of 0.249, the null hypothesis was retained. Therefore, no difference was found in Naval ROTC MIDN in prioritizing post-conventional thinking at the expense of stage 2/3 reasoning between the start of the study and at the third measurement a year later. AFROTC N2 scores were then evaluated using a non-parametric related samples analysis. Based on a Wilcoxon signed rank test p-value of 0.715, the null hypothesis was retained. As with NROTC, no difference was found for AFROTC in prioritizing post-conventional thinking at the expense of stage 2/3 reasoning between the start of the study and the third measurement. Figure 4 represents N2 score means for each group by trial, with standard deviations shown.

Figure 4. N2 score means with standard deviations for NROTC and AFROTC on trials 1 and 3.
Discussion

Given the extremely low subsequent completions of the survey by traditional students, they were excluded from the analysis. The discussion, then, centers around those aspects related to ROTC students (both AFROTC and NROTC). If a more significant sample size could be obtained and retained over subsequent testing for traditional students, the data might allow for better comparison between traditional student instruction and ROTC student instruction.

For ROTC students, no statistically significant trends were found. Based on the measurements being not statistically significant, it is likely that any variance seen within groups between test periods is due to chance. One logical conclusion, then, is that the developmental program used at ERAU NROTC (or at least the sophomore portion of it) does not produce measurable changes in post-conventional reasoning and the rejection of pre-conventional reasoning, as measured by the DIT-2 P score and N2 score respectively. One conclusion might suggest that the class room instruction, as well as that received at PT and Military leadership lab, is ineffective in producing measurable changes in post-conventional reasoning and the rejection of pre-conventional reasoning in ROTC sophomores. Alternatively, these instruction periods might produce a measurable change if they were the only things a student was exposed to (e.g., picture an ROTC student in a social vacuum), but the permanence of that change might be adversely affected by some other activity occurring within the sphere of college life for sophomore students. As an example, consider the individual who spends one to two hours a day in an atmosphere of instruction which fosters moral growth, but who spends the remaining 14 to 15 hours (i.e., given a 9 hour sleep schedule) in an atmosphere which in some cases might not reinforce that instruction, and in other cases might act in direct contradiction to it. Given this
case, even a fairly robust and effective ethics program might have difficulty producing positive measurable changes in moral growth.

In addressing the first conclusion (i.e., program ineffectiveness at producing changes on the DIT-2), there is room to speculate that the DIT-2 might be inappropriate for measuring moral development change. As discussed in the introduction, both Kohlberg’s stages and the DIT are not without their critics. However, the DIT appears to be the most widely used instrument for this purpose, indicating a reasonable likelihood for measurement validity. Additionally, as discussed in the introduction section, the other moral development measurement tools (e.g., the MJT, the MJI, the SRM) all appear to measure different aspects of morality or cognition, making them inappropriate for this study. The DIT-2, then, was the right measurement tool for this study. The lack of significant results suggests the likelihood that there was a lack of moral growth in terms of the indices measured by the test, not that the test was the inappropriate tool.

As concerns a counter-balanced change, the potential differences in groups may have been off-set due to the manner of survey distribution. The survey is available in both paper format and online, allowing it to be administered either in a classroom or over the internet. While internet distribution allows for ease of distribution and collection, it is not without its failings. Hartshorne, May, and Shuttleworth (1930) indicated that results for morality tests can vary widely depending on where the test was taken (e.g., at home, in a classroom, at church, or at a club), and with the widespread availability of internet access (i.e., both at the school and in town), it is difficult to say where the participants in this study completed the test. It is even possible (although there is no substantial reason to suspect) that someone other than the intended participant completed the survey, using the intended participant’s email. This would require conscious deception on the part of the student, so it seems highly unlikely. Still, insomuch as the
online administration allowed participants to take the survey whenever they had time and access to an internet connection, any differences between the groups may have been offset by the effects of different testing areas. Additionally, a participant may have completed the survey at different locations for different trials, which may have affected his or her results. The magnitude of this effect on these groups is difficult to determine, as is the frequency of its occurrence. A future study might be more effective if it were to choose one location and time-frame for all test periods. This would limit the generalizable nature of the results, but might improve the surety of the moral growth conclusions from those results, if only for Embry-Riddle NROTC.

The absence of a difference could also indicate that the samples in question are not representative of their respective populations, specifically in terms of their mental acumen and ability to grow morally. For ROTC students (i.e., both Navy and Air Force) this is unlikely; the selection process algorithm for cadet entry into the forces is used nation-wide, so the criteria for entry and retention are the same for each ROTC unit. Student applications are selected from all over the country, with the unifying selection characteristics being related to high ACT or SAT scores, high unweighted GPAs, and involvement in community affairs and high school leadership positions. The only thing unique aspect about these sample groups that is immediately noticeable is choice; whether as a NROTC scholarship student, or as a student paying his or her own way, every student at ERAU ROTC (both NROTC and AFROTC) would have had to indicate that Embry-Riddle Aeronautical University was one of his or her choices. There is, then, no reason to suggest that these students are less gifted academically, or less able to grow morally than any other group undergoing this same form of training/education. In fact, the mean scores for the two groups (i.e., NROTC and AFROTC) for the categories of interest are consistent with means obtained in multiple studies using the DIT-2. This is shown in Table 7, which is taken
directly from page 35 of the DIT-2 guide, and is available for purchase from the Center for Ethical Development (Bebeau & Thoma, 2003, p.35).

Table 7. Means, standard deviations, and sample sizes for DIT-2 participants by education level.

<table>
<thead>
<tr>
<th>Educational Level</th>
<th>Personal Interest (Stage 2/3)</th>
<th>Maintain Norms (Stage 4)</th>
<th>Post Conventional (P score)</th>
<th>N2 Score</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Standard Deviation</td>
<td>N</td>
<td>Mean</td>
</tr>
<tr>
<td>Grade 7-9</td>
<td>33.21</td>
<td>14.41</td>
<td>37</td>
<td>41.69</td>
</tr>
<tr>
<td>Grade 10-12</td>
<td>28.25</td>
<td>12.62</td>
<td>667</td>
<td>33.24</td>
</tr>
<tr>
<td>Voc/Tech</td>
<td>24.87</td>
<td>12.20</td>
<td>111</td>
<td>37.55</td>
</tr>
<tr>
<td>Jr. College</td>
<td>26.27</td>
<td>12.39</td>
<td>236</td>
<td>37.32</td>
</tr>
<tr>
<td>Freshman</td>
<td>28.53</td>
<td>12.32</td>
<td>2,096</td>
<td>33.57</td>
</tr>
<tr>
<td>Sophomore</td>
<td>29.27</td>
<td>12.35</td>
<td>1,028</td>
<td>32.36</td>
</tr>
<tr>
<td>Junior</td>
<td>27.36</td>
<td>12.77</td>
<td>1,333</td>
<td>32.93</td>
</tr>
<tr>
<td>Senior</td>
<td>24.80</td>
<td>12.51</td>
<td>2,441</td>
<td>32.40</td>
</tr>
<tr>
<td>MS degree</td>
<td>21.69</td>
<td>11.82</td>
<td>853</td>
<td>32.64</td>
</tr>
<tr>
<td>Prof. degree</td>
<td>19.76</td>
<td>11.28</td>
<td>1,582</td>
<td>31.41</td>
</tr>
</tbody>
</table>

Total: 25.48 (2.71) 10,553 32.73 14.00 10,553 36.74 16.05 10,553 35.67 16.23 10,553

Note: Of the 13,386 valid responses, 1134 (8.5%) were excluded because they did not report their educational level, and 113 (8.8%) were excluded who reported their educational level as Grade 1-6 (n = 5) or as ‘other’. Finally, of the12,139 responses, 1586 (13%) were excluded because English was not their native language.

Assuming, then, that the sample groups are representative of their populations at large, an alternative explanation for a lack of measurable change might be related to the efficacy of the ROTC instruction. This is not to say that the material or its manner of introduction is insufficient to produce measurable change in and of itself, although this may certainly be the case. Rather, perhaps the pervasiveness of the instruction, either in the form of its duration in relation to the whole college experience or in its intensity as related to the student’s receptiveness, is insufficient to produce measurable changes on the DIT-2. From a duration standpoint, it should be noted that NROTC students spend the majority of their academic day as traditional students. In any given week, an ROTC student will spend 3 hours in physical training, approximately 3 hours in ROTC related classes, and 2.5 hours in Military leadership lab. If 1.5 hours are added for various activities related to ROTC (e.g., transit, planning), the average ROTC student will spend about 10 hours a week on ROTC. Assuming (as an example) the student keeps a nominal
waking schedule of 0500-2200 on PT days (i.e., Monday, Wednesday, and Friday), and a less rigorous schedule of 1000-2200 on the other days, they will be awake for approximately 100 hours of week, so ROTC will only occupy approximately 10% of their time. This level of exposure, however well-directed and well-intentioned it may be, might be insufficient if the activities occupying the remaining time either do not facilitate this process or act in direct contravention to it. Simply put, of the great number of events that occupy an ROTC student during their sophomore year at Embry-Riddle, ROTC events account for a very small portion. If moral progress in terms of post-conventional reasoning is considered a valuable end for the ROTC program, two immediate solutions to the time dilemma come to mind. One solution is to increase the pervasiveness of the program for students through more frequent exposure to ROTC activities. Schools such as the US Naval Academy, Norwich University, and the Citadel have more militarily oriented education programs, so there is a more continuous exposure to that type of instruction. If ROTC instruction does lead to moral growth, increased exposure might help overcome contradictory forces at work in other aspects of the college experience. A study comparing Military academies to traditional ROTC schools in terms of DIT increases might be able to more fully answer this question. The second solution to increasing exposure is perhaps more realistic: continued moral emphasis in the form of moral instruction (i.e., classes) throughout all 4 years of the ROTC program. Since it is unlikely that those non-military schools hosting ROTC units (e.g., Embry-Riddle) will voluntarily consent to become more militant, including additional moral development classes (above and beyond NSC201 and NSC402) as part of the NROTC curriculum may help to increase moral growth.

Pervasiveness aside, from an intensity and receptiveness standpoint, the timing and intensity of instruction may not be ideal. Moral growth is a cognitive structuring and
restructuring process, involving both implicit and explicit processes. That being said, anything that decreases cognition during an instruction period is likely to lead to decreased effectiveness of instruction in terms of restructuring. For example, one of the periods of instruction is PT, which takes place from 0545-0700 three mornings a week. This happens to be a period when many people would still be asleep or in a mental trough (Leger, 1994; Van Dongen & Dinges, 2000), so the mental alacrity and receptiveness of the ROTC student in question might be significantly diminished. As concerns classroom instruction, Sophomore Navy ROTC classes take place on the same 3 days as PT, just at a later time. While the instruction is certainly set up to be engaging, it still relies on some of the traditional delivery methods found in other classes: oral lectures and powerpoint briefs. Therefore, while the instruction takes place in an ROTC dedicated classroom, it has many of the hallmarks of a traditional class. This traditional approach, coupled with the fact that the students will have been awake since 0500, may decrease absorption of the material by the student. If the student is not explicitly engaged, the moral growth that can be expected from moral instruction is likely to be small or non-existent. A study involving this same instruction with either differing sleep schedules, or more engaging teaching practices might be more able to assess the validity of this concern.

Coupled with the concern of pervasiveness and student receptivity is the idea of material retention. Put another way: the combination of classroom and external instruction was sufficiently stimulating, but its effects were lost during the semester when classroom instruction was more academic and less based on morality. While possible, this seems unlikely for a number of reasons. First, Navy ROTC compared to itself across 3 trials showed no statistically significant gains (or losses) between trials 1 and 2, 2 and 3, or 1 and 3. Second, Air Force ROTC showed no statistically significant differences as compared to Naval ROTC on either trial 1 or 3.
While it is possible that a statistically significant change might have occurred for AFROTC during trial 2, it seems unlikely.

Given that the study’s initial aim was to compare NROTC development to traditional student development, the exclusion (due to attrition) of traditional students in the final analysis is unfortunate. Table 2 bears out that the largest attrition (except for the case of trial 3 for NROTC) occurred due to screening by the Center for Ethical Development. The raw data itself shows that the predominant reason for purging results was that the participant didn’t complete the survey. Therefore, future studies of this sort of comparison should attempt to achieve larger ending completion rates for traditional students, either through reaching out to a larger base, or through conducting the surveys in a classroom so that the participant will feel some compulsion to complete the full survey.

If the problem of retention could be adequately addressed, a study designed to assess the moral growth of students during the whole ROTC program might be of interest. The Wattendorf (1981) study made a claim about ROTC development on the basis of comparing entry level ROTC students (i.e., Freshmen) with their senior counterparts. This relies heavily on the assumption that the senior group was equal to the freshmen group at the start, which may not have been the case. If sufficient time were available to the researcher, a longitudinal study assessing ROTC students from their Freshman year through program completion might be better able to assess the value of ROTC training on moral development. Such a study would be necessarily intensive, since it would prove difficult to attribute the moral development of 4 years to any specific element (ROTC instruction or otherwise), given the multitude of experiences in a typical college education. However, if it were to be undertaken, it might provide the necessary groundwork to more readily answer whether an ROTC education is fundamentally different than
a traditional college education in terms of moral development. It might also be able to answer another question: what is the latency between instruction and actual moral growth? A longitudinal study might be able to assess whether moral growth occurs quickly as a result of moral instruction, or occurs (and is measurable) for some time after the moral intervention has ceased.

All the concerns above being considered, the logical answer appears to be the correct one: after one academic year of education, there was no measured moral growth (i.e., in terms of the DIT) for the given populations. Given the wide use and acceptance of the DIT in moral development studies, it was the clear instrument to use. There is no reason to believe that the experiences of college life at ERAU are vastly different from the colleges reviewed in Sclaefi et al. (1985), where studies using the DIT showed measurable gains in moral growth. While daily and weekly exposure to ROTC might seem small in comparison to other activities, it lasts for the whole year, so there was certainly enough time for moral development to occur. Additionally, it would have been going on for a year for the participants prior to the start of this study. Natural troughs in cognition due to circadian rhythms might decrease the benefits received from 0545-0700, but that would not encompass the class times the students had, nor their leadership lab time period. Additionally, based on NSC201 and the leadership lab meeting the specifications of the Schlaefi et al. (1985) “Personality Development” class, the instruction should have been sufficiently engaging to produce modest but consistent results. Also, the absence of significant trends in NROTC across 3 trials, as well as between NROTC and AFROTC trials 1 and 3 argues against the probability of a moral gain (or loss) that was subsequently negated. Lastly, the low p-values obtained during analyses, coupled with the use of non-parametric statistics to compensate for low participant retention, makes it unlikely that a possible effect was masked due to high
variance and low participant numbers. In the end, the ROTC education received appears to mirror the instruction received in “Academic courses” as discussed in Schlaefi et al. (1985), whereby small gains in moral growth can occur, but are not guaranteed.

While the results suggest that this year of instruction did not produce measurable growth in moral reasoning, it might be premature on several counts to suggest that ROTC as a whole will not achieve that goal. The first reason is the number of years in the program, with a typical ROTC course of instruction lasting 4 years. Moral growth appears to be positively correlated to education, so a stagnant or unproductive year might be easily off-set by 3 productive years. The next reason is the NROTC capstone course: NSC402 Leadership and Ethics. This course occurs 2nd semester senior year, and is in line with the “Dilemma Discussion” courses discussed in Schlaefi et al. (1985), which produced consistently the highest increases in p-scores of all the course types reviewed. Lastly, the results of one study might produce a certain conclusion, but that conclusion might be more reliable if the results were replicated in other studies. If multiple studies were to come to the same conclusion, then there might be grounds for suggesting the ineffectiveness of ROTC at producing moral growth.

The results obtained also suggest some things about the DIT-2 which are not new, but are worth pointing out. First, the DIT-2 is said to have a test-retest correlation of 0.7-0.8, meaning that scores can vary for a participant simply between consecutive test periods. Table 3 and Table 5 in the results section bear this out, in that any given participant has a fair degree of variance in scores for the categories listed. Second, the DIT-2 requires a high degree of motivation to complete, which may be due to its length or its complexity. Table 2 in the results section shows a high degree of purged results between surveys started and those accepted, and the raw data shows this to be predominantly due to failure to finish. Finally, the DIT-2 can provide useful
information about a sample population as it relates to samples with similar characteristics. While the results of this study suggest that no moral growth occurred, it also suggests that the students in question are on par with their peers in terms of the categories of interest. For example, the average NROTC p score across the 3 trials was 31.05 with an average standard deviation of 9.97, which compares well to the DIT-2 validated score for college sophomores of 32.62 with a standard deviation of 14.77, as shown in table 7. The same is true of the N2 scores across 3 trials, as well as the p score and N2 scores obtained for AFROTC. The DIT-2, then, appears to be a good measure of moral attitudes, and can be useful to the researcher as long as the capabilities and limitations are understood.
References


Naval Service Training Command, United States Navy (2012). *Regulations for Officer Development (Rod) for the Nava Reserve Officer’s Training Corps (NROTC)(NSTCINST 1533.2A).* 85-90.


Instructor Information

CDR Steven Cincotta
NROTC Commanding Officer/Professor of Naval Science
(386) 323-8993 work
cincotts@erau.edu
ROTC Bldg Room 205

LT Paul Gillett
NROTC Training Officer/Assistant Professor of Naval Science
(386) 323-8995
gillettp@erau.edu
ROTC Bldg Room 204

Office Hours: As posted outside of office / per Outlook request.

COURSE POLICY STATEMENT

Purpose: Welcome! This course requires your leadership, preparation, and participation. You and your fellow students will prepare and lead many of the class sessions and your preparation for class is critical in our discussion-based seminar format.

Leadership is the most challenging and rewarding aspect of being a Naval Officer. While learning to drive a ship, operate a submarine, or fly an airplane is exciting and demanding, you will discover that it is much more difficult to become an effective leader than it is to become proficient at your particular craft. Leadership means dealing with human behavior, and people are much more complex than mere machines. In addition, the Navy is a dynamic organization that must deal with new and complicated leadership issues in an effective and forthright fashion. Over the past 25 years, the end of the Cold
War world has changed the Navy’s mission, its budget, and its manning. A new social awareness has changed the role of women in combat and caused the Navy to rethink some of its core values. Today, we are witnesses to changes to the “Don’t Ask, Don’t Tell” legislation and the removal of sexual orientation as a bar to service. To prepare you for the leadership challenges in the Navy, the Department of Naval Science offers you two leadership and management courses. They will introduce you to the theory and practice of leadership, ethics, and resource management.

L&M is a comprehensive study of organizational behavior and management with a special emphasis on situational leadership in the Navy. This semester, you will explore a variety of leadership and management topics including the classical theories of management, motivation, and communication.

In this course, you will develop your skills in organizational thinking and leadership problem solving. Your text discusses these skills in the context of four categories: Leadership is a Process - not a Position; The Leader; The Followers; and The Situation. We will examine leadership and management challenges in the context of the interdependence of and interactions between those four perspectives, using historical case studies and your missions in the NROTC Battalion and in the Fleet.

The content of this demanding course in leadership has been designed to develop skills and knowledge that will serve you well in the Fleet and beyond. This is the most challenging and potentially valuable course you will take in NROTC.

**Primary Texts**


**MCDP 5 Planning;** Krulak, C. C., General, USMC, Washington DC: 1997. PCN 142 000004 00 (On Blackboard)
MCDP 6 Command and Control; Krulak, C. C., General, USMC, Washington DC: 1996. PCN 142 000001 00 (On Blackboard)


Grading Policy

<table>
<thead>
<tr>
<th>%</th>
<th>Due Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>20%</td>
<td>Per Syllabus</td>
</tr>
<tr>
<td>15%</td>
<td>Oct, Nov</td>
</tr>
<tr>
<td>15%</td>
<td>Oct</td>
</tr>
<tr>
<td>15%</td>
<td>Per assignment</td>
</tr>
<tr>
<td>10%</td>
<td>Cumulative</td>
</tr>
<tr>
<td>15%</td>
<td>7 Dec</td>
</tr>
</tbody>
</table>

Total 100%

Reading Info Papers: Write a one-page summary detailing your understanding and individual opinion regarding the required reading for the topic discussed. Format is “Info Paper Format,” provided on blackboard. Papers are due on the day the theory will be presented in class.

Short Papers (format for both papers is an info paper format, like the homework only ~2 pages each)

Short Paper #1

Using the qualities highlighted during this course, write a 2-page paper reflecting on your leadership strengths and weaknesses. Include an example of an experience that helped you come to these conclusions. Additionally, using the provided web-based site determine which personality type you are. Discuss your Personality Type and how it has impacted your life choices to date.

http://www.personal.psu.edu/faculty/jj/5/j5j/IPIP/

Short Paper #2

Two-page info paper on an aspect of the curriculum section we have been studying(The Process; The Leader; The Followers; and The Situation) applied to a specific leadership/management case. Bibliography and endnotes are to be included for all papers on a page separate from the written work. This is not
just a summary of a leader’s biography (Lt Gen “Chesty” Puller) or a situation (Battle of Midway), nor is it a paraphrasing of Wikipedia or the texts.

**General guidance for papers**

**A-Range:** This work is insightful. It addresses the assignment in a way that indicates your comprehension of the assignment itself as well as an understanding of the underlying concepts. *Several key concepts from the course are consistently applied to the assignment, citing references where needed.* The message is communicated clearly, concisely, and directly and indicates an obvious amount of research.

**B-Range:** This work meets and, at times, exceeds the basic requirements of the assignment. The work indicates that you are beginning, at times, to think through and deal with the major ideas of the assignment. *One or a few key concepts from the course are occasionally applied to the assignment, with some citing of references.* The message is communicated with generally effective clarity, directness, and conciseness. Some unevenness in writing may be apparent.

**C-Range:** While this work offers little insight into the greater concepts of the assignment, it meets the basic requirements. The message, for the most part, is reasonably clear, concise and direct, although there may be unevenness in the writing. *If you throw the assignment together using whatever comes to mind with minimal application of course concepts or only cursory citation of references, this is the highest grade you can hope to possibly get.* Alternatively, this may be very insightful work obscured by substandard writing.

**D:** The basic requirements of the assignment are partially met; however, the message is not always communicated clearly. There is considerable unevenness in the writing.

**F:** The requirements of the assignment have not been met at a satisfactory level. It is not clear that you have understood the concepts of the assignment. The writing is not clear, concise or direct.

**General Considerations:** Written communication skills are a fundamental requirement for leadership success in the military. Therefore, proper GRAMMAR and PUNCTUATION as well as correct
SPELLING are critical. Have someone you trust (friend, roommate, classmate) proofread your papers before you hand them in. Your computer spell check and grammar check features are not enough to prevent errors. Your paper should be a "final product," something you would hand to your future commanding officer. Your written word will represent you, so always ensure your product is the best you can provide.

As a junior officer, your writing will often make the most significant impact on your commanding officer's impression of you, especially in large commands. Ensure your paper has a point and that it makes that point but does not belabor it. You do not have to be brilliant; just express your ideas clearly, concisely, and with impact!

**Class Facilitation:** See specific Student Facilitations Guide provided separately on Black Board.

**Quizzes:** The format is 10 questions, covering topics since the last quiz or exam, mostly short answer, multiple choice, fill in the blank, and true/false.

**Midterm:** The mid-term will be given during a regular class period in accordance with the class schedule. The exam will consist of short answer, multiple choice, and essay questions. Both the mid-term and final are cumulative.

**Final Examination:** The final will be a take home exam due the last day of class.

**Late Work**

The score received on a late assignment will be reduced by 10 points for each day the assignment is late. For example, if an item is due @ 1030 on a Monday for Section 1, then any late assignments turned in between 1031 Monday and 1030 Tuesday will be reduced by 10 points. An assignment turned in on 1031 Tuesday will lose 20 points. Thanks to the glories of email, the weekends & holidays count. If an assignment is due 2 Sep at 1030, then 12 Sep at 1031 zero points are possible for that assignment.
Cheating and Plagiarism

Basically, if you cheat / plagiarize, be prepared to fail the course, go to a PRB, and face university action (and if active duty, don’t forget the UCMJ and possibility of leaving the STA-21/MECEP programs!).

All work assigned outside of class may be freely discussed between students, and you are encouraged to study together. All work submitted for grade, however, shall have originated through a student’s individual effort and (unless otherwise cited or assigned as a group project) shall be original to the individual student or group. If you have any doubt whatsoever - ask.

All papers submitted in this course are subject to review through TURNITIN.COM where the text of the paper is compared to works compiled in their data base. Reviewed papers will automatically become part of this database, serving as source documents for the purpose of detecting plagiarism.

Use of Laptops

Laptops may be used in class EXCLUSIVELY for the purpose of taking notes regarding the class in session.

Attendance and Classroom Behavior Policy

Class participation will count towards the final grade. Punctuality is expected. A tardy student should enter the classroom without disruption and will see the instructor after class to explain tardiness.

Apart from the military courtesies extended to the instructors by NROTC students, the classroom behavior of all students should
be collegial. My intent is to run this class seminar style, which means students will be encouraged and expected to speak freely and challenge each other’s ideas and comments. Side-conversations should be kept to a minimum, solely for clarification or repetition of a missed point from fellow classmates, but discussions of subject matter will be shared with the entire class.

Discussions of controversial subject matter may arise in class. Students’ candid opinions are a valued contribution in an academic setting, however, inflammatory or offensive comments, bigotry, sexual, ethnic or racial slurs, avocation of illegal action, etc. will not be tolerated. A student (and/or the class as a whole) will immediately drop any such line of discussion when instructed to do so by the instructor or requested by any student. Any offended party is obligated to inform the instructor, either in public or in private.

Any views expressed by the instructor, unless specifically attributed otherwise, should be considered the personal views of the instructor and may not be representative of any official policy or viewpoint of the university, the US Navy, or the government.

**COURSE OUTLINE:**

<table>
<thead>
<tr>
<th>DATE</th>
<th>TOPIC</th>
<th>READING</th>
<th>ASSIGNMENT</th>
</tr>
</thead>
<tbody>
<tr>
<td>I. LEADERSHIP IS A PROCESS, NOT A POSITION</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(M) 27 AUG</td>
<td>COURSE INTRODUCTION N/A</td>
<td></td>
<td>NO HW</td>
</tr>
<tr>
<td>(W) 29 AUG</td>
<td>WHAT IS LEADERSHIP? CH 1</td>
<td>IP 1 DUE</td>
<td></td>
</tr>
<tr>
<td>(F) 31 AUG</td>
<td>LEADERSHIP INTERACTIVE CH 2</td>
<td>IP 2 DUE</td>
<td></td>
</tr>
<tr>
<td></td>
<td>John Paul Jones LE CH 2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(M) 03 SEP</td>
<td>NO CLASS – LABOR DAY</td>
<td></td>
<td>NO HW</td>
</tr>
<tr>
<td>(W) 05 SEP</td>
<td>CLASS EXERCISE</td>
<td></td>
<td>NO HW</td>
</tr>
</tbody>
</table>
II. FOCUS ON THE LEADER

(M) 17 SEP (8) POWER AND INFLUENCE
William A. Moffett
LE CH 17
***Group 3***

(W) 19 SEP (9) LEADERSHIP AND VALUES
Archer A. Vandergrift
LE CH 24
***Group 4***

(F) 21 SEP (10) ARMED FORCES OFFICER I
AFO CH 3, 5, 8, App 3
IP 8 DUE

(M) 24 SEP (11) LEADERSHIP TRAITS
Oliver Hazard Perry
LE CH 5
***Group 5***

(W) 26 SEP (12) LEADERSHIP BEHAVIOR
Merritt A. Edson
LE CH 36
***Group 6***

(F) 28 SEP (13) BASIC LEADERSHIP SKILLS
Lewis B. Puller
LE CH 18
IP 11 DUE
** Take personality test located at http://www.personal.psu.edu/faculty/j/5/j5j/IPIP/ **

III. FOCUS ON THE FOLLOWERS

(F) 05 OCT (15) MOTIVATION, SATISFACTION, CH 9
PERFORMANCE
Marc A. Mitscher LE CH 27 IP 12 DUE
***Group 7***

(M) 08 OCT (16) GROUPS AND TEAMS CH 10 IP 13 DUE
Slade D. Cutter LE CH 20
***Group 8***

(W) 10 OCT (17) BASIC LEADERSHIP SKILLS pgs. 481-536
Leighton W. Smith LE CH 49 IP 14 DUE

(F) 12 OCT (18) CLASS EXERCISE SP 1 DUE

IV. FOCUS ON THE SITUATION

(M) 15 OCT (19) CHARACTERISTICS OF THE SITUATION
Grace Murray Hopper LE CH 41 IP 15 DUE
***Group 9***

(W) 17 OCT (20) CONTINGENCY THEORIES CH 12 IP 16 DUE
***Group 10***

*19–22 OCT* NO CLASS – STUDENT FALL BREAK

(W) 24 OCT (21) LEADERSHIP & CHANGE CH 13 IP 17 DUE

(F) 26 OCT (22) CLASS EXERCISE QUIZ #2
(M) 29 OCT  (23) BASIC LEADERSHIP SKILLS pgs. 665-695  IP 18 DUE

(W) 31 OCT  (24) CLASS EXERCISE  LM CH 28  IP 19 DUE
CASE STUDY: Sir Ernest Shackleton

(F) 02 NOV  (25) CLASS EXERCISE  NO HW

(M) 05 NOV  (26) CRIMSON TIDE Part I  AFO CH 1  IP 20 DUE
Charles B. MacVay III  LE CH 30

(W) 07 NOV  (27) CRIMSON TIDE Part II  AFO CH 4  IP 21 DUE
Carl M. Brashear  LE CH 42

(F) 09 NOV  (28) CRIMSON TIDE Part III  AFO CH 6, 7, 9  IP 22 DUE

(M) 12 NOV  ****VETERAN’S DAY HOLIDAY****

(W) 14 NOV  (29) COMBAT LEADERSHIP  LM CH 29  IP 23 DUE
William B. Cushing  LE CH 9

(F) 16 NOV  (30) D-DAY & BLACKHAWK DOWN  LM CH 30  IP 24 DUE
CASE STUDIES

(M) 19 NOV  (31) Review, Special Topics,  SP 2 DUE
Leadership Dilemmas

*21-23 NOV  ****THANKSGIVING HOLIDAY****

(M) 26 NOV  (32) CASE STUDY: USS VINCENNES  LM CH 15  IP 25 DUE
& USS SARATOGA

(W) 28 NOV  (33) TWELVE O’CLOCK HIGH Part I  LM CH 24  IP 26 DUE
USS STARK / Under Fire

(F) 30 NOV  (34) TWELVE O’CLOCK HIGH Part II  NO HW

(M) 03 DEC  (35) TWELVE O’CLOCK HIGH Part III  NO HW

(W) 05 DEC  FINAL EXAM DUE; LAST CLASS  FINAL DUE